COMPUTER COMMUNICATION

IMPACTS AND IMPLICATIONS

The First International Conference on Computer Communication

Edited by
STANLEY WINKLER

OCTOBER 24-26, 1972 • WASHINGTON, D.C.

WINKLER

COMPUTER COMMUNICATION - Impacts and Implications

1972
ICCC

conference at a glance

	9 a.m.	10 a.m.	11 a.m.	12	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.	6 p.m.	7 p.m.	8 p.m.	9 p.m.
TUESDAY	1 **KEYNOTE SESSION** COMPUTER COMMUNICATIONS IN INDUSTRIALLY ADVANCED NATIONS *IBR Center*					2 RESPONSIBILITY *Lincoln West* [illegible] *Lincoln East* 4 COMPUTERS AND A TELEPHONE COMMUNICATION SYSTEM OF THE FUTURE *Monroe* 5 COMPUTER COMMUNICATIONS—AN EMERGING TOOL IN RESPONSE TO SOCIETY'S NEEDS *IBR East*					**CONFERENCE RECEPTION** *IBR East*		
WEDNESDAY	6 IMPACTS I *IBR East* 7 NETWORKS—RECENT DEVELOPMENTS *Lincoln East* 8 DATA NETWORK DESIGN PROBLEMS I *Lincoln West* 9 THE WIRED CITY *Monroe*			**CONFERENCE LUNCHEON** Speaker: Dr. Hans J. von Baeyer "CONFLICT IN COMPUTER COMMUNICATION" *IBR Center*		10 IMPACTS II *IBR East* 11 NETWORKS—HIGHER EDUCATION *Lincoln East* 12 DATA NETWORK DESIGN PROBLEMS II *Lincoln West* 13 THE ROLE OF COMPETITION *Monroe*						14 COMPUTER COMMUNICATION—THE QUIET REVOLUTION *IBR Center*	
THURSDAY	15 DATA BANKS AND INDIVIDUAL PRIVACY *IBR Center* 16 TELEPROCESSING—A COMPUTER UTILITY *Lincoln East* 17 COMPUTERS, COMMUNICATION & DISTRIBUTED HEALTH CARE SYSTEMS *Lincoln West* 18 SOFTWARE ASPECTS IN COMPUTER COMMUNICATIONS *Monroe*					19 SOCIAL CONCERNS *Lincoln West* 20 NETWORKS—PUBLIC DATA *Monroe* 21 COMPUTERS AND THE LIBRARIES OF THE FUTURE *Jefferson*							

COMPUTER COMMUNICATIONS:
IMPACTS AND IMPLICATIONS

The First International Conference on

Computer Communication

Edited By:
Stanley Winkler

Washington, D.C.
October 24 - 26, 1972

Copies available from:

Association for Computing Machinery
1133 Avenue of Americas
New York, New York 10036

IEEE Computer Society
8949 Reseda Boulevard
Northridge, California 91324

IEEE, Order Department
345 East 47th Street
New York, New York 10017

Designed and produced by Creative Communications, Associates, Inc.

ICCC 72 CHO 690-8C

PREFACE

It is almost two years since Reg Kaenel called a meeting of a small group of colleagues to discuss the organization of an international and interdisciplinary conference on computer communication.

The objectives set for the conference were: (1) to provide a free, international, interdisciplinary forum for technical and non-technical people concerned with computer communication, (2) to expose and debate the prevailing issues; to report and compare experiences; and to provide opportunities for interdisciplinary education, (3) to bring together in the conference proceedings a comprehensive document that presented the computer communication issues and techniques, and (4) to construct a meaningful experiment in computer communication aimed at revealing the merits of resource sharing and stimulating further resource sharing applications.

You now have before you the conference papers. The international character of the conference is readily discernable, as is its interdisciplinary nature. The Program Committee takes pride in what promises to be a very fine program.

A program and publication of this size is the result of dedicated hard work by many people. The names of most of these appear throughout the book. To those whose names do not appear for one reason or another, a sincere "thank you" is extended. While it is impossible to single out everyone whose contribution to this conference was essential, it is equally impossible for the editor not to acknowledge the guidance and support of the Conference Founder, Reg Kaenel, our General Chairman, Maurice Karnaugh, and the Chairman of our Executive Committee, Philip Enslow.

The support of the National Science Foundation was vital in enabling us to bring to the conference, participants who otherwise would have been unable to attend. We also appreciate the very real help of the sponsoring and cooperating societies who are listed on the acknowledgement page. Finally I wish to acknowledge the authors, speakers, and discussants who, of course, are really the conference.

Stanley Winkler

ACKNOWLEDGEMENTS

The first International Conference on Computer Communication (ICCC '72) is sponsored by:

ASSOCIATION FOR COMPUTING MACHINERY
IEEE COMPUTER SOCIETY
IEEE COMMUNICATIONS SOCIETY

with the cooperation of:

American Bar Association

American Federation of Information Processing Societies

American Society for Information Sciences

Association of Data Processing Service Organizations

Business Equipment Manufacturing Association

Comité Consultatif International des Téléphonique et Télégraphique

Data Processing Management Association

Federal Communications Bar Association

International Federation for Information Processing

National Academy of Engineering

National Academy of Science

National Institutes of Health

Society for the Application of Learning Technology

United Nations Educational, Scientific and Cultural Organization

Appreciation is expressed to the students from the American University, George Washington University, Johns Hopkins University and the University of Maryland who donated their time and assistance during the conference.

The Conference was supported by a grant from the National Science Foundation (NSF GJ 33239)

TABLE OF CONTENTS

PLENARY SESSIONS

RESPONSIBILITY, IMPACTS AND SOCIAL CONCERNS

NETWORKS AND RELATED PROBLEMS

Page

TECHNOLOGY

TOOLS IN RESPONSE TO A SOCIAL NEED

PLENARY SESSIONS

CONFERENCE LUNCHEON

TOASTMASTER:

Dr. Reg A. Kaenel
AMF Incorporated
Stamford, Connecticut, U.S.A.

DISTINGUISHED SPEAKER:

Dr. Hans J. von Baeyer, Director General,
Canadian Computer-Communication Task
Force, Department of Communications,

Ottawa, Ontario, Canada

"Conflict in Computer Communications"

KEYNOTE SESSION

Computer Communication in Industrially Advanced Nations: The Social, Political and Economic Dynamics of a New Technology

CHAIRMAN:

Dr. Maurice Karnaugh
IBM Thomas J. Watson Research Center
Yorktown Heights, New York, U.S.A.

KEYNOTE PAPERS:

Dr. Maurice Karnaugh, *"Opening Remarks"*

Yasuo Makino, Administrative Director of Telecommunications, Ministry of Post and Telecommunications, Tokyo, Japan, *"Data Communication in Japan"*

T. Larsson, Deputy Director General, Swedish Telecommunications Administration, Farsta, Sweden, *"Data Communication in Sweden – and Some Aspects of the Situation in Europe"*

KEYNOTE SPEAKER:

Dr. Clay T. Whitehead, Director, Office of Telecommunications Policy, Executive Office of the President, Washington, D.C., U.S.A., *"Data Communication in the United States"*

DISTINGUISHED GUEST:

Dr. S. I. Samoylenko, Vice-Chairman, Council for Cybernetics, USSR Academy of Sciences, Moscow, USSR

THEME: This session will explore the needs, genesis, status, and plans for computer communications in Japan, the U.S.A., the USSR and Western Europe. Sociopolitical and economic factors will be emphasized and their influences on the implementations will be examined.

OPENING REMARKS

Maurice Karnaugh
Thomas J. Watson Research Center, Yorktown Heights, New York, U.S.A

It is my pleasure to open this 1972 International Conference on Computer Communication by acknowledging our indebtedness to those who foresaw the need and who gave generously of their time and their creative energies: the founder of the Conference, Reg. Kaenel, the Program Chairman, Stanley Winkler, all other members of the Steering Committee and the Program Committee, the session chairmen, the authors and panelists, and the many volunteers who have assisted us. We have also been assisted by a National Science Foundation grant.

We are particularly grateful to those participants who have come from other countries and without whom this Conference would be much less valuable. International participation is no mere adornment to the Conference. It is a primary means toward achieving a diversity of interests and viewpoints. We are in the mainstream of revolutionary changes; and we need vision. Let us accept this challenge. Let us question old patterns of thought and expand our understanding.

Computer communication has many aspects. To the government of an emerging nation, it may seem to be another doorway into the international community and an opportunity, internally, to expand educational, medical and other services more rapidly than might otherwise be possible. An industrially developed nation, on the other hand, may view computer communication as, primarily, a tool for more efficient government, commerce, and industry in a highly competitive world.

There are functional differences of viewpoints also. Some young administrators may find this a field in which achievement and recognition may win the race with senility. Others may see it as a dragon with an insatiable appetite for investment capital. Lawyers will find a wellspring of legislative and regulatory puzzles. Ordinary telephone users see it, if at all, as yet another excuse for poor service. To potential users, it is both an opportunity and a bottleneck.

A brief consideration of the human condition may be useful in broadening our views. It has become painfully apparent in recent years that mankind has not achieved a viable relationship with this planet. We find ourselves in the midst of a violent transient period that began with the development of modern medicine and with the application of fossil fuel power to transportation and industry. The enormous energies thus released were uncontrollable in an ignorant and highly competitive socio-political environment: and they have brought about two centuries of unparalleled pillage, waste, and population growth.

Recovery from this crisis has barely begun in some parts of the world, while the fires rage uncontrolled in others. It is, therefore, much too early to predict whether the retreat into sanity will be orderly or castastrophic. We can, nevertheless, predict that progress in two technological areas will be necessary to an orderly recovery. These areas are power technology and information technology.

Despite this, we technologists should be among the first to point out that technology is not sufficient. There is, unfortunately, a cult that views technology as modern magic which, if we have enough faith, will indefinitely support exponentially growing consumption. The overwhelming evidence to the contrary has not yet been accepted by a public majority, nor has it been acknowledged by the spokesmen of government, industry, or labor. We must, therefore, point out the limitations of technology as well as its promise; and we must emphasize the proper use of our technological capabilities.

I have mentioned power technology and information technology as two keys to orderly survival. Power technology must find new answers to the problems of the depletion of fuels and the problems of pollution. Information technology, on the other hand, requires only modest natural resources and is not a major source of pollution; nor does it urgently require new scientific discoveries. It will be an essential tool to the regulation of any viable technology-based society that may emerge. It provides a means for computing values by projecting the consequences of well defined policies. It is capable of providing communication and control in the proper employment of scarce commodities or dangerous byproducts. It offers the hope of substituting communications for transportation wherever possible and of making necessary transportation much more efficient. It can offer cultural enrichment, improved educational opportunities, and better health services. It holds the promise that useful industrial activities will be performed at near optimal efficiency.

Obviously, these hopes cannot all be realized until society is ready for them. By then, every alternative will have been exhausted and time will be short. Let us do what we can to understand and define the problems; and let us do what we can to facilitate the necessary changes.

Computer communications is an important component of information technology, and we have many problems to consider. To begin with, the systems have got to be capitalized in real time and built with available skilled labor. How is this to be done in different countries? How does the flow of technology among the countries affect this process? How are priorities to be established on the basis of types of uses and locations of users? How are costs to be allocated? Who will operate the systems? How will they be regulated? Will there be competition? May private systems to be shared? How will intersystem compatibilities be assured?

Beyond these questions, we must examine potential applications and their requirements. And we must look into such vital policy questions as the right of access to data banks, privacy, and security. In addition, we shall explore some of the social impacts of computer communications.

The Keynote Session, which follows immediately, takes up the problems in providing and regulating computer communications in the industrially advanced nations, where these problems are currently most acute. It will be of interest to see how such essentially similar problems are addressed by the different countries represented.

DATA COMMUNICATION IN JAPAN

Yasuo Makino
Ministry of Posts and Telecommunications
Tokyo, Japan

Since the use of telecommunication circuits in Japan had been strictly controlled by the Public Telecommunications Law, provision of data communication services was not authorized for any entities other than NTT (the Nippon Telegraph & Telephone Public Corp.). The majority of enterprises who desired to operate data communication services, had to choose between setting up their own data communication systems by leasing computers from manufacturers or using the data communication services offered by NTT. It is expected that the data communication services by private enterprises will be increased, on and after the revision of the Public Telecommunications Law in Japan.

In Japan, the domestic public telecommunication services have been monopoliticaly operated by the single common carrier - NTT. Therefore it has been the carrier's duty to establish public telecommunication facilities, and set up, as a principle, terminal equipments.

While there was a growing need to connect computers, terminal equipments and others to the public telecommunication lines, it was prohibited for a user of computers to connect his own facilities (computers etc.) freely to the said lines.

Of course, it is permitted for "one" enterprise to use exclusively the public telecommunication line, but it is prohibited to use the same one public telecommunication line jointly between two enterprises, or to combine the leased line from NTT between two or more enterprises, who belong to a different category of industry each other.

These restrictions of law was to ensure the monopolitical operation of the carrier in the field of the public telecommunication services.

With the development of computer technology, these restrictive clauses of law became inadequate.

As many users of computer desired to use the public telecommunication lines freely, there occurred various arguments on the matter of "liberation of telecommunication lines" (which means to let it be free to use the public telecommunication line without any limitation) among political, governmental, economical and industrial circles.

From the point of governmental view, the M.P.T. (Ministry of Posts and Telecommunications), which is asked to keep the sound order of telecommunication operation in Japan, endeavoured to find a compromising arrangement between pros and cons on the matter of "liberation of telecommunication lines", aiming greater advantage of free use of lines, while avoiding the confusion of the telecommunication order caused by using the computer capability for message switching and line switching.

The concept of the revised law on computer/ communication is as follows.

(1) Telecommunication lines, computers, terminal equipments and others concerned to the data communication system are defined as a part of public telecommunication facilities, and are to be restricted by the public telecommunication law.

(2) Computers and terminal equipments connected to telecommunication lines are to be, as a principle, established by users.

(3) Users of telecommunication lines are, as a principle, prohibited to do message switching and line switching.

In this paper "Data Communication Service" means "Remote Access Data Processing Service" which is defined in the FCC's final decision of 1971 on the computer/communication problem.

Telecommunication Services in Japan

In Japan, domestic telecommunications circuits used for data communication service can be classified into two categories, i.e. (1) telecommunication circuits provided by NTT (Japan's domestic carrier), and (2) privately-owned telecommunication circuits installed by private enterprise under the restriction of existing regulations. In my country, inland public telecommunications service is exclusively operated by NTT, whereas overseas telecommunications service is operated by KDD (Japan's international carrier). Therefore, the use of telecommunication circuits provided by NTT for data communication service has been extremely rigid; namely, the conventional Public Telecommunication Law of Japan had the following problems:

1. Telecommunication circuits provided to users are limited to leased lines, and connection of a computer with the regular telephone network or regular telex network was not authorized.

2. Co-use of leased lines by two or more business enterprises was authorized only when the enterprises have close business relations with each other. Even when such use was authorized, a surcharge was imposed.

In accordance with the increasing demands for data communication circuits, efforts have been made toward a good use of and relaxation of the operating interpretation of the existing Law under the approval of the Ministry of Posts and Telecommunications. Such a variety of uses of telecommunications circuits for data communication was not even anticipated when the Public Telecommunications Law was established in 1952. Therefore, solution to the fundamental problems could not be obtained without revision of the Public Telecommunications Law.

The Ministry of Posts and Telecommunications initiated a study of a telecommunication circuits usage system for new data communication services in 1967; this resulted in a revision draft for the Public Telecommunications Law that was submitted to the Diet after thorough consideration and the opinions of various quarters were taken into account. The Law was passed on May 19, 1971.

Cutlines of the revised points of the law are as follows:

Two types of user contracts have been provided to cover cases where private enterprises wish to carry on remote data processing by means of their own computers connected to NTT telecommunication circuits. These are (1) the specified telecommunication circuit usage contract and (2) the public telecommunication circuit usage contract.

(1) The specified communication circuit usage contract

This contract applies in cases where the user wishes to connect his own computer or input/output equipment with a telecommunication circuit of a particular speed and a particular section designated by him.

In comparison to the former circuit usage lease contract with NTT, the conditions for the co-user lease are liberalized to a large extent, when it is recognized that it is necessary for two or more users to share the system in joint business operations.

With this revision, computer centers are now authorized to provide their customers with on-line computer services by leasing a leased line from the Corporation and connecting their computers with one end of such leased line and terminal equipment for such customer use on the other end of the line. Also, the conventional surcharge imposed for a co-user lease of a circuit usage contract was abolished. (See Table 5.)

(2) The public telecommunication circuit usage contract

This contract applies in cases where the user wishes to connect his computer or input/output equipment to a telecommunication circuit installed between the telephone or telex switching facilities and a location lesignated by him.

In the public telecommunication circuit usage contract, the customer-provided data terminal equipment can be connected to the public telephone network or telex network if the equipment conforms to the technical standards set forth by the Corporation with the authorization standard of the Ministry of Posts and Telecommunications, insofar as is necessary for the prevention of obstruction to the public telecommunication business.

This amendment was the answer to the user's requests for the liberalization of public telecommunication networks as to so-called "circuit problems".

In addition to the above changes persons who presently subscribe to telephone service may now install special terminal equipment (excluding CPU), and further use the lines they now have to provide, facsimile and electrocardiogram transmission via the regular telephone network. In this case, no further contract is necessary.

Data Communication in Japan

The number of actually operating general purpose computers was 11,237 sets as of Sept., 1971. This was a net increase of 41% over the preceding year. The value of these sets reached 1,024 billion yen, figured on the basis of sales value. This was a net increase of 38% over the previous year.

Of the total figures of all operating domestic and foreign machines, the number of domestic sets was 7,659 sets which corresponded to 68%. However, the value of these sets reached 562 billion yen, figured on the basis of sales value. This value was almost equivalent to a half of all domestic and foreign sets, which corresponded to 55%.

The past trends are listed in Table 1.

Data communication systems, (connecting computer directly to telecommunications circuits) can be classified into three major categories.

(1) A data communication system which employs privately-owned telecommunication.

(2) A data communication system which employs telecommunication circuits provided by NTT and terminal equipment or central computer unit installed by private enterprises.

(3) A data communication system, which employs computer and telecommunications circuits both installed by NTT. (Direct operation of NTT)

The growth of these three categories of data communication systems per annum is shown in Table 2.

Trends in computer industry

(1) Manufacturers

Japan's major computer manufacturers are mostly telecommunications and related equipment manufacturers. The technology obtained through the research and development of telecommunication equipment in past years has largely contributed to the rapid development of the computer industry.

At present, the computer market in Japan is occupied mostly by six domestic firms; Nippon Electric Co., Ltd., Fujitsu Ltd., Hitachi Ltd., Tokyo Shibaura Electric Co., Ltd., Oki Electric Industry Co., Ltd., Mitsubishi Electric Corporation, plus two firms operating on foreign capital, IBM Japan and UNIVAC of Japan.

As obviously seen from Table 1, Japan has increased its share of the domestic market by her domestic manufacturers.

According to a survey made by Japan Electronic Computer Company (JECC .... a rental organization jointly established by six domestic firms in 1961) in March, 1971, six domestic manufacturers and two firms operating on foreign capital have increased their share of the domestic market as shown in Figure 1.

The number of domestic operating systems has gradually increased over foreign ones, however, hardware and software technology in the large-scale high performance systems is still inferior to the foreign one, and thus the market share of domestic makes for large scale systems is smaller than that of the foreign machines due to the differences of capital investment. (See Figure 2).

In order to improve the hardware and software techniques, the development of "high performance computer", one of the large projects being undertaken mainly by the Governmental organization, was begun in 1966. This project has absorbed approximately 10 billion yen in research funds over a period of six years. On its completion, the computer will rank as one of the largest systems in the World.

The development of "DIPS" (Dendenkosha Information Processing System) project in NTT is also being carried out with the expectation that it will contribute largely to technological development of large-scale data communication systems.

In accordance with capital and trade liberalization, severe competition is being expected among domestic and foreign computer manufacturers. In such a case, competition power will be weakened when six domestic manufacturers are independently organized.

Therefore, it was determined to avoid unsuccessful competition, by re-grouping the existing six manufacturers into three groups with the guidance of the Government, i.e. - three groups - Hitachi Ltd. and Fujitsu Ltd., Nippon Electric Co., Ltd. and Tokyo Shibaura Electric Co., Ltd., and Oki Electric Industry Co., Ltd. and Mitsubishi Electric Corporation.

Data communication systems in NTT

Data communication systems provided in NTT are split into two categories.

One is designed at the user's specific request, which is represented by the domestic exchange communication service in the Nation-wide Local Bank Association System, the Motor Vehicle Registration System in the Land Transportation Bureau etc.

The other is designed for the joint use of a large-scale computer by a number of subscribers, where its service items are fixed and prearranged. This system is applied for the following three systems.

1) Calculation by Telephone System: This is provided for those subscribers who want to make calculations ranging from such a simple one as four rules of arithmetic, standard deviation and matrix, to the integration of the area of complicated figures with the help of stored library programs.

2) Scientific and Engineering Calculation System: This is provided for those subscribers who would perform high degree calculations in science and engineering by the use of programs entrusted to the subscriber's production.

3) Sales and Inventory Management System: This is provided for those subscribers who want to perform management and control of sales and inventory business with terminal equipment installed in the subscriber's premises such as shops, factories etc.

The present status of these systems is as follows

(1) Systems in operation

A total of twelve systems designed, installed and maintained by NTT is now in operation. The names and service opening dates are listed herein-after in chronological order.

1968; The Nation-wide Local Bank Association System and The Gunma Bank System opened their services, respectively.

1970; March, The Motor vehicle Registration System of The Land Transportation Bureau;
September, The Shizuoka Bank System, Sales and Inventory System in Tokyo, and Calculation by Telephone System in Tokyo commenced their services consecutively.

1971; January, Sales and Inventory System in Osaka;
March, Scientific and Engineering Calculation System in Tokyo;
June, Scientific and Engineering Calculation System in Osaka;
July, two banking systems, Kinki Mutual loans and saving Bank System and The Hokkaido Bank System, commenced their services, respectively.

The Tokyo and Osaka area financial Bank Association System, The Yokohama Bank System, The Chiba Bank System, The Tottori Nokyo System (or Agricultural Cooperative Association System) and Sales and Inventory Management System in Nagoya, commenced services in early 1972.

(2) Systems to be Inaugurated in the near future.

The Nation-Wide Inter-Bank Data Communication System, the largest system ever seen in Japan, is expected to open its domestic exchange service in April 1973, which is now in the process of programming.

In addition to this, system designs are now going on for the Chuo Sogo Bank System, The Taisei Mutual Financing Bank System, the seat reservation system for Towa Airline Service Co., Ltd., The Toyama Agricultural Cooperative Association System etc.

As to the joint use system, Scientific and Engineering Calculation System in Nagoya, and Sales and Inventory Management System in Nagoya, and Sales and Inventory Management System in Fukuoka and in Sapporo are now continuing their preparation in order to open their services in 1972.

The systems in operation and opened in 1971 are shown in Table 3. and Table 4.

Future Problems of Data Communications Policy

Although data communication in Japan has established firm ground for further expansion by the revision of the Public Telecommunications Law, it is still in the cradle and there are many problems to be studied and resolved for the achievement of sound development and expansion.

The basic points are as follows:

1. Development of National Technology

Japan has approximately 11,237 computers (as at the end of Sept., 1971), of which, foreign-made computers (mostly made in the U.S.A.) occupy 45% in terms of money, and 32% in number. In the field of data communication, U.S.-made computers are dominant, representing approximately 63% of central processing units.

If this trend continues and foreign-made computers come to play the greater part in information processing for the major firms of Japan, it will have a serious influence on the economy and on our society. Therefore, we have to accelerate the development of national technology as a basic policy, and in order to put this policy into effect, it is necessary to consolidate and stimulate research and development efforts for such items as follows:

a. Raising the level of software technology
b. Development of high-performance computers
c. Development of economical and diversified types of data terminals
d. Development of high-speed data communication circuits
e. Development of high-performance electronic switching equipments, making it possible to provide highly reliable switched networks and economically possible to provide various kinds of new and better data communication services

2. Standardization of Hardware and Software

In order to facilitate the interconnection of two or more data communication systems and to carry out data communication effectively and economically, standardization pertinent to the following items is necessary:

a. Standardization of hardware so that computer input-output devices, recording devices, etc. will be compatible with each other.

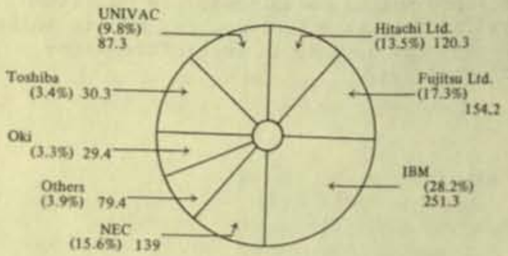
b. Standardization of software such as programming languages, data file methods, etc.

3. Promotion of Education and Training

Data communication cannot be developed rapidly and smoothly over a long period without the availability of skilled computer people, especially system engineers and programmers.

For this purpose, it is necessary to reconsider the existing education system of Japan and to consolidate new education systems of Japan and to consolidate new education systems both at school and for adults so as to meet new demands in the information-oriented era.

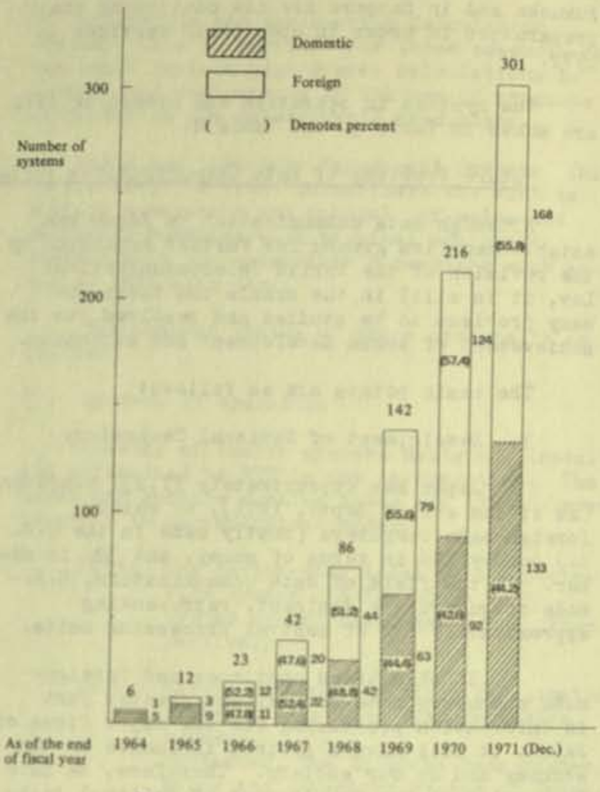
Figure 1 Market share of computers in Japan (JECC Survey, as of March, 1971)



Note: Others include Burroughs, NCR, Mitsubishi, and CDC. (Four Co.)

Figures in billion yen

Figure 2 Comparison of on-line data communication systems by domestic & foreign manufacture per annum



Note: A system that includes domestic & foreign machines in parallel is assumed to be two systems.

Table 1 State of Operating Systems

Top figure in each line shows number of sets, bottom figure shows value in ¥ million

Type		Mar. 1966	Sept. 1966	Mar. 1967	Sept. 1967	Mar. 1968	Sept. 1968	Mar. 1969	Sept. 1969	Mar. 1970	Sept. 1970	Mar. 1971	Sept. 1971
Japanese	Large	25 10,285	30 12,228	58 22,088	70 25,740	95 35,627	135 51,234	186 74,602	214 89,158	265 115,681	334 153,517	428 199,631	486 229,191
	Medium	450 46,519	509 52,653	612 62,724	701 72,396	831 84,326	1,011 100,580	1,183 117,427	1,350 133,373	1,655 165,147	1,942 193,849	2,299 229,926	2,579 261,827
	Small	197 4,904	252 6,083	534 13,103	676 16,191	887 20,345	1,086 24,706	1,311 29,217	1,556 34,189	1,824 40,089	2,073 46,048	2,358 52,730	2,531 57,828
	Very Small	401 2,553	417 2,720	544 3,509	627 4,057	671 4,229	738 4,889	858 5,603	1,016 6,666	1,214 7,967	1,447 9,551	1,626 10,698	2,063 13,395
	Total	1,073 64,262	1,208 73,685	1,748 101,426	2,074 118,386	2,484 144,528	2,970 181,411	3,538 226,851	41,36 263,388	4,958 328,884	5,796 402,965	6,711 492,986	7,659 562,242
	Increase over Preceding Year (%)	55.0	45.0	57.8	60.7	42.5	53.2	57.0	45.2	45.0	53.0	49.9	39.5
Foreign	Large	97 47,237	103 49,829	122 56,403	135 63,045	181 81,904	225 107,600	268 128,096	310 155,636	369 191,362	427 232,063	500 280,896	563 330,038
	Medium	439 53,464	479 59,273	500 60,420	544 64,406	557 65,668	580 68,308	629 74,466	685 79,407	721 80,899	788 86,337	871 93,806	935 102,241
	Small	328 9,272	369 10,422	236 6,511	282 7,768	324 9,059	357 9,950	433 11,808	462 12,448	605 15,688	722 18,241	822 20,142	895 22,753
	Very Small	0 0	0 0	0 0	0 0	0 0	0 0	2 5	8 43	65 326	217 1,208	578 3,391	1,185 7,245
	Total	864 109,974	951 119,524	858 123,335	961 135,221	1,062 156,632	1,162 185,859	1,332 214,376	1,465 247,536	1,760 288,275	2,154 337,850	2,771 298,235	3,578 462,277
	Increase over Preceding Year (%)	24.2	20.8	12.1	13.1	27.0	37.4	36.9	33.2	34.5	36.5	38.1	36.8
Total	Large	122 57,522	133 62,057	180 78,492	205 88,786	276 117,531	360 158,834	454 202,699	524 244,795	634 307,043	761 385,580	928 480,527	1,049 559,229
	Medium	889 99,983	988 111,927	1,112 123,145	1,245 136,802	1,388 149,994	1,591 168,889	1,812 191,893	2,035 212,780	2,376 246,046	2,730 280,187	3,170 323,732	3,514 364,068
	Small	525 14,176	621 16,505	770 19,615	958 23,960	1,211 29,404	1,443 34,656	1,744 41,025	2,018 46,638	2,429 55,777	2,795 64,289	3,180 72,872	3,426 80,582
	Very Small	401 2,553	417 2,720	544 3,509	627 4,057	671 4,229	738 4,889	860 5,608	1,024 6,710	1,279 8,294	1,664 10,759	2,204 14,089	3,248 20,640
	Total	1,937 174,236	2,159 193,210	2,606 224,762	3,035 253,607	3,546 301,160	4,132 367,270	4,870 441,227	5,601 510,924	6,178 617,160	7,950 740,816	9,482 891,221	11,237 1,024,518
	Increase over Preceding Year (%)	34.0	29.0	29.0	31.3	34.0	44.8	46.5	39.1	39.9	45.0	44.4	38.3

Table 2 Development of Data Communication System in Service

(as of the end of each fiscal year)

Items / Fiscal year		1964	1965	1966	1967	1968	1969	1970	1971 (Dec.)
No. of systems	Owned & operated by NTT	–	–	–	–	2	4	9	13
	Privately owned but circuits rented from NTT	5	10	20	35	75	122	188	265
	Privately owned and operated (including circuit lines)	1	2	2	5	7	13	17	21
Total		6	12	22	40	84	139	214	299
Rate of annual increase in (%)		–	100	83	82	110	65	54	40

Table 3 Systems designed at user's request

Systems	Services	Locations of Data Communication Centers	Dates of Service
The Nation-wide Local Bank System	Messages transmitted from the terminal of a bank with paper-tape, to the data communication center, are processed, and sent to the terminal of a called bank and printed on paper-tapes or ledgers. (Applications) On-line domestic exchange service between the 4100 branch offices of the 62 banks.	The Higashi-Ginza Telephone Exchange in Tokyo	October 1968
The Gunma Bank System	Same as above (Applications) On-line domestic exchange service between the 27 branch offices.	The Gunma Bank Data Communication Center in Maebashi City	August 1968
The Shizuoka Bank System	Same as above (Applications) On-line domestic exchange service between the 59 branch offices.	The Shizuoka Bank Data Communication Center in Shimizu City	September 1970
The Motor Vehicle Registration System of The Land Transportation Bureau	Messages transmitted from the terminal installed in each Land Transportation office, with the help of mark sheets, to the data communication center, are processed immediately. The results are sent back to the terminal and printed on motor vehicle registration sheets (Applications) Registration, inspection and administration service for more than ten million cars at 65 Land Transportation Bureaus throughout Japan	The 2nd Marunouchi Telephone Exchange in Tokyo	March 1970
1970 warld Exposition Data Communication System	Messages necessary for the control and administration are compiled at the center. They are processed and delivered to the variety of display equipment where they are needed. (Applications) Compilation, analysis and delivery of the following information (1) Number of visitors	Expo '70 Head Office in Osaka	March 1970 – September 1970

Systems	Services	Locations of Data Communication Centers	Dates of Service
	(2) Site Congestion (3) Vehicular Traffic (4) Lost Children or properties, and pertaining thereto announcements.		
The Kinki Mutual Loans & Saving Bank System	Necessary messages transmitted to the communication center by means of the teller equipment or the key-board printer, are processed almost immediately and all the appropriate ledgers kept in the center are updated. The reply is sent back to the terminal equipment and printed on the pass-book, the bond etc. (Applications) On-line banking system with some 50 branch offices, whose work includes handling ordinary deposite, loans, domestic exchange and supplementary business.	The Osaka Higashi Telephone Exchange in Osaka	July 1971
The Hokkaido Bank System	The bank teller inserts a pass-book into the terminal. Details of the messages are keyed in and sent to the center. Processing is made immediately and the result is sent back to the terminal and printed on the pass-book. (Applications) On-line saving bank service, with 40 branch offices of the bank.	Head office of the bank in Sapporo	July 1971
Tokyo Area Financing Association System	About same as that of Kinki mutual loans and saving bank system (Applications) On-line banking service with about 60 financial banks including 400 branch offices, which are members of the association. Its service covers ordinary deposit, fixed deposit, fixed saving account.	The Ohtemachi Telephone Exchange in Tokyo	March 1972
Osaka Area Financing Association System	Same as above (Applications) On-line banking service with about 30 financial banks including 120 branch offices. Its service covers ordinary deposit, fixed deposit, fixed saving account, current account loans.	The Kanmoku Telephone Exchange in Osaka	April 1972

Systems	Services	Locations of Data Communication Centers	Dates of Service
The Tottori Agricultural Cooperative Association System	Handling process is about the same as above. But in this system, in case a branch office does not require a terminal equipment, in view of the traffic of messages, the messages are transmitted to the data communication center with a fascimile equipment installed in the branch office, by way of another office providing a terminal equipment. (Applications) Deposit (ordinary, fixed, etc.), purchase and sales service, with about 70 branch offices belonging to unit agricultural cooperatives affiliated in the association.	The Tottori Telephone Exchange in Tottori City	Near future
The Yokohama Bank System	Same as that of Kinki Mutual Loans and Saving Bank System (Applications) The number of the branch offices is about 100.	The data communication center of the bank in Yokohama city	May 1972
The Sapporo Winter Olympics System	Registration of records, officers, athletes, etc.	The Shin-odori Telephone Exchange in Sapporo	February 1972

Table 4 Prearranged Systems in NTT

Systems	Location of Data Communication Centers	Date of Service
Calculation by Telephone System	Ohtemachi Telephone Exchange in Tokyo	September 1971
	The Shin-Toyosaki Telephone Exchange in Osaka	March 1971
Scientific and Engineering Calculation System	Ohtemachi Telephone Exchange in Tokyo	March 1971
	Shin-Toyosaki Telephone Exchange in Osaka	June 1971
	Nagoya Data Communication Center in Nagoya	Around the first half of 1972
Sales and Inventory Management System	Shin-Kasumigaseki Telephone Exchange in Tokyo	September 1970
	Ebisu Telephone Exchange in Osaka	January 1971
	Kikui Telephone Exchange in Nagoya	March 1972
	Naka Telephone Exchange in Fukuoka	Around the second half of 1972
	Odori Telephone Exchange in Sapporo	Same as above

Table 5. Systems operated under Specified Circuit Usage Contracts with NTT

Service		Number of users	Number of circuits					Total
	Service		A – 1 50 B/S	C – 2 200 B/S	D – 1 band use	D – 5 1200 B/S	D – 7 2400 B/S	
Banking	Deposit	21		260		1,151	9	1,420
	Money order	18	1,131	492		30	9	1,662
	Deposit money order	24	1,185	602		1,777	8	3,572
	Trust	2				31		31
	Sub-total	65	2,316	1,354		2,989	26	6,685
Management	Production, Sales & inventory management	93	158	972		98	42	1,270
	Reservation	6	522	73		15	3	613
	Management	5	2	17		10	11	40
	Others	1		19		25		44
	Sub-total	105	682	1,081		148	56	1,967
Government	Air pollution observation	13	6		349			355
	Scientific and engineering calculation	5	26	1		5		32
	Traffic control	12	64		1,010			1,074
	Others	11	986	73		21	2	1,082
	Sub-total	41	1,082	74	1,359	26	2	2,543
Stock dealings, scientific and transport management etc.	Stock dealings	10	131	597		10	49	787
	Transport management	6	236	44		17		297
	Engineering calculation	15	8	30		9	28	75
	Message communication	5	70	7		1		78
	Trustee calculation	5		10		1	5	16
	Insurance	5		11		5	1	17
	Others	8	2	113	20	14	1	150
	Sub-total	54	447	812	20	57	84	1,420
Total		265	4,527	3,321	1,379	3,220	168	12,615

(as at January 1st, 1972)

DATA COMMUNICATION IN SWEDEN—AND SOME ASPECTS OF THE SITUATION IN EUROPE

Mr. T. Larsson
Swedish Telecommunications Administration
Farsta, Sweden

Summary. The lecturer deals chiefly with the conditions in his home country, Sweden, but also with some aspects of the situation in Europe.

Telecommunications in Europe are often handled by PTT administrations on a monopolistic basis. An interaction necessarily arises between PTT, computer industry and users, all with their various problems.

There is a trend in Europe to the use of public data networks.

Characteristics of private functional data networks and public data networks are described, and the differences between the two types are analysed.

International standardization of data communication is necessary, particularly in Europe. European cooperation, however, is taking place not solely in the field of standardization. A very extensive market survey of the whole of Western Europe is being carried out by fifteen countries jointly.

1. General

1.1 Telecommunications administrations

I represent a small European country - Sweden. But in many respects the development in Sweden is similar to that in the rest of Europe, and the Swedish situation therefore may often be said to reflect the conditions in Europe. Under the government of each country there are ministries for telecommunications and other matters. In Sweden, for example, there is a Ministry of Communications with separate administrations for posts, railways and telecommunications. The Telecommunications Administration is responsible for all forms of telecommunications (telephony, telegraphy, telex, data, radio and TV distribution, but not program production). In most European countries posts and telecommunications are combined within one organization. In some countries there are also private telephone companies.

1.2 Computer industry

The Telecommunications Administrations come into contact with the computer industry and with data communication in the capacity both of user and of seller, a fact which in some cases creater problems. Apart from the Swedish computer industry (SAAB, L M Ericsson) most computer equipment manufacturers in Sweden are international, e.g. IBM, Univac, Honeywell Bull, ITT, Siemens etc. All of these large international computer companies are well established on the Swedish market, often in the form of subsidiaries. Similar conditions exist in several other European countries.

The computer industry is advancing very quickly, one reason being that the multinational computer enterprises have such a strong financial situation and can invest large sums in development and marketing.

1.3 Users of data communication

Users in Sweden are large customers both of the Telecommunications Administration and of the computer enterprises. Among the very large customers are government agencies, private national and multinational industrial enterprises, banks etc.

Industrial enterprises, banks and government agencies have a need both for data processing of their own and for communication with one another and, for example, use data stored in other data banks. The need for data communication is bound up with the efforts at rationalization and the requirement of efficient decision processes.

The multinational undertakings also have a great need for transmission and processing of data on an international basis. Cooperating enterprises are also dependent on each other's data, for example for decision-making and coordinated planning of their production.

Customers of a partially different type are municipal and regional services, e.g. labour market boards, health service, taxation authorities, police etc.

1.4 Relations between Telecommunications Administrations, computer equipment manufacturers and customers

The basic factor in the work within the computer field is the customers and their desires. If a customer wishes to use computers in his enterprise, he must procure the necessary equipment from a computer manufacturer.

In conjunction with his data processing a need fairly soon arises for telecommunications, as the customer finds it necessary to establish some cooperation with other units.

In the case of very complex systems, however, it is difficult for the customer himself to decide on the most suitable form of equipment. The computer industry is there to advise him, but at the same time he becomes tied to and dependent on the supplier. In the same way, the customer will be dependent, for his transmission problems, on the policy and equipments of the Telecommunications Administrations in view of the monopolistic nature of the telecommunications services in Europe.

Some interaction, though on an insufficient scale, occurs between computer manufacturers and Telecommunications Administrations. New customer requirements may lead to new products, which will be furnished by the computer industries. These new products are often based on data communication. A new transmission technique introduced by the Administrations brings new means for the use of computers. New computer equipments place new demands on transmission technique and applications, for the realization of which they are dependent on the Administrations.

The main difficulty in the interaction between the telecommunications and computer industries is that telecommunication networks represent large investments over a long period and are very complex. This means that new services and new transmission media cannot be introduced on a wide scale over a short period. As the Administrations are dependent on investment funds from their governments, it is in the last resort the priority alloted by the government which decides how soon new investments on the large scale that is required can be made. The picture for the computer industry differs to some extent. The investments neede to make are not so large. It is also financially stronger than the Administrations. The picture will partially change in the future when data communication becomes a service to the ordinary consumer. The large sums which will then be invested in data communication will not permit such fast changes as we have to-day.

The European Administrations will probably solve the data transmission problems through some form of public data networks, since solely private solutions, involving for instance problems of intercommunication in the "functionally" private networks, cannot be considered satisfactory. By "functionally" private networks I mean circuits leased by the Administration, which thus "function" as private networks. In the following these "functionally" private networks will just be called private networks. These public data network will in turn provide the computer industry with new possibilities of system design based on the necessary standardization. An effective interaction will then have started between the computer industry and the Administrations.

2. Development of data communication

2.1 History

The computer industry of to-day is usually international. This applies to the suppliers both of computers and terminals. In recent years special software houses and consultant offices with an emphasis on the use of data and on system investigations have grown up in large numbers. As a result of the very severe competition different constellations arise at frequent intervals. For the various national requirements, subsidiaries are usually established within the large computer companies.

In recent years comparatively new but very progressive branches have grown up also within the computer accessories industry, in particular data service bureaux of different kinds which undertake more or less specialized data processing, often via the telecommunications network. It is principally in the latter fields that a considerable pressure arises against the Telecommunications Administrations, as these service companies sometimes have special requirements, e.g. for short set up times and high speeds.

As most computer equipment companies have their origin in the USA or have very large user areas in the USA, the system are created primarily for American standards. This means that the systems produced will be adapted to American standards and to the organization and rate structure of the American Telecommunications Companies. The systems developed for the USA spread in due course to Europe and the modifications made to them must be based on the equipment built up for American conditions. As I have already pointed out, international standardization is necessary if the future development is to be favourable.

2.2 Customer requirements

The requirements placed by customers on the computer industry and the Telecommunications Administrations depend on the level of economic development, integration tendencies between different fields of activities, internationalization, rationalization etc.

Industry is moving to an increasing extent towards the creation of large units often with sub-units specialized production. Structural rationalizations are taking place on a very large scale and will probably continue, resulting in the closing down of less profitable enterprises. This leads to the establishment of increasingly large units within a branch of industry, which in turn places requirements on suitable forms of communication between the enterprises operating within the same branch of industry and also between the units of one large undertaking. The creation of economic blocs both within a country and between different countries means that structural rationalizations will not only be a common phenomenon within one country, but that they will often result in the formation of multinational enterprises or in cooperation between enterprises.

In such case telecommunications will be the necessary base in a chain of information and joint planning. The national and international use of computers must be foreseen as a basic requirement for planning purposes and for control of the production processes. There is, already a need today - and this will grow greater as time goes on - for the joint use of certain data banks with basic data required for a company's data processing.

With the present rapid rate of development it is difficult to foresee customers' reqirements for more than three years ahead. In order to improve the situation, a market survey called Eurodata was started in the Autumn of 1971, to be carried out simultaneously by fifteen European countries, and on the same premises in the various countries. I shall return to the question of this survey later.

2.3 Trend towards public data networks

The first modems were introduced in Sweden in 1962. There was already a great interest in data communication at that time, which has since continued. But it was not until around 1967 - 68 that the interest became concretized in the form of some large orders to the Swedish Telecommunications Administration.

Up to that time data transmission had taken place chiefly on the public telephone network, but now that large real-time systems were being planned, chiefly for banking purposes, it became clear that the demands, chiefly for short set up times and fast transmission, could not be satisfied by the public telephone network. It therefore became necessary to make use of point-to-point private networks. Some of these networks became very extensive.

A definite disadvantage of the private networks is their lack of intercommunication facility. This lack is due to a number of factors, the different structures of network, different transmission speeds, different line procedures etc.

From the point of view of the Telecommunications Administrations the maintenance of the private networks may also be fairly complicated, with divided maintenance responsibility, many different types of equipments connected to the networks, etc.

If we look at the various private networks that have grown up in Sweden, one sees that practically all of them follow the same geographical pattern, covering the places where there are concentrations of population. A number of parallel networks grow up throughout the country, each with its maintenance routine.

It would be natural to attempt to replace the many private networks by a public network so designed as to satisfy the growing demands of customers. In this way there would be facilities for intercommunication between different systems, and planning and maintenance would be simplified for the Administrations. But it must be pointed out that the parallel between a public data network and a public telephone network cannot be drawn too closely. The traffic interests are in principal different. In the data case the traffic interest is concentrated to traffic between the terminals of the system and its central computer. The traffic between different systems is as yet of less importance.

Within practically the whole of Europe the Telecommunications Administrations are considering how the problems of data transmisson should be solved in the long run.

Practically all countries have arrived at the conclusion that the long-term solution must be some form of public networks, which permit intercommunication on a national and international basis between the various customers. The present customer requirements of intercommunication facilities between different private networks cannot be satisfactorily met owing to the lack of a standard for these networks.

For some years there has been a trend towards public data networks, but acceptable solutions for them in the various countries require agreement on standardization. In the meantime, probably, different interim or experimental networks will grow up. Later, it is hoped, it will be possible to fit these latter into a more permanent solution. The reason for first introducing an experimental network is that one cannot afford to risk investing in a system which will not later comply with an international standard. Other ressons why we are at present holding back is that we know too little about the customer requirements which exist, and that we have not been able to make satisfactory cost analyses. The various interim and experimental networks, however, will provide better means for assessing the requirements of future data networks and will also provide valuable experience in the fields of development, technique, operation, maintenance and economy.

3. Public versus private networks. Differences

As I have already said, the private networks were introduced chiefly because the public telephone network (and telex network) could not satisfy the accentuated demands made by users. It is perhaps especially the long set up times (of the order of 10 - 20 seconds) which were inadequate. Actually the technique is largely similar, with modems connected to the telephone circuits, as is done also in the private networks.

The modems which the Swedish Telecommunications Administration can supply correspond roughly to those supplied by other European Administrations. A list of these modems will be found in fig 1.

A schematic illustration of various equipments which in principle can be used in a transmission system with modems and the division of responsibilities between the Administration and its customers is shown in fig 2.

Telex and telegraph circuits can, of course, also be used for data transmission, but experience shows that the interest in this type of transmission has not been especially great except in certain special applications (message switching). The reason is probably the comparatively low transmission speed (usually 50 bits/s), and in certain cases the restrictions on the transmission code.

Private networks often include concentrators, which concentrate the traffic from a number of terminals into the central computer. Two principles have been used for connection of the terminals, the multidrop principle and the loop principle. The two principles are shown in fig 3. The various terminals are connected by means of polling signals either from the central computer or from the concentrators. The procedures vary depending on the supplier of the computer.

The Swedish Administration, like several other Administrations, has made a careful study of the possibilities of introducing a public data network. The principles both of line switching and packet switching (message switching) have been considered. The studies have shown, however, that the line switching principle is probably the one which would offer the best service at the lowest price for the majority of customers.

The study carried out within the Telecommunications Administration was built up from a model for Sweden with three data switching exchanges to which concentrators could be connected so that the whole country could be covered. An investigation was made of a synchronous network, an asynchronous network and a hybrid between the two.

Customers with heavy traffic and/or in need of higher speeds could be connected direct to the concentrators. Otherwise the customers would be connected through the local network via modems to the concentrators. Asynchronous data transmission should be used in the local network up to the customers. Speed between the concentrators and the data switching exchanges should be 48 kbits/second in synchronous time multiplex. See figure 4.

Some of the characteristics being studied for the data network are:

- rapid set-up time
- flexible speed of transmission
- flexibel codes and code conversion
- definite secrecy
- low fault rate
- customer categories

- abbreviated dialling

- hot line

The detailed system study of a possible Swedish data network covers a number of different speeds. To enable customers to choose their own speeds there is a division into classes (50 - 300 bits/second, 600, 2400 and 9600 bits/second). The set up time will be short (≈ 100 milliseconds) and the congestion low (≈ 1 p.c.). Operating reliability will be greater than in the public telephone network; among other things essential equipment in the network will be duplicated.

It has been decided to etablich an experimental network with about 2000 customers at the end of the seventies. The present experimental network is expected to be put into operation in 1973 - 74. The experimental network will be concentrated to the three largest cities, Stockholm, Gothenburg and Malmoe. It will be built up mainly around a data switching exchange in Stockholm and, in addition, concentrators in Gothenburg and Malmoe. If the network turns out well, and if some type of international standard or at least a European standard exists when the experiment is ended, a decision will probably be made to introduce a public data network. The final design of the network will depend on whether a standard exists and on how the experimental network and its characteristics turn out.

The introduction of a public data network will result in demands for a certain standard, for instance concerning the interface equipment sending data, signalling and procedures for sending. That makes it possible to build into the network operational supervision equipment complying with high customer demands. In broad outline one can probably predict that all types of operational supervision will be built into software or hardware in the data network system.

The technical standard applied in the Swedish experimental network will link up with international standard drawn up by the international standardization bodies, CCITT, ISO etc. If no international standard exists, a nationl standard must be drawn up in cooperation with the users. Special cooperation will take place between the various Scandinavian countries.

It may be foreseen that traffic between the public data network and the telephone and telex networks will be introduced at a later time. As regards the telephone network one may foresee, in particular, traffic from simple terminals connected to the widespread telephone network. But as the public data network is designed solely for digital transmission, and for serial transmission, terminals of pushbutton type with parallel transmission of signals and audible response are not likely to be used. A video telephone service will not initially be introduced in the data network.

If a data network is to be attractive from the customers' point of view, it must not have a bit error rate greater than 10^{-6}.

For the Telecommunications Administration a fundamental demand must be reliability, with long total MTBF, and that the equipment can operate at a reasonable cost. An important condition as regards reliability of operation is that, at an early stage, stress is laid upon the fact that the exchanges in a data network must function unattended day and night. The MTTR for the exchange, excluding travell time etc., should be not more than 30 minutes.

The share of faults caused by the exchange in the stream of bytes should not exceed 10^{-7} measured over about 10^9 bits in the busy hour traffic (measured at the interface between the local network and the exchange).

Congestion caused by technical faults should be disregarded in relation to the traffic congestion and for that reason must be less than 0,1 p.c. The maintenance facilities for the system must, of course, include for instance blocking/unblocking of equipment, calling into service of diagnosis programs, connecting/disconnecting of equipment for testing of circuits and switching equipment and other similar functions.

A public data network will require fairly heavy investments and will not be profitable during its early period. The cost estimates that have been made in Sweden show that a private network with high utilization would today be less expensive than a corresponding service in a public network, provided that there is no need for intercommunication with other systems. The need for future intercommunication between systems is also difficult to assess, so that a correct cost analysis cannot be made.

The public sector in Sweden has a number of data teleprocessing systems under implementation or planned. For some fifteen of these systems an evaluation has been made as to whether a common data network would be profitable. Owing to the lack of standardization such a network must be based on message switching. The evaluation shows roughly the same cost for a common network as for individual networks. Here again, despite well defined systems, there was a great difficulty in assessing the need for intercommunication. It was decided to await internal tests to be made by the Administration on a public data network in 1973 - 74 and for the time being to build the necessary networks in the form of separate private networks.

At the introductory stage, accordingly, the public network cannot offer lower charges to customers than a private network. As the public network is extended, however, the cost of a service will be lower than that offered by a private network. There will also be the possibility of additional service facilities and a better maintenance service. If a very large number of private networks are implemented, there are likely to be considerable maintenance problems. The main advantage of a public data network will, of course, be the intercommunication facility both nationally and internationally.

The disadvantage of a public data network is that data processing systems, in which the computer connects terminals by polling, will be difficult to establish. A partially different system philosophy will therefore be necessary. This requires, too, that the Telecommnuications Administrations are well familiar with system engineering questions and can provide the necessary service to customers in this respect.

Plans for public data networks exist, as I have said, in several European countries. In general they relate to the introduction of networks of line switching type, but packet switching has also been studied.

Let me give you a few examples.

Federal Republic of Germany: An asynchronous network with line switching is to be started on trial in 1972.

France: A synchronous network with line switching (PCM) is being studied. A temporary solution based on a special telephone network and modems will be introduced first.

Spain: A network of message switching type is in operation as from 1971.

United Kingdom: A synchronous network with line switching is being studied in the first place, but a study is also being made of packet switching.

4. Efforts at international collaboration in the data communication field

As always with a new technique are innumerable problems, especially at the start. Data communication is no exception. Owing to its international aspects many solutions must be sought on an international basis. A number of international organizations are also working on data communication questions - standardization, policy questions and market questions.

Examples of strictly international standardizations are: ISO (International Standardization Organization), IEC (International Electrotechnical Commission), ITU (International Telecommunication Union), CCITT (Comité consultatif international télégraphique et téléphonique), IATA (International Air Transport Association), SITA (Société Internationale de Telecommunications Aeronautiques), IFIPS (International Federation of Information Processing Societies), IFAC (International Federation of Automatic Control).

The best known of these are presumably CCITT and ISO.

Examples of organizations in Europe working on standardization questions within the data communication field are CEPT (Conférence européenne des Administrations des postes et des télécommunications), ECMA (European Computer Manufacturers Association) and CENEL (Comité Européen de coordination des Normes Electrotechniques).

An international body, which in recent years has devoted some attention to data communication from the general policy angle is OECD (Organization for Economic Cooperation and Development). Data communication questions have also been brought up in connection with technological cooperation projects within the EEC.

One of these technological projects (project 11) is a European data communication network between certain universities and research centres. This network, which will work on the "packet switching" principle, is reminiscent of the ARPA network. Secondary networks can be connected to nodal centres (fig 5). Nodal centres will exist in Italy, France, Switzerland, the United Kingdom, and within the EEC administration. Norway, Sweden, Portugal and Yugoslavia have also joined the project.

Another technological EEC project is technological forecasts in the telecommunications field (project 20). A preliminary short-term forecast of data communication has already been made.

5. Eurodata

An interesting example of European cooperation is the market study within the data communication field being made jointly by fifteen European PTTs.

Market studies on data communication have been made in several European countries. In some cases they have been very ambitious. A lack, however, has been that they dealt solely with national aspects, which is hardly satisfactory in the international situation of today.

In the Spring of 1970, therefore, discussions took place within CEPT on the prospects of carrying out a joint market study for Western Europe. A specification for the study was written in the Autumn of 1970, invitations for tenders were issued in the Spring of 1971, and a contract was signed with the firm selected in the Autumn of 1971. Fifteen countries are taking part in the study. Some 50 industrial sectors and about 200 geographical areas are being investigated. The period for the forecasts will be 1972 - 1985 with the emphasis on the years 1972, 1976, 1980 and 1985. The study is being carried out by PA International Management Consultants, London, with Quantum Science Corporation, New York, as subcontractor, and for Italy the firm Italsiel. The result will be presented in the form of four reports on different aspects and a final summarizing report. A sketch of the research program is shown in fig 6.

This complicated study, probably the largest of its kind hither to, at all events in Western Europe, has necessitated a special organization within the fifteen sponsoring countries.

The study is to be completed by the Spring of 1973 and has therefore now reached halfway. Although no results can yet be presented, the following points nevertheless deserve mention:

5.1 Present Status of the Study

At the beginning of September the main programs of field research are virtually complete, together with the supporting economic studies in each of the fifteen countries. The field research programme has involved some 7000 respondents throughout Western Europe and is providing very detailed information about their existing data communications usage and future needs, and about the infrastructure of their organizations to provide the basis for calculation of future demands for data transmission. Interviews have been conducted with manufacturers of computing and telecommunications hardware in Europe, in Japan and the USA.

From a research viewpoint, the feature of the Eurodata Study which distinguishes it from most other projects of this type in the data communications field is its emphasis on data traffic patterns. It is concerned not only with computing systems and terminals but with the volume, characteristics and direction of the data traffic flow along over 10,000 links between the 200 telecommunications planning regions designated by the participating countries. For this reason the research programme has involved not just a sampling, but a complete census of all current and imminent users of data communications in Europe. The resulting data base will be the most comprehensive of its type every assembled and it is of course designed to provide a basis for continuous updating, as future variances from the original forecasting assumptions come more sharply into focus.

Whilst it is not yet possible to present detailed results from this project, some general patterns and trends have emerged which are of interest.

5.2 Some key features

Before discussing the trends in terms of industries and applications, it is worth mentioning that the response of organizations in industry and government in all countries to the Eurodata Study has been remarkable. Nearly all of the roughy one thousand firms who were asked to contribute personal interviews readily agreed to do so. Although the cynic might question the motives underlying this desire not to offend the Telecommunications Administrations, the realist is encouraged by the genuine desire to encourage the provision of good data communications facilities.

5.3 General Structure of Data Traffic in Europe

In Western Europe there are between 1500 and 2000 organizations who use the data communications services of the PTTs as part of their own in-house EDP systems. A further 3000 organisations use remote computing and terminal facilities provided by data processing service companies. Of these user organizations, an estimated 10 p.c. account for over 80 p.c. of the total volume of traffic. This emphasis on massive use of data communications by a relatively small number of key organizations is thrown into further relief by the fact that half the modems in use in Europe are operated by only forty users.

In the commercial and industrial sectors, multinationals dominate the scene. In discrete and process manufacturing, for example, approximately 90 p.c. of the modems are used by companies who are members of multinational groups.

Amongst the larger European countries, the United Kingdom is the main user of transmission-based systems in most industrial sectors. There appears to be no single reason for this dominating position, and it is anticipated that it will not continue for much longer.

5.4 Industries and Applications

At present the major applications of data transmission are found where large volumes of relatively standardised transactions are handled in situations where customer service is at a premium. Thus airlines, banks and manufacturing and retail sectors are well to the fore. Reservation applications, customer account maintenance and advanced systems for order entry and inventory control are the main applications.

5.5 Trends

Growth is expected due to further penetration in these areas, and also in the data processing serivce industry, in government and in the insurance sector. The factors promoting growth are the enlargement of the EEC, government policies of regionalisation and the falling relative costs of data processing systems components. However, there are several checks and balances in this situation and in Europe generally the following factors are likely to influence the growth pattern:

- The lack of experienced staff to write application software;
- The need for international standards in hardware, software and transmission codes;
- The impact of low cost minicomputers. Already one major European multinational is decentralising data processing to the plant level and providing each establishment with its own minicomputer to run its own affairs;
- The increased cost-consciousness in all data processing applications, forced by recent economic reverses in many European countries.

5.6 Major Markets for Data Transmission Equipment in Western Europe

Federal Republic of Germany, France, Italy and the U.K. are likely to provide the highest volume of demand. The Nordic countries also, with their high G.N.P. per capita and long lines of communication between parts of the same organization, are expected to be a key market area.

Replacement demand by airlines, banks and data processing bureaux will be augmented by new demand arising in, for example, the insurance and retail, wholesale sectors.

Generally the need is for equipment and software which is tailored to meet the user's requirements in each key market area. This applies particularly to systems, applications software and terminals. With concentrators, multiplexers, modems and front-end processors there is of course greater scope for the exploitation of standard products.

6. Conclusion

If a comparison is made of the data communications in Sweden and Europe with those in the USA, it may undoubtedly be said that the USA is far ahead of Europe in the number of data communication applications. It may also be said that the Europeans have had the advantage, at an early stage in their development, of being able to study the development in the USA and can therefore perhaps advance more quickly towards more sophisticated solutions. Among other things it appears as though the tendency to public data networks is very strong

The strong position of the PTTs and the Telecommunication Companies and the necessity of creating an international standard will lead to a development in the direction of standardization and compatibility between different systems, which will be of advantage to all parties (not least the customers).

Charges for modems

	Initial fee	Annual fee
Serial modem 200 bauds	211 $	261 $
Without automatic switch phone/data	211	287
With automatic switch phone/data		
Serial modem 600/1200 bauds		
Without return channel, without synchronising	211	287
Without return channel, with synchronising	211	362
With return channel, without synchronising	316	371
With return channel, with synchronising	316	447
Serial modem 2400 bauds	316	590
Serial modem 4800 bauds	632	278
Serial modem 40.8 kilobauds	1686	2613
Parallel modem 75 characters per sec.		
Transmitter	211	147
Receiver	421	822
Terminal for audio response	211	26

Figure 1.

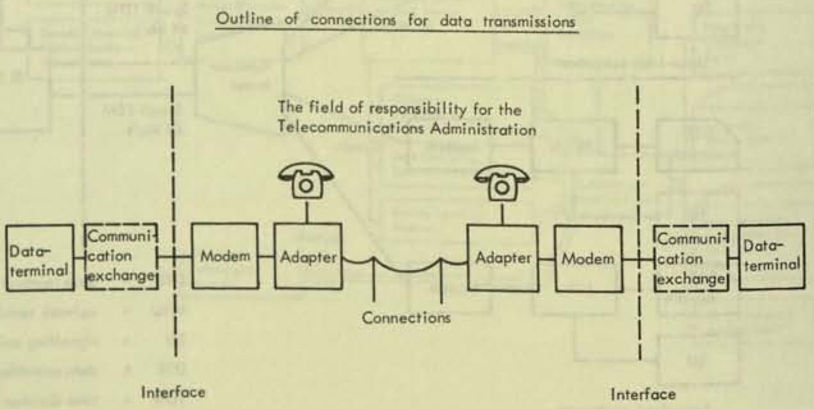


Figure 2.

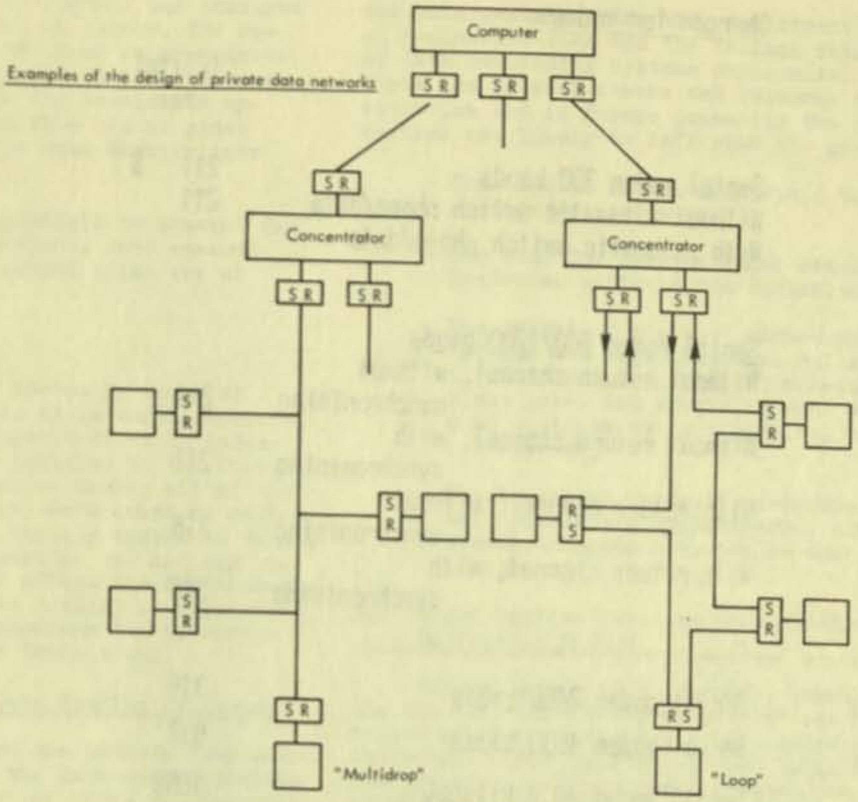


Figure 3.

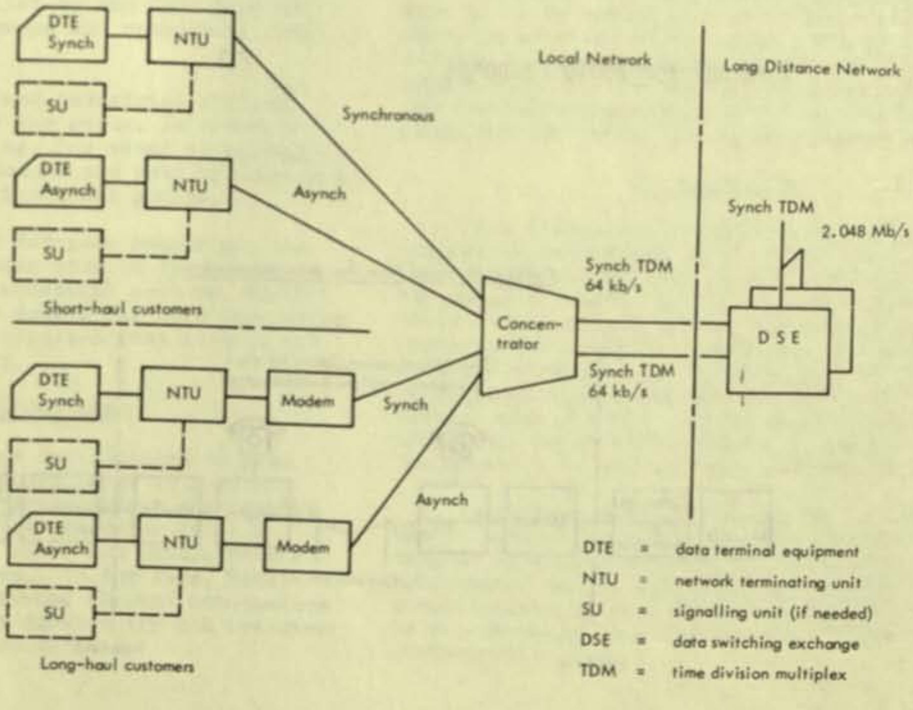


Figure 4. A local public data network according to the Swedish data network study.

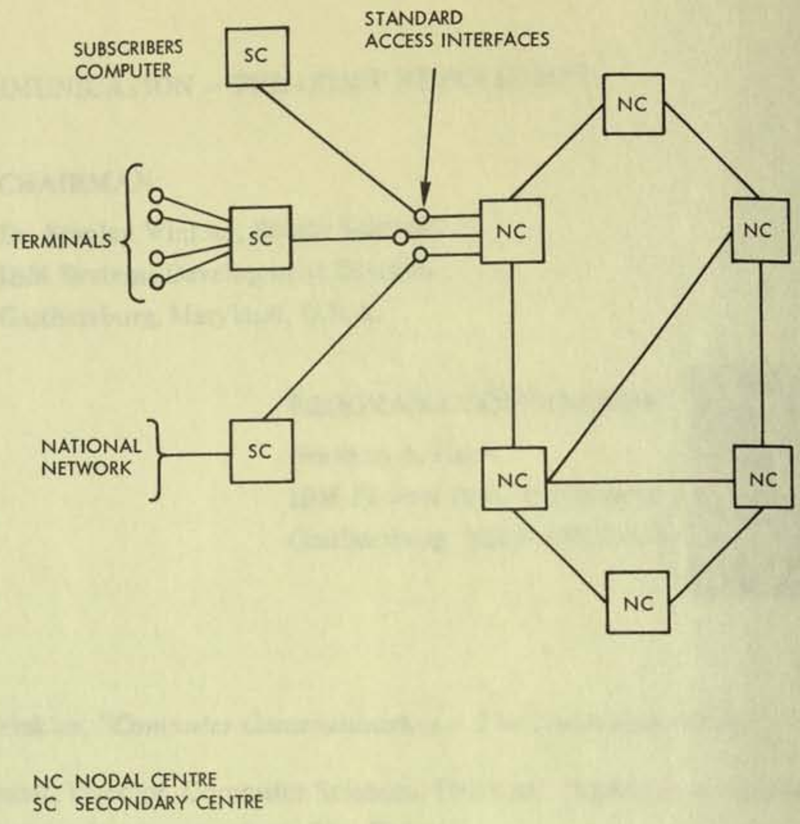


Figure 5. European data communication network according to Project 11.

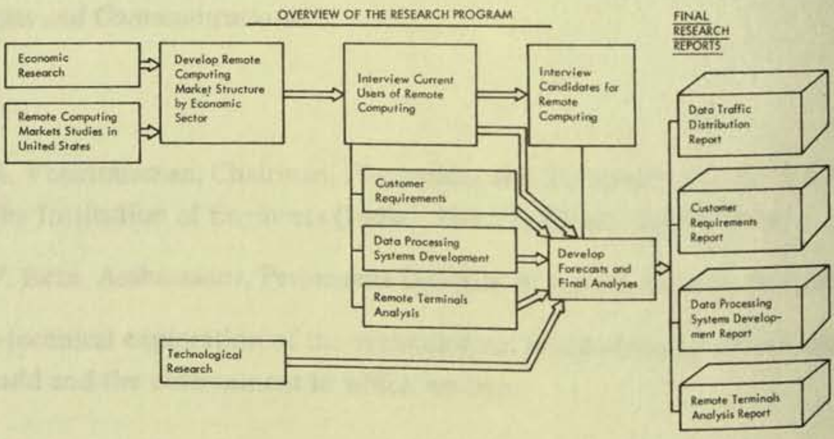


Figure 6.

COMPUTER COMMUNICATION – THE QUIET REVOLUTION

CHAIRMAN:

Dr. Stanley Winkler, Senior Scientist
IBM Systems Development Division
Gaithersburg, Maryland, U.S.A.

PROGRAM COORDINATOR:

Norman A. Heck
IBM Federal Systems Division
Gaithersburg, Maryland, U.S.A.

PAPERS:

Dr. Stanley Winkler, *"Computer Communication – The Quiet Revolution"*

Dr. Carl Hammer, Director, Computer Sciences, UNIVAC, Washington, Washington, D.C., U.S.A., *"Computer Communications: The Future"*

Gordon B. Thompson, Manager, Communications Studies, Bell-Northern Research, Ottawa, Ontario, Canada, *"Three Characterizations of Communications Revolutions"*

SPEAKER:

Kaoru Ando, Managing Director, Fujitsu Ltd., Tokyo, Japan, *"Social and Economic Implications of Computers and Communications"*

DISCUSSANTS:

Professor K. Venkitaraman, Chairman, Electronics and Telecommunications Engineering Division, The Institution of Engineers (India), Sivaram Nivas, Kerala, India

Professor F. Reza, Ambassador, Permanent Delegate of Iran to Unesco, Paris, France

THEME: A non-technical exploration of the technological innovations which are changing the nature of the world and the environment in which we live.

Perhaps it is too early to assert that Computer Communication is the innovation which will be comparable in its effect on the next century to the effect of the Gibbsian innovation on the last century. But the assertion can be defended cogently. Note that we use the words "Computer Communication" not "Computers and Communications". This is not just a semantic preference, but a conscious choice intended to indicate that it is not the appending of communications onto a computer, but rather the intimate entwining, amalgamation, if you will, of computers and communication into Computer Communication. The essential characteristic of this entwining is the pushing out in space over communication links of computing functions. Distributed Computing is a more accurate and descriptive term for this phenomenon.

Distributed Computing or Computer Communication has already burst upon us. In its simplist aspect it takes the form of remote job entry. A deck of cards is inserted in a reader at a remote station, transmitted to a distant computer, processed and returned to be printed out at the remote station. This scheme has the great advantage of being relatively inexpensive yet effective, and is popular wherever the elapsed time between submission and completion of a job is not critical. The remote job entry device is in actuality a small computer. Many of them are programmable and can perform computing functions at the remote location. On the next level are interactive terminal systems. The terminals range from simple teletypes to complex CRT devices.

The accoustic coupler, which enables every individual telephone to serve as a communication link, has provided the impetus for development of light weight, inexpensive, portable terminals. The extension of service from across the street to around the world by satellite, is a fascinating engineering feat, but conceptually trivial.

The expansion of Distributed Computing appears to be in three directions. The first is the intelligent terminal. Developments in terminal devices are occuring and the projected performance for such terminals is little short of astounding. For example, such terminals can be instructed (programmed) to accept and record or reject messages; and then evaluate and store for later delivery; or forward the recorded messages. The guarding of entrances where the terminal verifies the identity of an entry seeker and sends a record of admittance to a central file is feasible. Composing and editing with a light pen on a cathode-ray tube display unit is sufficiently familiar to be commonplace. The limiting factor in deciding whether or not to implement an application is no longer technological capability but economic feasibility.

The second direction is the network. If a terminal which is also a small programmable computer can be connected to a larger control computing system, there would be no reason not to be able to interconnect computers of any size. To say that the idea is simple in concept is not to minimize the difficulties of connection, control and programming. These difficulties are being solved in both obvious and ingenious ways. Again cost is very much a factor. The possibilities for application seem limitless. All at once, every computer, anywhere in the world, is potentially accessible to anyone with a terminal and a telephone. One can think of a library system where a user can, from his own location, tap the resources of all the libraries in a city, a country, or even a continent. Or, equivalently, an attorney can reach the legal records of any jurisdiction. A Wired City or a Computer Utility is no longer wild speculation.

The third direction in which Distributed Computing is spreading is in the area of sensory devices with inherent computing functions. The concepts of process control are being adapted to new fields. Sensor-based systems are extensively used in factories, petroleum refineriés and chemical plants. Similar systems are proposed which will control traffic in the large city, which will supervise and insure the safety of air travel, and which will monitor and protect the environment.

No single example or group of examples can convey the feeling of vast, revolutionary change. It is the all-encompassing pervasiveness of computer communication that is producing an effect much greater than the sum of the individual effects of which it is composed. Quietly, almost imperceptively, our way of life and its quality is changing. These changes will include the kind of work we do; the way in which we will do it; the way in which we will travel; the kind of homes in which we will live; how we will spend our leisure; how our children will be educated or trained; and all the other aspects of life as we know it. It is the totality of these changes which constitute the quiet revolution.

"Omnia mutantur, nos et mutantur in illis". (All things change, and we change with them.)

- Matthias Borbonius

COMPUTER COMMUNICATIONS: THE FUTURE

Carl Hammer, Ph.D.
UNIVAC
Washington, D.C., U.S.A.

Summary

Trends in electronic technology are best understood by a critical examination of the past, a statement about the present, and an extrapolation of our experiences into the future. This approach leads us from the ENIAC of 1945 and the UNIVAC of 1955 to large machines of 1965 and the colossal systems of the seventies. Concurrently, applications grew in complexity and scope, ranging from simple mathematical computations and data processing to simulation, war gaming, and real-time systems. Meanwhile...the cost of hardware has steadily decreased in terms of cost per unit operation while the speeds of system components have increased by many orders of magnitude. However, the most remarkable gains have been rolled up in the efficient use of these machines by development of operational systems, including remote terminals, communications, and special-purpose input-output devices.

Electronic systems of the future, therefore, will likely be larger, faster, and still cheaper to operate. They will also be more reliable than today's machines, furnishing uninterrupted service to multitudes of users with time-sharing and real-time capability. In short, they will be designed to provide computing power, similar to electric power, resembling closely the utility concept of the latter. But the most spectacular advance is likely to be their ability to do the job where it needs doing, reversing completely the logistics of yesteryear and today. Thus, we will be able to cope with the dense data trails created by man and his complex society, effecting a transition from an automated to a cybernetic society. Shortly, teleprocessing systems will take on most of the tasks of rote and drudgery which nature and society now impose upon us. The outlook is thus very bright as we learn how to make intelligent use of our not-always-so-intelligent and often maligned machines. As Norbert Wiener put it so aptly some thirty years ago, future systems will enable us to "make human use of human beings".

1. The Age of Teleprocessing

When Christopher Columbus set foot on the New World in 1492 he did not know where he was nor how he got there. It took the world years to learn of his courageous journey and decades to readjust. When Neil Armstrong stepped upon the surface of the moon in 1969, he knew exactly where he was, having followed a precisely preplanned journey, and the whole world was watching him! In less than five hundred years man's ability to communicate had advanced from its most primitive forms to an astonishing level of sophistication.

Early telegraph links, in Napoleon's time, had signal speeds of about two characters per second. The ability to combine up to six communications channels in one physical link came into being in 1874 with a scheme invented by Jean Maurice Emile Baudot. Two years later, Bell spoke his first sentence over the telephone. In 1913 vacuum tube repeaters were introduced in telephony. Other developments in electronics followed quickly: by 1918 the first carrier system permitted several voice channels to occupy a single pair of wires. High capacity coaxial cables started to replace wire-pair cables in the early 1940's; today they can carry thousands of telephone channels. Microwave radio links were first installed in 1946; they can now accommodate more than 10,000 telephone channels and even more in the future. The 1960's brought us satellites and high-speed waveguides; the 1970's may bring the lasers. In a hundred years, the capacity of our communications links has risen from fifteen to a billion baud, from two to over a hundred million characters per second.

As civilized man evolved from his primitive ancestry, he developed a voracious appetite for large masses of data, recording observations about his individual or collective activities with ever greater precision and detail. He thus created an historical data trail whose density continually grows with his needs. Concurrently, he also developed systems and machines to process this data avalanche more rapidly and efficiently. His demands for computing power soon exceeded by far the best available electro-mechanical calculators of the 'thirties. The concept of machines capable of storing both data and computing "recipes" internally, born in John von Neumann's mind, started us on the road toward developing electronic data processing machines with immense power. The very first such device, the ENIAC of 1946, gave man at once a leverage factor of ten-thousand and today's super-computers provide us with a mind-amplifying advantage of a billion-to-one.

The progress we have made, spectacular as it may seem, is the result of planned and continuous investments in research and development. Advances in technology are accomplished by an incremental and inductive process which has no deductive basis. Thus scientists and engineers plan certain experiments and design the required equipments, only to learn upon implementation that they did not foresee all possible contingencies. They improve their experiments or equipment designs in a second round, incorporating the experience gained from the first. This incremental process of invention and innovation has a well-defined gestation period and does not allow for leap-frogging. Advertisements in the trade press, proclaiming that "tomorrow's terminal is here today", are typical of the callous way in which marketeers attempt to "snow" their customer. The fact is that tomorrow's technology is founded upon that of today!

2. Technological Considerations in Electronic Systems

We have made significant advances in information processing and data transfer as a result of dynamic developments in computer and communication technologies. Computing centers are being harnessed to remote user terminals and connected to other centers by communication channels. In so doing, computing power is being made economically accessible to a multiplicity of users who no longer require a computer installation of their own in order to have access to the most modern information and data handling systems.

This developing relationship between computer and communications technology is possibly the most important event of our times. By themselves, the communications or the computer industry alone were capable of bringing about changes in business, government, and in our whole way of life. But the two technologies complement each other; in combination their power is more than additive! Computers control our immense communication switching centers and assist in managing the enormous capacity of new transmission links into useable channels; communications in return, make available the power of computers and the information in data banks to millions of users in remote locations.

This computer-communication marriage is all the more important as we learn to use third generation machines more efficiently and begin to introduce fourth generation machines. During the 1970's telecommunications revenue from machines conversing with machines will surpass that of people talking to people. This is an astounding development considering the fact that the world's largest corporation, A T & T, grew to its present size serving only human communication!

2.1 Basic System Ingredients

Three major components of science and technology have contributed significantly to the development of our new teleprocessing systems:

(i) Hardware. Historically, of course, we started with computers; we still differentiate the roles played by central and terminal electronic facilities. In data transmission systems we recognize two additional, important hardware items: Modems and Communication Links.

(ii) Software. Software now plays an entirely new role. It furnishes high-level management for the central facility hardware complex, lower-level management for the terminal facilities, and provides communications between these two management levels, assigning priorities and allocating communications resources to multitudes of users sharing them.

(iii) Brainware. Design and development of operational hardware and managing software has drawn heavily upon a host of sciences: Operations Research, queueing theory, mathematical statistics and probability theory, linear and dynamic programming, even management and social sciences. With their help we are designing the architectures of improved operating systems with stress on scientific management of resources, security, and error detection/correction in computer and data transmission subsystems.

Typical teleprocessing systems could not have come into their own without modulator-demodulator subsystems, or modems, which superimpose digital signal information upon analog carrier waves. Computer output circuits usually produce rectangular pulses and modems enable these pulses to travel along communication lines by converting them into optimal analog forms. Sophisticated modulation schemes are used to multiplex several signals together for joint transmission over the same channel. Their hardware implementations differ in cost, speed, reliability, and tolerance to noise and distortion.

Communication lines in terrestrial data transmission systems are furnished predominantly by the common carriers. A signal traveling along a leased line can follow a variety of paths to its destination. When it disappears into the socket in the plaster wall we often don't even know which way it will travel. It may go by itself on a wire circuit, or it may be huddled with hundreds of other signals on a coaxial cable or a microwave link. It may even race 22,300 miles into space to be beamed back by a satellite.

The largest common carrier in the U.S. is, of course, the American Telephone and Telegraph Company; its subsidiaries and associated companies operate more than eighty percent of the telephones installed in the United States. They make available a variety of services, means, and media to their customers. The most prominent trade-off factors usually considered are cost and speed; pertinent tariffs are complex and difficult to understand. Perhaps the two most important points to keep in mind are, (i) that the present lines were almost exclusively designed for voice communication, and (ii) that by using pulse code modulation, voice and other analog signals can be converted into a stream of bits looking remarkably like computer data. Therefore, it appears certain that future circuits will be designed to transmit very high-speed pulse trains carrying voice, television, facsimile, and other data in a uniformly coded form.

2.2 Real-Time and Time-Sharing Systems

The most promising, synergistic offspring resulting from the union of computer and communications systems is the development of real-time systems. We need not quibble here about the exact definition of this term, if indeed one can be found. But we can point out three characteristic functions of real-time systems. First, they are communications-oriented, making extensive (and hopefully efficient) use of data transmission networks. Second, they connect interactive remote terminals or sensors to computers, facilitating a meaningful dialogue between men and machines. This dialogue must proceed at reasonable speeds; it must also remain uninterrupted and terminal responses must be timely so that human participants can continue their train of thought without frustration. Third, to get the job done, the terminals must have interactive components, such as keyboards, displays, lightpens, joysticks and sketchpads.

Design and implementation of real-time systems are more formidable tasks than were faced in the earlier era of building electronic batch-processing systems, for at least four reasons: (i) Real-time systems tend to be larger in magnitude and scope than their batch-processing ancestors. They are usually developed for very complex applications, thereby compounding engineering and human problems. (ii) Real-time systems are difficult to install and their complete checkout is a never-ending task. Hardware and software must complement each other while polling user terminals or sensors, responding to demands for facilities or software systems, and maintaining managerial control over the entire system. (iii) Real-time systems concepts are still relatively new. Frequently they are built for applications which batch-processing could not handle. (iv) Real-time systems are still subject to catastrophic disruptions as neither hardware nor software have reached the state of perfection attained by the more mature batch-processing systems.

Crucial design problems have arisen in many areas. In the case of remote terminals, the bandwidth of the intermediate communications link is the limiting factor for effective data transmissions. Therefore, most recent designs aim at providing some "intelligence" at the terminal site, thus reducing both communications requirements and workload at the central facility. The problem how "smart" a terminal ought to be to make the overall system optimal is far from being solved although the amount of "smarts" available at the remote stations has increased continually as hardware costs are declining and reliability rises.

Another design problem arises in the communications subsystem. For example, if we link satellite data transmissions to terrestrial systems mixing of the media may lead to almost unresolvable conflicts. The design of a terrestrial transmission system using microwaves will have certain optimal design parameters for block data size, error detecting/correcting procedures, acknowledgment of transmitted data etc. A totally different set of parameters will obtain for coaxial cables laid on the ocean floor, or for high frequency systems in space. Linking two or more such systems together invariably causes us to alter individually optimal procedures or parameters, making them at best suboptimally optimal, so to speak.

2.3 Engineering Principles

Over the short history of designing and building teleprocessing systems, a number of fundamental principles have emerged. Engineering design considerations considered most important and most subject to change as the state-of-the-art improves include:

(i) Hardware Reliability. This aspect has probably received the greatest attention because it is so obvious.

(ii) Redundancy. Where anticipated component hardware failures might exceed permissible limits (as in real-time systems, for example), compensatory redundancy on all systems levels is employed. Since no man-made device is perfect and has unlimited life, statistical principles are employed to "guarantee" minimal performance over long periods of time, of appropriately designed redundant systems.

(iii) System Availability. Improved system reliability leads directly to greater availability of system resources. The most important innovation in this area is the soft-fail concept. If a system component deteriorates, monitoring circuitry is now available to detect system degradation and to develop appropriate messages for human observers and as inputs to other electronic components.

(iv) Error Detection and Correction. What hardware redundance does in the macrocosm of equipments and devices is accomplished by very sophisticated operational techniques in the microcosm of redundant data communications. Minimally, this might include "parity bits" to complement every transmitted set of data bits. This procedure permits the detection of only the simplest errors but it makes no provision for locating or correcting errors. More complicated coding schemes add two or more redundancy bits to each binary data word. Thus we can design arbitrarily powerful codes with whose help we can detect and/or correct any number of possible errors. But we observe that any increase in error detecting/correcting capability raises the data transmission overhead, defined as the extra number of bits which are added during transmission to each basic raw data word.

3. The Outlook

The rapidly increasing rate of change in all of society's activities forces us to look more penetratingly into the future than ever before. Our style of decision-making is changing as social scientists and environmental specialists begin to exploit recent advances in technology. We can no longer plan solely for short-term profit, ignoring the adverse side-effects of the means for obtaining it. Growing public awareness of the effects of technology upon society and the accompanying public and political debate indicate that the long-term must become our standard planning horizon. Industry and government alike must be made aware of complementing and competitive technologies; they must learn to understand relevant political, social and economic impacts and be able to plan for contingencies.

Our industrial and political leadership is becoming aware of the need to reduce uncertainty in short-term and long-range planning. So many inter-related factors must be evaluated that decision-making is becoming a far more complicated task than ever before. This is clearly reflected by the acutely rising consciousness of accountability in setting policy and goals, both nationally and internationally, and by attempts to establish a balance between short-term existence and long term-growth continuity.

New techniques have been developed to aid our over-burdened decision-makers, to provide them with quantitative information about the implications of decision alternatives. One such new tool is technological forecasting. It allows us to incorporate the scientific method into formalized management systems for long-range planning and resource allocation.

3.1 Technological Forecasting

This powerful management tool can provide management with indications of the growth of technology as measured against estimated social and economic needs. Technological change has become a staple characteristic of modern life, leading to a need for management to predict, with a level of confidence, possible and probable technical achievements in a given time frame, assuming a specified level of support. Technological forecasting still remains an art, rather than a science, as its credibility is heavily dependent upon the people conducting the forecasts, the method of obtaining relevant data, the manner of performing the analysis, and the format of presentation. Of the five available methodologies (Intuitive, Trend Extrapolation, Trend Correlation, Analogy, and Normative), the intuitive DELPHI technique has possibly received the greatest acclaim.

Projected advances in technology will dramatically impact computer communications in these three areas:

(i) Hardware. Reliability, maintainability and cost-performance ratios will continue to improve with a great increase in component density and attendant reductions in equipment size, weight, and power consumption. Intelligent terminals will make their appearance both in space and on the ground, greatly reducing the load on communications bandwidth requirements. Sensors and cameras with increased resolution will allow us to gather better and more timely data, and to transmit and process them in real-time.

(ii) Software. The most significant trend will be the gradual elimination of systems software through replacement by mini-computer modules, thus reducing significantly design and development costs of future teleprocessing systems. Further standardization of applications software will increase transferability of modular programs between computers and terminals. Micro-programmed machines with more flexible, high-level languages will greatly reduce the time and cost required now in software development efforts.

(iii) Brainware. Significant advances in modulation theory, error detecting/correcting codes, hardware systems architecture, and other areas impacting future data communications systems are anticipated only if relevant basic and applied research receive appropriate support from government and industry. A severe cutback in funding such projects is more than likely to result in an erosion of our technological leadership with predictably dire consequences in terms of our posture in the world scientific community.

A typical example of technological forecasting can be taken from the area of terminal devices and displays. Most equipments in this category use conventional cathode ray tubes for viewing by human operators. Very much like the television sets with which we are familiar, display devices are bulky and require a good deal of power. These devices would be much more effective if today's cathode ray tubes could be replaced by flat screens which people could stand up on their desks, hang on the wall, or even mount on ceilings. Technological forecasters predicted in 1967 that before 1975 we would develop such flat screens. It is remarkable that the tradepress very recently carried a news item according to which the first experimental flat screens are being tested, with mass production now envisioned for 1974! What impact this event will have on the television and picture-phone market can hardly be fathomed. But it is a fact that research and development in this field were adequately funded, with the result that in the eighties few people will remember what TV looked like today.

3.2 Challenges of the Seventies

Networks consisting of satellites, surface-probing sensors, and remote terminals will be most useful in the study of our natural resources. With their help, we can take inventory of the oceans and the air; of what, where and how well forests and crops are growing; and of the condition of the soil and its potential productivity. Thus we can develop regional, national, or global predictions of crop yields and livestock inventories, or patterns of fire, insect, and disease damage. Stream and river flows, excess surface water, pollution, and glacial action can be monitored to help us plan better irrigation and flood control systems, or to develop and maintain water resources and to contain erosion.

In order to manage world resources more effectively, adequate data must be gathered and converted into relevant information. For centuries, man has collected such data on the surface of the earth only. More recently, aerial observations and satellite surveys have broadened our field of view, exponentially increasing the amount of data collected and the value of the derived information. With the introduction of mass acquisition of data and sophisticated computer processing, we may well be able to stem the tide of our diminishing natural wealth and of the pollution of existing resources.

The introduction of electronic data communication systems into our society has already caused profound changes in its structure and organization. Real-time systems and time-sharing make the power of the computer available to untold thousands at their desks and even in their homes. A steadily growing number of commercial users rely on global and spatial electronic communication systems, as much as scientists employ them to obtain, store, and disseminate large quantities of data and pictures. And yet, this is only the beginning; the full impact of electronic systems upon human society will take at least another decade before we can properly assess it.

We must understand that we are in the midst of a transition from an automated to a cybernetic society. By the end of this decade, electronic systems and especially communications will affect practically every aspect of human behavior. Information utilities and real-time systems will be available to the public in the same sense that other utilities today service our homes and offices. Linked global and spatial communication systems will serve government officials, businessmen, scientists, students, even our children, furnishing them the ability to "converse" with computers as readily as they now talk by telephone.

As we perfect intelligent remote terminals, the intensity of the data trail thus created will undoubtedly increase beyond all imagination. Relevant information will become available as inputs to sophisticated simulation models whose outputs will aid our decision-makers in the search for viable alternatives and optimal solutions to many man-made problems, some of which may affect indirectly all of mankind. Our efforts to build models for testing of "global" decisions, as in the "Club of Rome", can only be successful if international model-makers have access to an appropriate data base through teleprocessing. These efforts must succeed if man is to survive -- they will succeed if we have the ability to communicate. With apologies to Norbert Wiener, we conclude that homo sapiens must learn to make intelligent use of our not-yet-very-intelligent machines -- before his time runs out!

4. Selected References

R. U. Ayres, Technological Forecasting and Long-Range Planning, McGraw-Hill, 1969.

G. B. Bernstein and L. A. Feidelman, Data Processing Forecast 80, Forecasting International Ltd., Fairfax, Virginia, USA, 1971.

M. J. Cetron and J. D. Goldhar (Editors), The Science of Managing Organized Technology, Gordon and Breach, 1970.

P. J. Denning, Third Generation Computer Systems, ACM Computing Surveys, Vol. 3, No. 4, December 1971.

R. G. Dewitt, Nationwide Digital Transmission Network for Data, Telecommunications, September 1971.

I. Flores, Computer Organization, Prentice-Hall, 1969.

J. W. Forrester, Urban Dynamics, MIT Press, 1971

J. Martin, Future Developments in Telecommunications, Prentice-Hall, 1971.

SPECIAL PROJECT

Participating Demonstrations of a Multi-Purpose Network Linking Dissimilar Computers and Terminals

CHAIRMAN:

Dr. Larence G. Roberts, Director for Information Processing Techniques,
Advanced Research Projects Agency
Department of Defense, Arlington, Virginia, U.S.A.

PROJECT COORDINATOR:

Dr. Robert E. Kahn*
Bolt, Beranek and Newman
Cambridge, Massachusetts, U.S.A.

PAPER:

"Participating Demonstrations of a Multi-purpose Network Linking Dissimilar Computers and Terminals"

EXAMPLE DEMONSTRATIONS:

"Multi-Computer Support of Cross-Country Interactive Graphics", Bolt, Bernaek and Newman, Cambridge, Massachusetts, U.S.A. Institute of Technology, Cambridge, Massachusetts, U.S.A., The MITRE Corporation, McLean, Virginia, U.S.A.; and University of California, Santa Barbara, Goleta, California, U.S.A.

"An Aid for Collaboration for Geographically Distributed Groups", Stanford Research Institute, Menlo Park, California, U.S.A.

"MACSYMA–An Interactive Formal Mathematics System", Massachusetts Institute of Technology Cambridge, Massachusetts, U.S.A.

THEME: Conference attendees will exercise a variety of computers and terminals in a range of applications. The computers are widespread geographically and under diverse managements.

---

*Now at Advanced Research Projects Agency, Arlington, Virginia, U.S.A.

PARTICIPATING DEMONSTRATIONS OF A MULTI-PURPOSE NETWORK LINKING DISSIMILAR COMPUTERS AND TERMINALS

The Special Project of the International Conference on Computer Communication (ICCC-72) is a demonstration of a computer network that has successfully melded computers and a telecommunications network to achieve resource sharing. To illustrate this achievement demonstrations of data retrieval, combined use of several machines, real time data access, interactive cooperation, and simulation systems via computer communication networks are shown at ICCC-72. In particular, an opportunity for conference attendees to gain first hand experience with the use of a computer network has been provided.

The Special Project Committee of ICCC-72 has arranged for access to the ARPA Computer Network (ARPNET) from the Georgetown Ballroom of the Washington Hilton Hotel for the duration of the conference. The ARPANET is an advanced technology store-and-forward message switching network of more than 25 computers.

A map of the ARPANET showing the geographic distribution of network nodes is shown in Fig. 1. All links between network nodes are at least 50 kilobits/second. Connection to the ARPANET for the purposes of the Special Project of ICCC-72 was accomplished by creating a temporary node in the network, at the Washington Hilton, between the TINKER and ETAC nodes. The temporary node at the Washington Hilton comprises:

1. A Terminal Interface Message Processor (TIP) that provides both a store-and-forward message switching function in the ARPANET and terminal access to the network.

2. Connection of the TIP to the ARPANET via two 50 kilobit/second wideband telephone circuits.

3. A variety of thirty or so terminals all connected to the network through the TIP. Terminal types connected to the TIP for the Special Project include: impact printers with keyboard, thermal printers with keyboard, CRT displays with keyboard, impact and impactless line printers, etc.

The Special Project highlights the beginning of an era that will allow a user to command execution of a program that exists on a computer installation that is geographically distant from him; in turn, the program may accept or demand input data command execution of a program, and/or use specialized hardware at any installation connected to the network. It also demonstrates that a computer network can eliminate the need to implement programs again and again for each installation or the need to maintain copies of identical data-sets at each installation they are used.

Many organizations that have host computers connected to the ARPANET or obtain computational capability via the ARPANET are providing demonstrations. While all the demonstrations cannot be described here, demonstrations that are a representative cross section are:

A demonstration of the On-Line and Logo Systems. This is particularly interesting because the Intercomputer Network Development Group at the Mitre Corporation, McLean, Virginia, uses, through the ARPANET, a facility that exists at M.I.T.'s Project MAC for inputting, editing, assembling, and debugging programs for an intelligent terminal possessing a minicomputer. The network is also used to transmit a loadable version of the program to the minicomputer. To facilitate reloading, the initial load via the Network is copied onto tape using a cassette style tape recorder. The programs thus far implemented for the minicomputer are programs that specifically tailor the intelligent terminal to interface to some of the other computers on the Network, notably the On-Line System at the University of California at Santa Barbara, existing on a PDD-10 Tenex system, at Bolt Beranek and Newman, Cambridge, Massachusetts.

A demonstration of the Network Information Center (NIC) designed and developed by the Augmentation Research Center at Stanford Research Institute at Menlo Park, California. The NIC provides a general information service to the ARPANET Community. The services offered by the NIC have as their basic objectives: to help people with problems find resources --people, systems, and information -- available within the network community that meet their needs; to help members of a geographically distributed group collaborate with each other. In fact, the service that aids collaboration was used for planning the Special Project. In addition to a demonstration, conference attendees are provided an opportunity to use some of the NIC facilities.

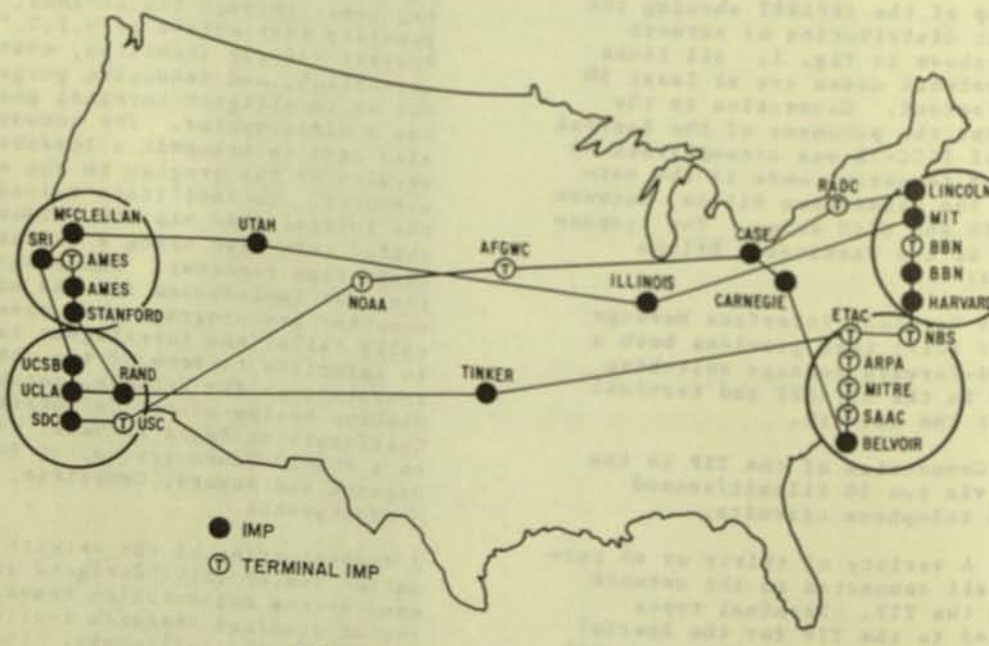
A demonstration of MACSYMA, Project MAC's SYmbolic MAnipulation system developed by the Mathlab Group at M.I.T.'s Project MAC in Cambridge, Massachusetts. MACSYMA exists on a PDP-10 at M.I.T.. It provides a capability for symbolically manipulating algebraic expressions involving constants, variables, and functions. The system provides capability to differentiate, integrate, take limits, solve equations, factor polynomials, expand functions in power series, plot curves, etc.. In addition to a demonstration of MACSYMA, attendees are provided with an opportunity to use it.

A demonstration of a Multi Computer Route Oriented Simulation System called MCROSS, developed by the Computer Science Division, at Bolt Beranek and Newman of Cambridge, Massachusetts. Some of the unique aspects of the system are: that several computers on the network can cooperate to simulate air traffic in adjacent airspaces and the network is used to hand-off aircraft information as aircrafts cross airspace boundaries; the aircraft situation display of the simulated airspaces can be viewed at any network node; the parts of the simulation can move from one computer to another throughout the course of a simulation without disrupting the simulation.

While it is evident that the major benefits of computer networks are yet to be fully realized, one benefit, resource sharing, is one of the major goals of the ARPANET. To achieve resource sharing, the ARPANET has provided a facile means by which computer resources that are geographically distributed are shared by users who themselves are geographically distributed. To illustrate this, capabilities and unique resources -- in the form of programs, data sets and/or specialized hardware --of many of the computer systems on the network are demonstrated.

Acknowledgements:

The ICCC-72 Special Project Committee would like to acknowledge: All the terminal equipment manufacturers who graciously contributed the terminal equipment for the duration of the conference; the cooperation of the American Society for Information Sciences (ASIS) for their help in making the Special Project successful; the cooperation of ATT Long Lines for their aid in creating the temporary ARPANET node; the Advanced Research Projects Agency of the Department of Defense for coordinating the Special Project and providing a computer network; and finally, the more than 50 people who were actively involved in making the Special Project happen.



ARPA NETWORK, GEOGRAPHIC MAP, AUGUST 1972

Figure - 1

RESPONSIBILITY, IMPACTS AND SOCIAL CONCERNS

[illegible]

[illegible] A demonstration of [illegible] system developed by the [illegible] Group at [illegible], a project [illegible] in Cambridge, Massachusetts. [illegible] It provides a capability for [illegible]. The system provides capability [illegible]. In addition to a demonstration of [illegible], attendees are provided with an opportunity to use it.

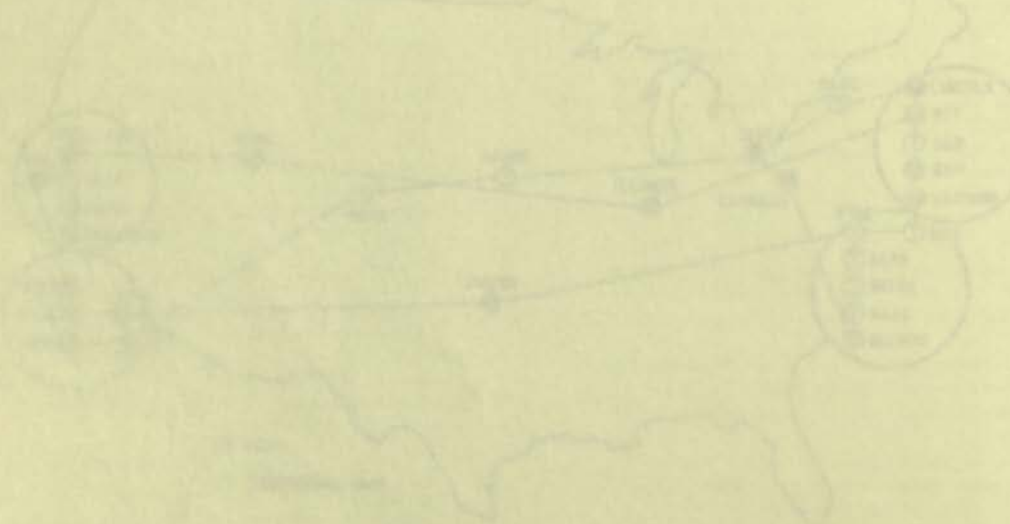
A demonstration of a Multi Computer [illegible]

[illegible] network under the guise of the time-sharing [illegible] from the computer as [illegible] and aware of [illegible] without disrupting the simulation.

While it is evident that the major benefits of computer networks are yet to be fully realized, one benefit, resource sharing, is one of the major goals of the ARPANET. To achieve resource sharing, the ARPANET has created a facility by which computer resources that are geographically distributed are shared by users who themselves are geographically distributed. To illustrate this, capabilities and unique resources -- in the form of programs, data base and/or specialized hardware -- of many of the computer systems on the network are demonstrated.

Acknowledgements:

The ICCC-72 Special Project Committee would like to acknowledge: all the terminal equipment manufacturers who graciously contributed the terminal equipment for the duration of the conference; [illegible] the American [illegible] Special Project [illegible] the cooperation of [illegible] for their aid in [illegible] the ARPANET under the [illegible] Advanced Research Projects Agency of the Department of Defense for [illegible] the Special Project and providing [illegible] network; and finally the [illegible] of people who were actively involved in making the Special Project happen.



[illegible]

RESPONSIBILITY

CHAIRMAN:

Dr. Herbert Maisel, Director, Computation Center and Associate Professor of
Computer Science
Georgetown University, Washington, D.C., U.S.A.

PROGRAM COORDINATOR:

Louis Feldner, Common Carrier Bureau
Federal Communications Commission
Washington, D.C., U.S.A.

PAPERS:

Dr. Herbert Maisel, *"Responsibility for the Humanistic Use of the Information Revolution: Where Will the Battle be Fought?"*

Commissioner Robert E. Lee, Federal Communications Commission, Washington, D.C., U.S.A., *"The Role of the Federal Communications Commission"*

Professor Victor C. Ferkiss, Department of Government, Georgetown University, Washington, D.C., U.S.A., *"Computers and Communications: Boon or Bane? Information Technology, Freedom and Social Control"*

SPEAKERS:

Professor Joseph Weizenbaum, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A., *"Responsibility of the Technologist as an Individual"*

Yoneji Masuda, Director, Japan Computer Usage Developement Institute, Tokyo, Japan, *"Responsibility and the Information Revolution: An Overseas Perspective"*

THEME: Everyone concedes the tremendous impact that computer communications will have in our society in the coming decades. Some say that this impact is already substantially upon us. A new technology with enormous impact necessarily modifies many of our day-to-day activities. It is important that this modification benefit mankind and especially that it promote, in Norbert Wiener's words, "The human use of human beings." This implies responsibility—the need to establish principles and guidelines and to channel these technological developments in accordance with these principles and guidelines.

RESPONSIBILITY FOR THE HUMANISTIC USE OF THE INFORMATION REVOLUTION: WHERE WILL THE BATTLE BE FOUGHT?

Herbert Maisel, Georgetown University, Washington, D.C., U.S.A.

The second industrial revolution is now in progress. The widespread use of digital computers, the enhancement of communication networks, the large capacity and high speed of auxiliary storage units, and the development of inexpensive terminals with television capabilities imply a revolution in information processing. The impact of this revolution is already upon us. However, its effects on the day-to-day activities of the layman are still marginal. These effects will increase substantially in the next two or three decades. If we are to make this revolution one that mankind will look back on joyfully then we must assure that these effects are as beneficial as possible. We, all of us, have a responsibility to do this. This broad commitment is the subject of this panel. Each of us on the panel will attempt to make this commitment more specific. My objective is to try to answer the question: where will the battle for the humanistic use of the information revolution take place?

A note of alarm will be sounded, quietly by some of us, more stridently by others. This is only natural in a call to action. There is always a need to shout in order to alert the passive. If we sometimes appear to be too pessimistic and too negative, remember that there is much to be done and too few people willing to do it.

Pioneers in information processing, especially Norbert Wiener, believed that the principle impact of the second industrial revolution would be widespread unemployment and that the assurance of the human use of human beings should be our major responsibility.[1] This objective is still an appropriate one. However, the issues have become much more complicated. Privacy and the threat to privacy posed both by corporations and government have become important considerations. Also, the development of those applications that can be of greatest public benefit is not assured. Finally, we must indeed see to it that this second industrial revolution results in a smooth transition from a work to a leisure oriented society.[2] It may be that this last transition, the one that so greatly concerned our pioneers, will be most easily accomplished. This could occur because the transition will be made over several decades so that a radical recasting of individual ethics may not be necessary. The issue then may not be acceptance of a leisure ethic but rather the nature of our increased leisure activities and the potential restriction of our freedoms. Our responsibility is to assure that all human beings benefit to the maximum extent possible from this information revolution and that our freedoms are not curtailed by it.

An interesting further problem that may arise as a result of the second industrial revolution is the dehumanization of our self-image. The first industrial revolution encouraged us to think of ourselves as machines. This self-view was neither correct nor beneficial. We may now come to view ourselves as merely information processors. Professor Weizenbaum has emphasized this potential impact of the computer.[3] He will discuss our responsibility both in avoiding this self-view and in promoting a more healthy, humanistic view.

It would be reasonable to expect that in the United States the principle arena in which the fight will be waged to maximize the potential benefits of the information revolution would be in the Federal Government; perhaps most especially in the Federal Communications Commission. Will the Commission regulate the communication networks and their information processing users in such a way that the public good is assured? It would appear that a clear cut "yes" cannot be given in answer to this question. First of all, the Commission has in the past separated the communication lines from the data processing services that these lines interconnect and seems to be taking the position that its regulatory interest ends at the terminus of the communication lines. Second, the Commission has always emphasized efficiency and adequate service at reasonable charges in its regulatory activities. Much less emphasis has been given to the impact of these communication facilities on the public. Commissioner Lee will say a great deal more about this, so I will merely remark that the arena in which the fight will be fought to assure maximum benefits from the information revolution may very well not be the Federal Communications Commission.

Where, then, will this battle be fought? It would appear that new legislation will be required to do such things as ensure privacy and develop effective educational applications. (The educational potential of the information revolution is discussed in [4].) The major arena may be the Congress of the United States. However, in recent years, the Office of the President has assumed a greater and more direct responsibility for the conduct of our national affairs. Although this has been especially pronounced in the areas of intelligence, defense, and foreign affairs, it has also been true in domestic activity. Our science and educational policies are not entirely the result of legislation. This means that another important arena is in the Office of the President. In order to insure responsible information systems policies in both the Congress and the Office of the President, most o us must rely on the political process and an informed public. Our major responsibilities lie probably in informing the public and their representatives.

The principle impact of the Federal Government on the information revolution may turn out to be independent of legislation and of the Office of the President. It may result from the totality of applications of information systems made and promoted by the various government agencies. Important steps in such diverse applications as the recording of scientific information[5], simulation[6], and education[7]* have been taken by government agencies. In this arena, responsibility lies less with the political process and the public's concern and more with the individual technologists and their concern; in other words, more with those of you who help develop applications in government agencies and who have the authority to review contracts and grants for such applicaions.

Traditionally, the development of new knowledge and the innovative application of existing knowledge has been one of the major activities of our educational institutions. This is true of the information revolution but probably to a lesser extent than might be expected. Although many applications have been and will be developed at universities, many more are the product of government agencies, non-profit corporations and profit-making organizations. Certainly academicians will have an impact and must keep in mind their public responsibilities. But the research activities of our educational institutions will probably not be a principle arena in the battle for the human use of computers.

The principle arena of this battle may be the board rooms, the intermediate management levels, and the applications programming staffs of our largest corporations. Unless the profit motive is supplemented by public concern, the information revolution may do little more than substantially increase both the volume of information flow and the extent to which executives can review and control the functions of their organization. The employees of that organization, the customers for that organization, and the public at large may find that their life is no better as a result of these changes. In fact, serious restrictions on the freedom of the market place, increased standardization of products to the point where their humanistic appeal is secondary, and the exchange of information about people without their knowledge may be increased to an alarming degree. This potential for increased consumption of information and increased centralization of authority that can result from the information revolution must be offset by a concern for the public good as well as the corporate good. Legislation will probably provide an inadequate answer. With appropriate legal and technical assistance, corporations with sufficient motivation have always managed to do very nearly much as they please.[8] We must awaken the social consciousness of our corporations. The chances for success in this task are related to the last arena for the battle that I will discuss, our educational institutions.

The educational process must emphasize the potential good and the potential harm that can result from the information revolution. (Professor Ferkiss will have a good deal more to say about both the good and the harm that he sees in this revolution.) This emphasis should result in the development of a socially responsible attitude. Educators especially those at the college level, have been very concerned with providing appropriate breadth in their programs. Perhaps an appreciation for humanistic values, the societal impact of technology and the importance of a beneficial marriage of technology and human needs are appropriate objectives for scientific technological and business oriented undergraduate programs. In meeting these objectives, the appropriate breadth may be assured. Of course, all this must be done without sacrificing the depth of these programs.

Education, however, is no panacea. We return again to the simple truth that no matter what our backgrounds or our jobs we, all of us, have a responsibility to assure that the effects of the information revolution are truly beneficial to mankind.

References

1. Wiener, N.: The Human Use of Human Beings, Houghton Mifflin Company, Boston, Massachusetts, 1950.

2. Hilton, A.M.: "An Ethos For the Age of Cyberculture," Proceedings, Spring Joint Computer Conference, pp. 239-253, 1964.

3. Weizenbaum, J.: "On the Impact of the Computer On Society," Science, Volume 176, pp. 609-614.

4. Parker, E B and Dunn, D A.: "Information Technology: Its Social Potential," Science, Volume 176, pp. 1392-1399.

5. Feiger, M.: "Information Services: A Guide to Federal Offerings," Science, Volume 153, pp. 722-724.

6. Maisel, H. and Gnugnoli, G.: Simulation of Discrete Stochastic Systems, Science Research Associates, Palo Alto, California, 1972, Chapters 13-15.

7. Hammond, A.L.: "Computer-Assisted Instruction", Science, Volume 176, pp. 1005-1006 and 1110-1112.

8. Green, M J. with Moore, B.C . Dr., and Wusserstein, B.: The Closed Enterprise System, Grossman Publishers Inc., New York, N.Y., 1972.

---

*Actually most of the advances in computer assisted instruction are being made at universities, however, government funding underlies most of these projects.

THE ROLE OF THE FEDERAL COMMUNICATIONS COMMISSION

Robert E. Lee

Federal Communications Commission

Washington, D.C., U.S.A.

We are here today to consider the subject of authority and responsibility, from various vantage points, with respect to the burgeoning computer communication industry. I will concentrate on one branch of the U.S. Government, the Federal Communications Commission, in its role of authority and responsibility to the industry and to the public.

As a regulatory agency, the FCC's overall responsibility and authority has been defined by Congress in the Communications Act of 1934, as amended. Section I of the Act states that:

> "For the purpose of regulating interstate and foreign commerce in communication by wire and radio so as to make available, so far as possible, to all the people of the United States a rapid, efficient, Nationwide, and world-wide wire and radio communication service with adequate facilities at reasonable charges, for the purpose of national defense, for the purpose of promoting safety of life and property through the use of wire and radio communication ... there is hereby created a commission to be known as the 'Federal Communications Commission,' which ... shall execute and enforce the provisions of this Act."

In short, the measure of how well the Commission carries out its authority and responsibilities must be calculated against a yardstick already specifically provided by the Congress, that yardstick being the Communications Act of 1934 and its subsequent amendments. Of course, there is a degree of latitude in interpretations of such an Act, but the framework within which the the Commission works is perhaps more fixed than may be the case for some other segments of our society addressing themselves to this topic.

As you doubtless are aware, one way in which the Commission carries out its responsibility, is to hold Public Inquiries. These Inquiries are specifically designed to elicit public comment on important issues. One pertinent example of this kind of action is the Notice of Inquiry entitled: <u>In the Matter of Regulatory and Policy Problems Presented by the Interdependence of Computer and Communication Services and Facilities</u>. Issued in November 1966, this Inquiry is also known more briefly as Docket No. 16979.

Through Docket No. 16979, the Commission elicited public comment concerning potential problems and issues related to the <u>interdependence</u> of computers and communications. The questions raised in the Inquiry, and the subsequent responses and replies, ranged from consideration of the current and future trends related to the computer and information processing industry; to the adequacy of existing legislation to deal with industry trends; to the need for any new measures "... required by the computer industry and common carriers to protect the privacy and proprietary nature of data stored in computers and transmitted over communication facilities..."

The Commission received a great number of responses from the communications industry, from the data processing industry, from commercial users of data systems, from government agencies, and other parties. In all, over 3,000 pages of material were submitted to the Commission in response to the initial Inquiry.

This initial 1966 Inquiry was followed by a Supplemental Notice of Inquiry in 1967. In 1969 a Report of Further Notice of Inquiry was issued, followed in 1970 by a Tentative Decision and Notice of Proposed Rule Making and an oral argument. The Final Decision and Order was adopted by the Commission in March 1971.

The Commission's Final Decision stated first, that "... we are not proposing, at this time, to regulate data processing, as such, ..." and second, because of the growing interdependence of the computer-communication industry, the need to insure "... appropriate regulatory treatment..." required a maximum separation of activities which are subject to regulation from non-regulated activities involving data processing.

The Final Decision also answered questions raised during the Inquiry regarding the Commission's jurisdictional authority. The Decision stated that the Commission's jurisdiction extended "... to all communication common carriers insofar as they are participants in the provision of interstate communications services."

In certain areas of the Inquiry the Commission took a wait-and-see attitude. In attempting to assess the need to regulate "hybrid service" the Commission decided it did not have, at that time, the Solomon-like wisdom needed in order to separate hybrid communications from hybrid data services. The Commission decided, instead, to treat hybrid service issues on a case-by-case basis until it had much better visibility into how such systems would develop.

Although I concurred in part with the Final Decision, there were certain portions of that Decision to which I dissented. In particular, I disagreed with the degree of separation and the rules set up to deny common carriers access to computer services from carrier data affiliates. Those of us who dissented felt that the Commission majority in this instance were imposing restrictions "... because of some vague and unsupported feeling that a common carrier would not be able to resist the temptation to engage in 'improprieties' in managing its relationships with a data affiliate." I do not want to dwell on this point here, since our main theme is "responsibility" but I did want to make it clear that some of us disagreed with this aspect of the Final Decision.

Using this brief overview of the Computer Inquiry as our reference point, what can we say as to the Commission's future responsibility vis-a-vis the growth of computer communication interdependence?

One lesson that emerged from the Inquiry was that the Commission henceforth would be concerned with the interdependence of two areas affecting our society: computer and communications. The Commission now has an increased responsibility to be aware of, and sensitive to, developments not only among the communications carriers, but also to developments in computer systems. Where these two areas intertwine and impact upon the nation's communication facilities, the FCC must be alert to assess this impact and take action in accord with the requirements of the Communications Act.

With this added responsibility has come another requirement: the Commission must establish wider and closer contact with industry, and with sectors of the government and public, which heretofore may not have had much contact with the FCC. The Inquiry highlighted this need by virtue of the number of non-communications-related organizations which participated in the Inquiry.

The Inquiry also makes clear a need for the Commission to be alert to technological and user developments in the data processing industry. Technologies such as data terminals and computers -- which just a decade ago might not have appeared to relate to the scope of the Commission's activities -- will now have to be assesed for their relationship to the nation's communication services and facilities.

Like many other sectors of government, the Commission finds that the use and spread of computer communication systems have added to its responsibilities. But I believe that while the Commission has the authority and responsibility to probe the issues and problems raised by the growth of these systems and these technologies, there is a concomitant responsibility on the part of the Commission to tread carefully when it seeks to extend the domain of regulation into the area of computer communications.

COMPUTERS AND COMMUNICATIONS: BOON OR BANE?
INFORMATION TECHNOLOGY, FREEDOM AND SOCIAL CONTROL

Victor C. Ferkiss

Georgetown University

Washington, D.C., U.S.A.

Throughout history the power of men over their fellow men has been exercised through the differential possession of or access to scarce resources: physical strength, land, capital or knowledge primarily. Even if one rejects some of the more simplistic prophecies about "post-industrial" society, we are undoubtedly entering into an era in which knowledge of certain kinds is such a scarce resource upon which changing structures of social power are arising. One area of knowledge, control over which is increasingly vital, is that generated by electronic information storage, retrieval and computation systems.

The computer defined in its broadest sense creates new forms of scarce, power-generating information. It permits some to make calculations and acquire knowledge which can be denied to others, especially within certain time frameworks. It permits some to store and retrieve specific information not so easily accessible to others, if at all. It decreases the ability of some to protect information - such as personal histories - and makes that information more accessible to others for purposes of surveillance and control. It makes it possible through complex and subtle planning mechanisms to influence the behavior of persons without their being aware of and/or able to devise means of counteracting such influences.

Such new computer-generated power has the potential for radically altering the balance of power in society as a whole. As long as some have access to certain information systems and others do not, the former have greater social power just as if some had access to printed material, while others though literate did not. As long as some know how to use computers for their ends and others do not, the latter are at a disadvantage, as if some were literate and others were not. Those groups or institutions, such as government, which have the ability to command the services of those able to use the new technology and to control access to it have power advantages over groups and institutions which do not. At the present stage of computer development and use, special advantages therefore accrue to more powerful and/or wealthy groups and governments vis a vis others, the people as a whole or individual citizens as such.

Is this tendency of information technology to consolidate power inherent in the technology and thereforce possibly inevitable, or is it a matter itself of social choice? It can be argued that it is primarily the latter. If society chooses to do so, ability to deal with computers can be made as widespread as literacy. Access to and control of information systems can be made more widespread, both as regards input and output. Despite technical problems, privacy of computer-stored data can be maintained as much as that of any other recorded data, while systems can be devised to make information about the activities of governments and private groups more accessible to average individuals than under current conditions, especially as simpler means of querying computers are developed.

The choice as to whether the development of information technology will lead to a greater concentration or dispersal of power in society is itself essentially a social choice. Professionals in the field can greatly influence this choice one way or the other depending on their own attitudes and activities.

IMPACTS

CHAIRMAN:

Andrew J. Lipinski
Institute for the Future
Menlo Park, California, U.S.A.

PROGRAM COORDINATOR:

Gordon B. Thompson, Manager,
Communications Studies
Bell-Northern Research
Ottawa, Ontario, Canada

PAPERS:

Andrew J. Lipinski, *"The Impact of Wide-Spread Computer Communication Systems and their Use"*

Barrington Nevitt, Director, Innovations and Training, Ontario Development Corporation, Toronto, Ontario, Canada, *"Computer at Wits End Leads to Process Pattern Recognition"*

Professor D.A. Dunn, Engineering-Economic Systems Department, Stanford University, Stanford, California, U.S.A., *"Alternative Future Computer-Communications Markets"*

Professor David W. Conrath, Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada, *"Measuring the Computer's Impact on Organizational Structure"*

N. D. Hill and T. F. Watling, International Computers Limited, London, England, *"The Impact of LACES"* (London Airport Cargo EDP Scheme)

Dr. Thomas L. McPhail*, Chief, Social Environment Planning Unit, Department of Communications, Government of Canada, Ottawa, Ontario, Canada, *"How the Public Perceives the Computer: Some Social-Psychological Dimensions"*

Dr. Shigeru Watanabe, The University of Tokyo, Hoichi Itaya, Masayuki Shimada and Kazuo Yagi, Japan Society for the Promotion of Machinery, Industry, Tokyo, Japan, *"Computer-Assisted Instruction—A System and its Assessment in Japan"*

Dr. Hans J. von Baeyer, Director General, Computer-Communications Task Force, Ottawa, Ontario, Canada, *"Impact of Computer Communications as Tools in the Social and Economic Development of Canada"* (abstract only)

SPEAKER:

Kotaro Shimo, Ministry of International Trade and Industry, Tokyo, Japan, *"Computer Communication Assisted Community Project in Japan".* (presented by Toshio Kitamura, Japan Trade Center, New York, New York, U.S.A.)

DISCUSSANTS:

Professor J. C. R. Licklider, Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.

Professor Nathaniel Macon, Department of Mathematics and Computer Science, American University, Washington, D.C., U.S.A.

Raymond G. Fox, IBM Federal Systems Division, Gaithersburg, Maryland, U.S.A.

R. John Lansdown, Computer Arts Society, British Computer Society Specialist Group, London, England

THEME: The Impact of Widespread Computer Communication Systems and Their Use.

---

*Now Consultant, Computer Systems Division, The MITRE Corporation, McLean, Virginia, U.S.A.

THE IMPACT OF WIDE-SPREAD COMPUTER COMMUNICATION SYSTEMS AND THEIR USE

Andrew J. Lipinski
Institute for the Future, Menlo Park, California, U.S.A.

The impact of computer communication systems on most segments of our society to date has been slight because:

(1) data communications are in an early stage of development,

(2) the variety of system approaches reflects the typical early experimentation period of immature technology,

(3) fast-changing technology has to accommodate itself to less rapid institutional changes, and

(4) the "human" component of the system has received, thus far, little attention.

The papers of the two sessions, Impact I and II, illustrate some of the problems, solutions, and aspirations of the creators of an essential element of the postindustrial society--computer communication systems. The Present period corresponds to the "sweat and grime" period of early industrial society. The impacts of computer communication systems, both beneficial and injurious, are as yet only dimly perceived, as the effects of an industrial society were equally dimly perceived in Europe during the latter part of the last century. These two sessions are intended to make a small contribution toward improving our perception of the future.

COMPUTER AT WITS END LEADS TO PROCESS PATTERN RECOGNITION

Barrington Nevitt
Ontario Development Corp.
Toronto, Canada

Summary

We have developed a communication theory for machines, not people - for matching the old, not for making the new. Communicating with people is by representation, not replica. We have been trying vainly to reduce all human wisdom to "bits" for computer programming. Meanwhile, "software," or the organization of information and the design of structures, has been replacing its "hardware" embodiment at an ever accelerating rate. Information overload has led to process pattern recognition. The Greeks invented Nature by abstracting it from total existence. Greek ability for abstract thought was an unexpected result of their writing technology. Until recently in the Western world, we have continued to look only for "causal connections" in abstract Nature. We have ignored the "resonant responses" of hidden existence to both "hardware" and "software" innovations. The advent of satellites has suddenly forced ecological awareness upon us with total responsibility for programming Planet Polluto. In scrapping man-made Nature we have retrieved the "buzzing confusion" of existence. Today, the groundrules change after every play and "hardware" thinking collapses. As change itself becomes the main staple, percepts replace concepts for discovering "where its at." New knowledge replaces old experience. By using all our senses to make inventories of physical, psychic, and social effects, we can recognize process patterns in the Maelström But we must now learn to hear with the EYE and see with the EAR to make sense and program evitable Fate in ECO-LAND.

Rhyme and Reason

> Aphorisms, representing a knowledge broken, do invite men to inquire farther; whereas Methods, carrying the show of a total, do secure men, as if they were at farthest.
> (Francis Bacon)

> The exact word does not mean the word which exactly describes the object itself. It means the exact word which brings the effect of the word before the reader as it presented itself to the poet's mind at the time of writing the poem.
> (Richard Aldington)

> Man is the measure of all things. There is no other measure.
> (Tobias Danzig)

Magic and Reverse Magic

> To define is to kill. To suggest is to create.
> (Stéphane Mallarmé)

> Computer systems, not men, will first realize humanity's old dream of universal language, and the subtleties and nuances of human thought will risk being mediated through the restricted and standardized symbols of computer communication.
> (Robert Theobald)

> While artists like James Joyce can make a resonant universe with two words, computer programmers try to match universes of human knowledge and perception to the two-bit wit of their machines.
> (Marshall McLuhan and Barrington Nevitt)

Communicating with People is by Re-presentation, Not Replica

The MEANING of a word is not what it says, not merely its definition, but what it does in its context. Meaning for a person flows from his relationship as figure to the totality of physical, psychic, and social facets of his world as ground. "Junior" who cannot read, learns the power of words by their effects. He knows which four-letter words drive mama up the wall! Poetry can communicate before it is understood, so can music. Today, mime, and rhythm, and gesture are replacing words in "silent languages," old and new. The inexpressible lies in the gap between the spoken and the unspoken. What do we know of the grammars of our communication media?

Technologies are extensions of man that create human environments of both services and disservices. The totality of its effects, now and ever is the MESSAGE of any technology. The user is its CONTENT. During the 1940's, Harold Innis[1,2] was the first to show how all human artefacts act as media of communication. The products or figures that we make produce environments or grounds that remake our "human nature" through continuing interplay. What do we know of the differing effects of communication technologies upon ourselves? Do we yet have a valid theory of communication that includes people?

Communication by Making the New

The artist in any field is engaged in exploring and creating effects by works of art that will re-present his INTENT for people. It is a metamorphic process of making the new, not merely of matching the old. The artist learns to recombine the old and familiar to communicate his own unique experience through his chosen medium. The poet, like the actor, "puts on" an audience as the CONTENT of his art in order to sharpen their perception and human awareness. French symbolist poet Baudelaire,[3] writing in the 1850's, "put on" his audience in the classic phrase: "Hypocrite lecteur, mon semblable, mon frère (Hypocrite reader, my likeness, my brother).

W.B. Yeats[4] describes the artist's process of retrieving the old cliché for making the new archetype:

> Those masterful images because complete
> Grew in pure mind, but out of what began?
> A mound of refuse or the sweepings of a street,
> Old kettles, old bottles, and a broken can,
> Old iron, old bones, old rags, that raving slut
> Who keeps the till. Now that my ladder's gone,
> I must lie down where all the ladders start,
> In the foul rag-and-bone shop of the heart.

Spoken language is a technology which extends all of the human senses. Every word is an invention, a breakthrough, which accumulates untold human perceptions through use. Perception as such is a proportion among proportions apprehended in our sensory life. Percepts are the direct sensory experience of our encounters with the world. Concepts are never percepts, but are packages of endlessly repeated percepts - fossilized percepts, as it were, which frequently hinder invention and obscure insight. Humour and poetry release hidden perceptions by sensory interplay and resonant recall. Mark Twain knew that "the right word is the difference between lightening and lightening bug," for it makes sense by involving all the senses. Silent reading of written speech, on the other hand, limits words to the visual sense.

Communication by Matching the Old

In contrast to the artist, the specialist in any field replaces percepts with concepts and eliminates sensory interplay by visual assumptions and logical connections. Scientific terminology by striving toward "one meaning for one word" still further intensifies this process of abstraction. The information theory of both telecommunication and computer reduces the "message" to BITS - signals for yes-or-no choices - as the ultimate "content" of any communication channel whatever. It is a theory of replica for matching the old that treats the new, the unmatched, and the unmatchable alike as "noise." This theory of communication deals with outputs as variant transmissions of inputs. It is for "machines," not people. At wits end, it entirely ignores the "message of the birds" who daily demonstrate that the outputs of every living organism are always transformations of the inputs.

The Medium as the Message

> I like the way he nods. He really communicates.
> (Chairman of the Board)

Human sensory responses are never replicas of sensory impacts. E.H. Gombrich[5] notes:

> What is called "synesthesia," the splashing over of impressions from one sense modality to another, is a fact to which all languages testify. They work both ways - from sight to sound and from sound to sight. We speak of loud colors or of bright sounds, and everyone knows what we mean.... There is touch in such terms as "velvety" voice and "a cold light," taste with "sweet harmonies" of colors and sounds, and so on through countless permutations.

Dr. Herbert E. Krugman,[6] General Electric Company research physiologist, reports his unexpected observation that: "the basic response of the brain wave is clearly to the media and not to the content difference within T.V. commercials, or to what, in our pre-McLuhan days, we would ordinarily have called the commercial message." He thus verifies the McLuhan hypothesis that THE HUMAN PSYCHIC RESPONSE NEVER MATCHES THE MEDIA CONTENT.

Dr. Krugman demonstrates that the brainwave response is unique for each medium and quite independent of its "commercial content." It is also invariant for the viewer's personal preferences, opinions, or value systems. It is a process of making rather than matching. He confirms the paradoxical McLuhan aphorism that, The medium is the message, by another paradox that: The viewer is "transported" by the communication process, whether in print or T V. or any other medium. Dr. Krugman concludes: "The old theory was concerned with the fact that the message was transported. The new theory must be concerned with the fact that the viewer is transported, taken on a trip, an instant trip - even to the moon and beyond."

"Etherealized" by electric media, we become discarnate minds. For "we are there and they are here," instantly. We are reshaped and transformed unconsciously by every medium. The consciously contrived program or intent, like the other media involved in any communication process, ride "piggyback" and are incidental to the message. The message is always the totality of effects, whether intended or not. Partial inventories of the effects of communication media, based on studies made by Marshall McLuhan and his associates at the Center for Culture and Technology, University of Toronto, are given in a previous essay[7].

Languages and Cultures are Media that Transform Perception and Make Different Worlds out of Existence

> One duck to another: "I quack, therefore I am."
> (Cartoon caption in The New Yorker)

Every period of history has dominant metaphors that not only carry meaning across from one mode of being to another but also change it totally. Yesterday's mechanical and today's general system models relate outputs to inputs only as quantitative variants, never as qualitative transformations. They are abstractions "put on" existence as figures without grounds, for they ignore the psychic and social consequences of man-made innovations upon man himself. How did this "royal divorce of thought and feeling come about"?

The Greeks Abstracted Artificial Nature from Total Existence

About 2500 years ago, the discipline of the Greek alphabet[8] began to impose its new visual order upon the "buzzing confusion" of the old aural-oral world. Literate Greeks invented Nature and created cosmos out of chaos. Pre-literate Greeks identified technologies with their gods as services and put the disservices mythically into Pandora's box. Technology, like every codification of human energy or awareness, is myth - encapsulated human drama.

Literate culture produced the psychic detachment - the repeatable "second look," that created Western philosophy, science, and industry. Adam Smith and the classical economists equated the market with nature itself. Every value was for exchange, and every soul had its price. James Joyce described the entire process mythically in one word - ALFORABIT.

Western Civilization Substitutes the Eye for the Ear and Reduces All Order to Visual Order

> Love my label like myself.
> (James Joyce)

Visual space is both the hidden assumption and the "natural" space of Western man. It imposes a visual framework of separate centers with fixed boundaries, clear perspectives, and private points of view upon all existence. Visual space structure is continuous and either contains or excludes definable components or specified properties. It demands yes-or-no matching of "case-hardened" categories. It transforms "magic" into science or "reverse magic." It eliminates all but visual premises "scientifically" by reducing diversity and uniqueness to the uniformity of common denominators. In visual space everything is logically connected, and sequentially ordered in its proper time and proper place. It is the space of Euclid, Descartes, and Newton. It is the mind's EYE.

For 2500 years we have continued to search for the "causal connections" of Greek Nature as abstract figures or isolated concepts in visual frameworks. Meanwhile, we have totally ignored the "resonant responses" - the new hidden grounds or environments "caused" by our arts and artefacts as they transformed us ECO-logically. What have media, and markets, and production lines, and bureaucracy, and transportation, and computers, and satellites done to human nature? By failure to study and to anticipate their psychic and social consequences, we have allowed ourselves to be transmuted into faceless tentacles of our own technological extensions. We have abdicated the power of choice to

Dr. Strangelove's technological imperative: "If it can be done, it must be done!" For we have no epistemology of experience. Are we pure empiricists, like the criminal who said on his way to execution, "This will teach me a good lesson!" Are we "motivated somnabulists," afraid to question our own hidden assumptions? Are we saying in effect: PRIVATE PREMISES - KEEP OFF? Every man has the right to defend his own ignorance!

What the Blind Can See

> A man may see how this world goes with no eyes.
> Look with thine ears.
> (Shakespeare)

In his book And There Was Light[9] Jacques Lusseyran, who was accidentally blinded as a child, expresses his dismay at those who insist upon "ocular proof" and "scientific objectivity," while ignoring acoustic awareness:

> When I came upon the myth of objectivity in certain modern thinkers, it made me angry. So there was only one world for these people, the same for everyone. And all the other worlds were to be counted as illusions left over from the past. Or why not call them by their name - hallucinations? I had to learn to my cost how wrong they were.
>
> From my own experience I knew very well that it was enough to take from a man a memory here, an association there, to deprive him of hearing or sight, for the world to undergo immediate transformation, and for another world, entirely different but entirely coherent, to be born. Another world? Not really. The same world rather, but seen from another angle, and counted in entirely new measures. When this happened, all the hierarchies they called objective were turned upside down, scattered to the four winds, not even like theories but like whims.

Scientific objectivity, the separation of subject and object, is achieved only in short spurts by specialists who suppress all but the visual sense. Human understanding, on the other hand, comes in short bursts of making sense with all the senses. TRUTH is not something we match; it is not a label; it is something we make in a continuing encounter with the world, as the world remakes us.

Mental Set: Western Eye and Eastern Ear

Ernest Fenollosa[10] exposes the contrasting natures of Eastern and Western "explanations:"

> In Europe, if you ask a man to define anything, his definition always moves away from the simple things that he knows perfectly well to abstractions. Thus, if you ask him what red is, he says a colour. If you ask him what a colour is, he tells you it is a vibration or refraction of light, or a division of the spectrum. And if you ask him what vibration is, he tells you it is a mode of energy, or something of that sort, until you arrive at a modality of being or non-being... beyond your depth and his depth. But when the Chinaman wanted ... to define red ... he put together the abbreviated pictures of rose, cherry, iron rust, flamingo ... The Chinese ideogram for red is based on something everyone knows.

Ezra Pound adds that any language written in this way simply had to stay poetic. Fenollosa's essay recalls Francis Bacon's distinction between writing "in Method" and "in Aphorism" Fenollosa and Bacon reveal the basic difference between visual and audile-tactile sensory bias in written language. The same bias is apparent in Japanese flower arrangement: whereas we arrange flowers in space the Japanese arrange the ma or space between flowers. Artists like Shakespeare, Joyce, and Eliot learn to use language both as a visual package for exposition and as an audile-tactile probe for exploration in the complementarity of EYE and EAR.

The Structural Pattern of Acoustic Space is that of Centers Everywhere with Boundaries Nowhere

Acoustic space is inclusive and totally related, but never connected. It is neither a map nor a framework, but a field that has no perspectives, no goals, and no points of view whatever. Acoustic space structure is discontinuous with resonant interfaces or gaps between processes having neither inside nor outside. It is totally involving, like a musical wrap-around. It is the mind's EAR.

The primitive man makes his own space wherever he is. Recently a geologist who had lost his bearings in a Canadian forest, said to his indian crew: "We're lost" Laughingly they replied: "We not lost. Wigwam lost." If you don't know which end is up, ask an astronaut. Each sense makes its own space. We are all new primitives in the electric world.

Perception Leads from Polarization of Absolutes to Complementarity of Processes

> Not only all knowledge, but all feeling, is in perception.
> (T.S. Eliot)

> The child is father of the man.
> (William Wordsworth)

The child remakes the father just as the father makes the child in a continuing process. Rigid definition implies matching of absolutes that permit quantitive variations only. Process complementarity, on the other hand, implies making combinations that lead to qualitative transformations. Concepts ISOLATE whereas percepts RELATE figure and ground.

Affluence creates poverty (not physical hardship). Millionaires on a life raft are not poor but suffer hardship. Instant poverty is the psychic effect of seeing via media what THEY have that WE don't have. In today's electric world, nothing is in camera, everything is on camera. More than a decade ago Indonesia's President Sukarno confirmed that, Hollywood had helped to build up the sense of deprivation of man's birthright that is playing such a large part in the national revolutions of postwar Asia.

Jobs create unemployment (no role-player need apply Mother is never unemployed, only "women's lib" can transform her multiple role for human satisfactions into fragmented jobs for pay. Tribal man was always employed until Civilization brought both nature and "natural" man to market.

Knowledge creates ignorance (not less information). There was no bad grammar or bad spelling before literacy.

The young create the old, just as whites created blacks (by classifying people). Now the "white man's burden" has become the white man!

Order or neg-entropy creates disorder or entropy (not chaos) just as matter makes "empty" space (not void). Nature loves vacuums where the action is, while classifiers reduce them to absolute voids where nothing can happen by definition.

Nils Bohr's complementarity of "wave" (acoustic analogue) and "particle" (visual analogue) for representing "atomic" interaction patterns, exemplifies the complementarity observed in every process involving continuous interplay of simultaneous actions. Such are the plots and sub-plots of human drama. Figure-Ground complementarity, whether stated or hidden, is essential for all human communication. Nothing has a meaning alone.

The New "Ground" of Electric Information Speedup Flips the Old Mechanical "Figures"

> In the last century, we have increased our speeds of communication by 10^7; our speeds of travel by 10^2; our speeds of data handling by 10^6; our energy resources by 10^3; our power of weapons by 10^6; our ability to control disease by something like 10^2.
> (John Platt)

In nature all exponential growth curves lead either to a "flip" or a "flop." When the old Spitfire went through the "sound barrier," the elevator controls reversed their effects.

Alvin Toffler[11] shows how today's information speedup is leading to breakdown in decision making everywhere in the industrial world. But the response of the "tech-fixer" to unanticipated human consequences of technological innovation is always more technology. The latest nostrum is the old plague intensified. Robert Boguslaw[12] notes that, unable to cope with people, "the new utopians are concerned with non-people and with people substitutes." The logic of MAD (Mutual Assured Destruction) now brings everybody to less than twenty minutes from annihilation. In his report to the Supreme Soviet of the USSR on July 10, 1960, Foreign Minister Gromyko[13] confirms that "the command and control of arms are becoming increasingly autonomous, ... from the people who create them ... The decisions made by man depend in the last analysis on the conclusions provided by computers." At wits end, the computer becomes surrogate for man's mind in deciding man's existence. The computer can now do better than ever what must never be done at all.

Assembly line thinking, like an old jalopy, flops at electric speed, for concepts block perception. With no valid theory either of change or stability the human creator of technology becomes its creature. FUTURE SHOCK IS CULTURE LAG.

Software Etherealizes Hardware

> It is no longer resources that limit decision.
> It is decision that makes resource. This is the fundamental revolutionary change.
> (U Thant)

Today, information pours unceasingly and simultaneously upon us from every side. Communication of "software" is replacing transportation of "hardware" by "etherealization" - doing more and more with less and less. "Software" is neither data, like a telephone directory, nor mere computer programming; it is the organization of information and the design of structures. As the "hardware" embodiment of "software" vanishes, thinking becomes doing. The mental and the manual merge as man becomes information hunter and processor. The relation between "hardware" and "software" is not fixed, but is in a state of perpetual flux and metamorphosis (like figure and ground, or experience and knowledge)

Dialogue is a process that enriches through diversity by making new "software," not by merely repeating old formulas. Dialogue bypasses the assembly line, and the market, and the "Laws of Economics." Software unlike hardware is wealth that increases with exchange and use by transforming ignorance into knowledge.

The Hidden"Rim Spin"

The telephone bypasses the memo and all other "channels of communication" prescribed by the organization chart. Monopolies of knowledge and control diminish in proportion to information speedup. Xerox leaks have now become public purges. The faster "rim spin" envelops, erodes, and transforms the slower. Today's hidden "rim spin" is a child of the electric telegraph invented by nineteenth-century artist Samuel Morse. Telegraph lines and cables, following the old transportation routes, converted our planet into a "wired city." The etherealization of "wire" into "wireless" created today's "magnetic city" of radio and T.V. This invisible "rim spin" has transformed the old world of visual connectedness into a new world of audile-tactile gaps and resonances - a "global theatre" of instantaneous, all-at-once awareness.

Suddenly, via satellites, we have reached a new "aerial perspective" that has made everybody everywhere intensely aware of "Planet Polluto." While we have been blindly remaking old Mother Nature, she has been remaking our human nature. We have inadvertently taken over nature's role and can no longer avoid full responsibility for the consequences of our actions on Spaceship Earth. All fixed premises have gone with the ether. With neither plot nor plan nor any intent whatever, the effect of speeding up "civilized" technology in visual space has been to create a new "tribal" electric world of acoustic space. ALL HOLDS ARE BROKEN as past, present, and future merge in mythic NOWNESS.

Yesterday's "reality" was in visual space and meant matching the old. Today's "reality" is in acoustic space and means making the new.

Pushed to extreme, abstract Greek Nature has retrieved the barbaric babble of total existence as the EYE once more yields precedence to the EAR. Pre-Socratic philosopher Heracleitus (540-480 BC) during the shift in human sensibilities from tribal EAR to civilized EYE was fully aware that "you cannot step twice into the same river for fresh waters are ever flowing in upon you." In today's electric world, as for Heracleitus, only change itself is stable.

There is nothing old under the sun. It is precisely the instant dimension of electric living that constitutes both the problem and the solution. The hang-up is to transfer the instant character of the new information environment to the old assembly line. The old figure of "change" has become the new ground, while the old ground of "stability" has become the new figure, with new groundrules after every play. The old game ends, where the new sport begins, in the flip from conserving stability via feedback to anticipating change by process pattern recognition and feedforward.

Understanding is neither a point-of-view nor a value judgement. Understanding requires simultaneous apprehension of all facets of any "live" situation whatever. Edgar Allan Poe,[14] symbolist poet and "father of the detective story," provides the clue to the Maelström. Only by recognizing the pattern of the whirlpool's effects and discarding all preconceptions did the hero of his story learn to reprogram fate and escape "sure death." Like Hercules in cleansing the Augean stables, he found the solution not in external criteria, but in the problem itself. Information overload as breakdown leads to process pattern recognition as breakthrough.

Every Process Pushed to Extremes of Quantity, Speed, or Intensity Reverses Its Effects

More than 4000 years ago Chinese sages recognized the process pattern of chiasmus in their I Ching[15]: "Whatever has reached its peak must turn back." In the nineteenth century, German philosopher Georg Hegel IMPOSED chiasmus categorically upon abstract Greek Nature as "The Transformation of Quantity into Quality."

Pushed to extreme, law becomes ordure and virtue becomes vice, just as food becomes poison, and existence becomes "Existentialism." Extreme adaptation leads either to destruction (like the dinosaur) or to metamorphosis (as a new form). The repeated visual image becomes the resonant audile-tactile mosaic - a magic spell. Visual structures become acoustic, just as acoustic structures become visual in the constantly changing figure-ground relationships not of abstract logic, but of existence itself.

Nothing exceeds like excess. When a parrot breeder finally succeeded in developing a bird weighing three hundred pounds, he taught it to say, "Polly wants a cracker." The parrot instantly added,"AND I MEAN NOW!"

Electric information speeds EXPOSE chiasmus in power, in wealth, and in community. The chiasmus of speedup itself is slowdown - the return to human scale. The new "software" ground has reversed the old "hardware" groundrules. THEREFORE, PLAN VERY CAREFULLY AND PREPARE TO DO THE OPPOSITE.

The Ouroboros Pattern: Figure-Ground Merger as Monster

When the figure swallows the ground it produces a monster. Tribalism as the ground of pre-literate humanity was a monster of "Rhyme without Reason" that swallowed man's private identity. Civilization, the artefact of literate man, when pushed to extreme, becomes a monster of "Reason without Rhyme" that devours all humanity. No matter how well-intentioned the monolithic state, or the uncontrolled business monopoly, or the all-embracing system eventually becomes monstrous, for interplay requires a figure-ground interface. The Greek ouroboros or snake that ate its own tail is the ancient symbol for a world that survives by endlessly devouring itself.

In contrast, the mythic Phoenix constantly renewed itself through rebirth from the ashes of its own funeral pyre. The Phoenix process pattern re-presents multiple figure-ground reversals or the chiasmus of chiasmus.

Whereas in Visual Space the Connection Becomes the Hangup, in Acoustic Space the Action is in the Gap

> For your information, let me ask you a question.
> (Sam Goldwyn)

In the relation between wheel and axle from "seize-up" to "play" to collapse, the wheeling ceases unless the gap is right. The "play" and the "action" are one. As Nobel prize winner, Linus Pauling has shown, the only material connection left in the material universe is the "resonant chemical bond" of discontinuous interacting quanta. Whereas "missing links" stimulated nineteenth-century re-searches, "resonant gaps" are the beat and rhythm of twentieth-century compositions. Questions, not answers, have become our main resource.

Normal Science as a Paradigm Trap

Normal science organizes knowledge, and labels rather than processes. Classification as such hinders the organization and the sharing of ignorance through dialogue that can transform breakdowns into breakthroughs. T.S. Kuhn[16] discusses some hidden implications of normal science which is a "paradigm trap" preventing discoveries:

> No part of the aim of normal science is to call forth new sorts of phenomenon.... One of the reasons why normal science seems to progress so rapidly is that its practitioners concentrate on problems that only their own lack of ingenuity should keep them from solving.

Scientists make discoveries in acoustic, not visual space, as people rather than as specialists. Concepts always impede percepts. Discoveries are made in resonant gaps, not in logical connections. But the mere visualizer is satisfied only if he can see what is resonating.

"Minerva's owl takes its flight at twilight" - the vision at the magical interface between sleeping and waking. The Incas had one word for both poet and inventor - hamavec. Mission management of research and development has no place for either, for artists and inventors are the ultimate enemies of everything established. While inventors create new products and processes that transform environments, artists make new perception that changes all the social groundrules. We might well ask what discoveries have not been made by "hams"?

The Displacement of Concepts or the Rear-View Mirror

> To most men, experience is like the stern lights of a ship, which illumines only the track it has passed. (Samuel T. Coleridge)

The use of concepts alone to discover "where its at" leads only to hindsight. Donald A. Schon[17] shows that "the tendency either to obscurantize or to explain away novelty reflects the great difficulty of explaining it. The difficulty comes in large part from our inclination, with things and thought alike, to take an after-the-fact view."

Our rear-view mirror displays not only the vanishing ground of a nostalgic past but the oncoming figure of a future terror. Philosophy is hindsight disguised as foresight, for classification always rejects the breakthrough of process pattern recognition.

The Cliché-Archetype Pattern

> Try somehow to get some money. Someday it may be valuable.
> (current wit)

In the cliché-archetype process, obsolescence occurs at the peak of popularity and use. Books were made obsolete by electric information i.e. old book forms were scrapped as new varieties proliferated in greater quantities than ever. Similarly, print supplanted writing, but there is more writing today (long-hand and short-hand) than in all times past. When enveloped by the new, the old form becomes art form. What is new in any innovation are its effects. When obsolescence outstrips innovation, invention becomes the mother of necessities, new and old.

Every new technology retrieves a much older one, just as every new mode of knowledge brings back some more ancient form. The electric world has retrieved

ESP (extra-sensory perception), ancient medicine (acupuncture), and occult arts (zen, yoga, and astrology). When the latest model (car, radio, or TV) displaces its predecessor, the old structure becomes highly visible for the first time, either through breakdown and repair or as junk. Its components may be re-used (old spare parts or new "found art") and if not destroyed they may reappear as "camp," or ultimately as museum pieces. OLD TRASH BECOMES NEW TREASURE AS THE CLICHÉ RETRIEVES THE ARCHETYPE.

Archeologists and astronomers "predict" the remotest past by using the latest technologies (radio-active dating and satellite astronomy). Why do scientists treat Velikovsky as a rejected medium rather than as a possible probe in his humanist approach to problems of astronomy? As one famous astronomer put it: "If Velikovsky is right, we must all be crazy!" But "proof" of sanity is available only to those discharged from mental institutions. THE FUTURE IS NOT WHAT IT USED TO BE, NEITHER IS THE PAST.

The Displacement of Percepts or the Bridge for Wits

> The role of art is to create means of perception by creating counter-environments that open the door of perception to people otherwise numbed in a non-perceivable situation.
> (Marshall McLuhan and Harley Parker)

Today's "survival kit" requires new perceptual training and fresh wits, rather than new concepts. The poets and artists are engaged in attempts to intensify perception of the present and to anticipate the effects of new technologies by constant exploration. Like inventors, instead of discarding the old traditional forms of culture, they are constantly striving to imbue the old components with new and youthful vigour. THE FUTURE OF THE FUTURE IS THE PRESENT. The art of probing the present, in order to "predict" what has already happened, requires careful and deliberate dislocation of ordinary perception. The "artist" in any field seeks to attain fresh vision through the juxtaposing of ordinary things in extraordinary ways. Like James Joyce, he explores "Where the hand of man has never set foot." The "artist" plays the field of clichés to retrieve and to create archetypes as fresh figures in the new grounds. Ignorance, noise, and the midden heaps of history are his raw materials for making discoveries.

The Comic Converts his Hangups and Grievances into Jokes as Percepts that Lead to Instant Understanding

> Gravity, a mysterious carriage of the body to conceal the defects of the mind.
> (Laurence Sterne)
>
> Jesters do oft prove prophets.
> (Shakespeare)

Jokes are percepts, never concepts. Perception inspires levity, not gravity. The clown tests the environment to show the emperor "where it's at;" he is an audience researcher minus preconceptions. An inventory of jokes or gripes can reveal causes of friction and breakdown in any human situation whatever. Wits choose their jokes and puns deliberately to break down conceptual barriers and break through to fresh insights.

Newton did not discover "gravity" but "levity." Although everybody knew that apples fell downward to the earth, only Newton recognized that they also felt the urge to fall upward to the stars. Such "action-at-a-distance" by gravitational "resonant bonds" is still a hangup for those who can conceive only visual connections. For the man of visual bias objects do not create time and space, they are merely contained in time and space. Understanding Einstein's Relativity demands a change from mechanical models of rigidly connected things to audile-tactile re-presentations of resonant intervals between processes. People, unwilling or unable to explore the process patterns of current happenings, resort to gloom-pouring and moral indignation, the age-old substitutes for understanding.

You cannot package live processes. Measurement kills what you love. T.S. Eliot[18] sums up: "That's all the facts when you come to brass tacks: Birth, and copulation, and death." For the insurance company, death is the ultimate measure of life. THEREFORE, COUNT YOUR BLESSINGS, BUT DON'T TRY TO EVALUATE THEM.

The Paradox of Communication

The effect of increasing access to speedier communication has been to lessen human tolerance and mutual understanding everywhere. Communication speedup has created a crisis of identity that is a simultaneous quest for change and conformity. Why have our hopes failed?

> The rim spin of the electric world annihilates the very image of one's self. The young are deprived of both identity and goals .... The UNPERSON is the inevitable result of improved communication. When all barriers of private consciousness are overcome, the resulting collective form of awareness is a tribal dream .... The fragility and insecurity of tribal life lead to violence as a quest for identity in pre-literate and post-literate societies alike.[19]

We have developed a theory of communication, which fits "general systems," but which leads to consequences quite opposite to our intent when people are fitted to it. For our theory has ignored both the conflict and the complementarity of EYE and EAR. Have we become what we behold?

Bridges from Petty wits to Giant Brains

Computer users have sometimes observed that "information overload leads to pattern recognition." However, such "pattern recognition" is not mere identification of visual forms; it is rather making sense of all sensory inputs by human wit. Computer programmers of the future will no longer be confined to two-bit logic in visual space. They will also be engaged in orchestrating the computer's "auditory imagination" in acoustic space. Poet T.S. Eliot[20] elucidates:

> What I call the "auditory imagination" is the feeling for syllable and rhythm, penetrating far below the conscious levels of thought and feeling, invigorating every word: sinking to the most primitive and forgotten, returning to the origin and bringing something back, seeking the beginning and the end. It works through meanings, certainly, or not without meanings in the ordinary sense, and fuses the old and obliterated, and the trite, the current, and the new and the surprising, the most ancient and the most civilized mentality.

Auditory imagination is the mind's ear - the complement of visual imagination. Can we simulate an intelligence of "Rhyme and Reason" with anything less?

Marshall McLuhan has suggested that consciousness itself results from both the constant interplay and the instant retrieval and scrapping of all sensory impressions. Consciousness would thus be a process whereby the brain continually forms archetypal modules of information, while converting them through instant use into clichés that are immediately discarded and replaced by new ones. Can we construct such a conscious computer?

The ancients knew that mimesis is the process by which all men learn. Mimesis is not matching, but remaking the experience of a culture by using all the senses. Without senses to use, the computer's two-bit wit pushes all breakthroughs to breakdowns. It is for matching, not mimesis. Can the computer ever learn to make sense except as an extension of man's brain?

From Job as Specialist to Role as Comprehensivist

> Hold to the now and here from which all future plunges into the past.
> (James Joyce)

The job of the specialist is to push his paradigms back to get history and forward to get science. But the role of the comprehensivist is to make resonant bridges between old and new perception, the civilized EYE and the tribal EAR. The comprehensivist uses concepts, models, and theories, not for programming, but for probing the "now and here." He lives simultaneously in "stereo-perspective" on both sides of Alice's Looking Glass - beyond the hangups of technology and ideology. He plays the role of "Finn-again-awake" revealed by James Joyce[21] in the myth of Finnegan's Wake. The new science is of percepts, not concepts.

At electric speeds we can no longer afford to wait for experience by feedback. We must know the services and disservices to be expected before undertaking any major innovation. In the contrast and comparison of developing situations, we can perceive the relevant process patterns by making inventories of their effects. Process pattern recognition is the "causality" of our electric age. But the patterns lie inside our problems themselves, not in outside criteria. We can TAKE TODAY and become prophets by being the first to perceive what has already happened. With this knowledge we can reprogram "inexorable Fate" by feedforward.

As knowledge replaces experience, Everyman becomes a measure for all things in ECO-LAND. Participatory Democracy becomes Anticipatory Democracy when perception takes command. TODAY WE CAN CHOOSE EFFECTS TO PRECEDE CAUSES, FOR THOUGHT TRAVELS MUCH FASTER AND FARTHER THAN LIGHT.

References

1. Harold Innis, The Bias of Communication (Toronto: University of Toronto Press, 1964)

2. — Empire and Communication (Toronto: University of Toronto Press, 1972)

3. Charles Baudelaire, Les Fleures du Mal (New York: French and European Publications, 1961)

4. W.B. Yeats, "The Circus Animals' Desertion," Poems (New York: Macmillan, 1957)

5. E.H. Gombrich, Art and Illusion (New York: Pantheon Book, 1961) pp. 366-7

6. Herbert E. Krugman, "Electro-encephalographic aspects of low involvement: Implications for the McLuhan hypothesis" (Lake George, New York: Meeting of the American Association of Public Opinion Research, May 21 - 23, 1970)

7 Barrington Nevitt, "Communicating with people through media" (THE, No.1, Northern Electric Research and Development Laboratories, Ottawa, Ont.)

8. According to David Diringer, The Alphabet (New York: Philosophic Library, 1948) the Greek alphabet was the first to use symbols that had no sematic meaning either visual or acoustic. The unexpected and largely ignored effect of this Greek technological innovation is that it produced the psychic visual bias that gave Western civilization its unique character.

9. Jacques Lusseyran, And There Was Light (Boston: Little, Brown & Co., 1963) p. 112

10. Ernest Fenollosa, "Essay on the Chinese Written Character," cited by Ezra Pound, The ABC of Reading (New Directions Paperback) pp. 19-22

11. Alvin Toffler, Future Shock (New York: Random House, 1970)

12. Robert Boguslaw, The New Utopians (Englewood Cliffs N.J: Prentice-Hall, 1968) p. 2

13. Cited by Herbert F. York in "ABM, MIRV, and the Arms Race," (SCIENCE, July 17, 1970)

14. "A Descent into the Maelström" in The Complete Tales and Poems of Edgar Allan Poe (Modern Library, Random House, 1938)

15. I Ching (or The Book of Changes) Translated from Chinese and German by R. Wilhelm and C.F. Baynes (London: Routledge and Kegan Paul, 1951)

16. Thos. S. Kuhn, The Structure of Scientific Revolutions (Chicago: University of Chicago Press, 1962) pp. 52-53

17. Donald A. Schon, Displacements of Concepts (London: Tavistock Publications, 1963) p. 4

18. "Sweeny Agonistes" in The Complete Poems and Plays of T.S. Eliot (London: Faber and Faber, 1970)

19. Marshall McLuhan and Barrington Nevitt, Take Today: The Executive as Dropout (New York: Harcourt, Brace, Jovanovich, 1972) pp.259-264

20. T.S. Eliot, The Use of Poetry and the Use of Criticism (New York: Barnes & Noble, 1955) p. 63

21. James Joyce, Finnegan's Wake (New York: Viking Press, 1959.

22. The Modeling of Mind: Computers and Intelligence, edited by Kenneth M. Sayre and Frederick J. Crosson (New York: Clarion Book, 1968). An inventory of computer hangups useful for organizing ignorance to make discoveries.

ALTERNATIVE FUTURE COMPUTER-COMMUNICATION MARKETS*

D. A. Dunn
Stanford University
Stanford, California, U.S.A.

Summary

The potential of computer-communication systems as a means for offering a wide variety of new services is explored in terms of technology, cost structure, capability to capture the full value of information services for creators of such services, and market structure. The market structure under which computer-communication systems will eventually operate in the U.S. is strongly dependent on the regulatory environment under which cable television develops and the regulations governing competition between the telephone carriers and cable television system operators. Several realistic regulatory alternatives which could be adopted would result in widely differing competitive environments. The provision of some form of nondiscriminatory access to at least one of these networks in every community is essential to the operation of a competitive market in information services.

1. Introduction

The market for computer-communication systems is one part of the market for information services generally. The information market is characterized by several unusual features. First, information in marketable form is usually costly to create. Second, it is usually much less costly to duplicate it than it is to create it. Third, it is difficult to effectively transfer limited rights to use information, such as the right to read a book without the right to copy it, and it is therefore difficult for the creator of information to appropriate the full value of his creation. These features of the information market make it unprofitable for artists and entrepreneurs to enter some parts of the market, while other parts, in which the market functions effectively, are among the most profitable parts of the world economy.

The particular technology involved in the transfer of information from the creator to the purchaser-user is a critical element of the market structure. Some technologies are well-adapted to allow the information creator to appropriate the full value of his creation. Other technologies have inherent difficulties in this regard. The technology of computer-communications potentially offers a high degree of appropriability to information creators who use it to offer information services to users in their offices and homes. Consequently, the part of the information market which uses computer-communication systems as the vehicle of information transfer is potentially an effective market and hence attractive to information creators and entrepreneurs.

In order to realize the full potential of computer-communication technology the appropriate legal institutions must be created to go with this technology. Some institutional arrangements will favor the operation of a market while others can severely constrain the development of the market for new information services. By limiting the development of the market for new services, such institutional arrangements can also limit the development of new systems and technology.

In this paper some aspects of computer-communication technology that affect marketability of services and the basic cost structure of information services are briefly considered. Then several alternative U.S. legal institutional arrangements believed to be possible are compared with respect to the effects that these choices could have on the future development of the markets for computer-communication systems and for information services offered over these systems.

2. Computer-Communication Systems

In this paper the definition of a computer-communication systems is as shown in Fig. 1 which is a diagram of a system of this type with a central studio with live origination, tape, and computer facilities connected to a number of user terminals through communication links. Each user terminal includes a video and/or audio receiver and also audio and/or data input devices. Data input is by means of some type of keyboard and/or a light pen or equivalent means for indicating the user's choice among two or more items shown on the CRT display. At the central studio a switching computer is used to poll the users and to route their messages to a variety of information sources which may be humans or computers which contain central data files.

Such a system may be viewed as a library with a wide variety of information services available, packaged in many forms. A user seeking information on a particular subject such as computers may wish to obtain information in the form of a novel about a computer, a movie made from the novel, a short course in programming, a research paper, a photograph of a particular computer with its specifications, music composed by computers, etc. Some of the media that are readily available through a receiver system of the type shown are listed below.

1. Printed page - viewed as a still picture on the CRT display and/or received as a hard copy.
2. Still photographs - viewed as a still picture on a CRT display and/or received as a hard copy.
3. Sequential still photographs - viewed on the CRT display at rates up to 30 per second (at which point of course motion video is obtained) and/or received as hard copies or as a videotape recording.

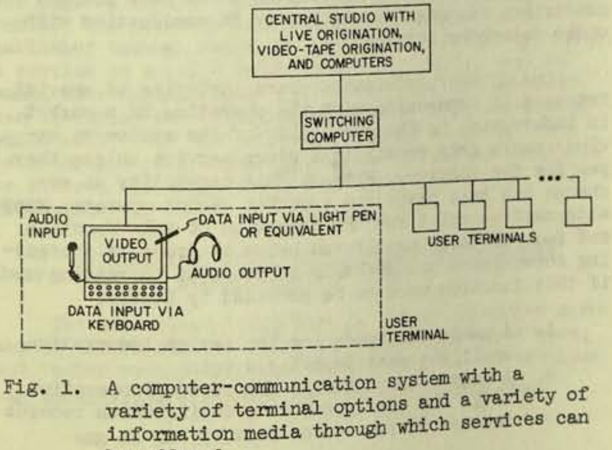


Fig. 1. A computer-communication system with a variety of terminal options and a variety of information media through which services can be offered.

---

*This work was supported by National Science Foundation Grant GR-86.

4. Motion video viewed on the CRT display and/or received as a videotape recording.
5. Audio received on earphones and/or recorded on tape.

The ability to obtain either a hard copy or a videotape recording of what the user has seen on his CRT is an important addition to each service. The costs of these service options are strongly dependent on whether the copies are made without significant delay at the users' terminal or at the central studio and mailed to the user with an attendant delay of a day or more (but at greatly reduced cost). Various video quality options exist as well, such as the number of lines in the raster, color, screen size, etc.

For the user input the same options are of course available as for the studio, but the cost of some of these options is high. A list of media options for the user input terminal which are likely to be available to the home user in the coming decade at modest cost is as follows.

1. Alphanumeric keyboard or twelve button pads
2. Light pen or mechanical equivalent to indicate user choice of items listed on his CRT by direct positioning of an indicator on the screen
3. Audio (microphone).

This list of the media now technically feasible allows us to visualize a variety of different computer-communication systems providing a wide variety of services, many of which do not now exist. It is not the purpose of this paper to explore the costs of the many technological alternatives suggested by such a list, but it is evident that cost is one of the critical factors in projecting the future market for such systems along with performance characteristics [1], [2].

Most of the system cost-performance tradeoffs involve both the terminal designs and the cost and characteristics of the communication network which connects users with the central studio [3], [4]. Three major network alternatives exist: the public telephone network, the cable television network, and private-line networks. Each of these has special characteristics which make certain kinds of system designs especially attractive. For example, the near-complete national two-way interconnectability of users with voiceband lines through the telephone network makes many systems immediately possible through this network that can only be considered as feasible a decade or more in the future, if the cable television network is used. Similarly, certain broadband information-services are likely to be most economical if offered either in conjunction with entertainment services provided over the cable television networks of the future which have limited subscriber response capability or in conjunction with video telephone service.

One of the performance characteristics of special interest in connection with the operation of a market in information is the capability of the system to exclude users from receiving a given service unless they pay for the service. Without this capability no exchange can take place and a market cannot operate. Many alternatives exist for the provision of this function and four of these are listed below in order of increasing convenience and probably in order of increasing cost if this function were to be provided by itself.

1. A means for rendering the system inoperative until the user deposits a coin.
2. A means for rendering the system inoperative until the user actuates a device which records the use of the system locally, as in gas or electric power meters. Such a metered system need only be read monthly, so that the need for two-way communication for this purpose alone is very limited.
3. A means for rendering the system inoperative until the user sends a short data message over his telephone. Such a system could be used to allow a relatively simple pay-television system to be built which uses existing one-way cable television networks. This alternative also covers uses of the telephone network for information distribution.
4. A means for rendering the system inoperative until the user sends a short data message over the network being used to send information to him. If the network being used is the telephone network, this alternative is the same as the preceding one. If the network being used is the cable television network, provision must be made to carry data signals upstream in order to provide this option.

In addition to performing these basic exchange functions, it is necessary for any successful system to provide information desired by users in a form and in a medium that users want sufficiently that they will pay enough to cover the costs of creating and distributing the information. Computer-communication systems have both a wide range of technological options, as discussed above, and the possibility of being made available in essentially every home and office. Thus, the medium can (to a considerable extent) be matched to the message, and the market can be as large as necessary to allow the costs of creation to be distributed over a sufficiently wide audience to make creation costs less than other costs and to make total costs less than the price at which the information can be sold.

3. The Cost of Information

The cost structure of information production is very similar to the cost structure of manufactured goods like automobiles, but there are significant differences in the relative magnitudes of the cost components. These differences in relative magnitude cause the market for information to have several unusual features.

The cost components to be considered here are as follows.

1. The cost of creating the first unit = C_1
2. The cost of preparing the second unit in a form suitable for reproduction = C_2
3. The cost of reproducing an additional unit identical to the second unit = C_3
4. The cost of delivering a unit to a user, including the cost of an exchange mechanism whereby the user can be denied the information until payment can be assured = C_4

The total cost of one unit of the good, C_t, when the good is shared among n users is then

$$C_t = \frac{C_1}{n} + \frac{C_2}{n} + C_3 + C_4 \ . \qquad (1)$$

In the context of manufactured goods such as automobiles, the first term is the design cost, the second term is the development cost, the third term is the production cost, and the fourth term is the marketing cost. In context of an information unit such as a book, the first term is the cost of the author's time and the cost of the typing the manuscript, the second term is the cost of editing, typesetting, and producing page proofs, the third term is the production cost, and the fourth term is the marketing cost.

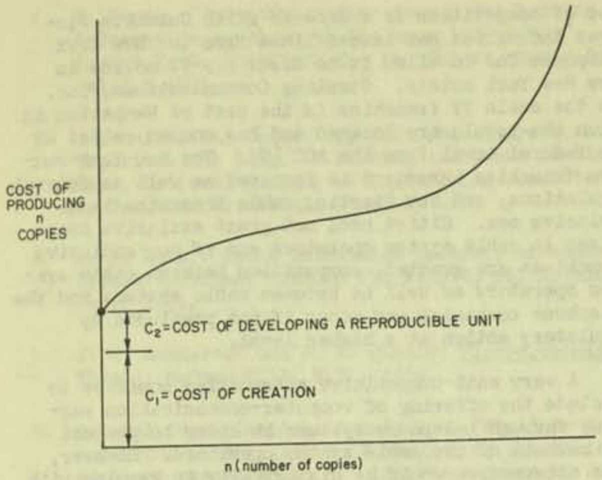


Fig. 2. Cost of producing n copies of a unit of information as a function of the number of copies produced, n.

Figure 2 is a plot of total cost, nC_t, for n units of a good, whether a manufactured good or information, as a function of the quantity n. This cost curve has a shape which is the same for almost all goods, with an initial cost, independent of quantity, here equal to $C_1 + C_2$, and region of declining marginal cost followed by a region of increasing marginal cost at very large quantities [5]. The region of increasing marginal cost is met in manufactured goods at a quantity for which, for example, a raw material is required in such large quantities that its price increases. In information production such an effect might be encountered when a given plant capacity began to be reached, causing excessive equipment breakdowns and increased labor costs.

Figure 3 compares these four cost components on a relative basis for hypothetical cases of a typical manufactured good such as an automobile and an information unit such as a book in quantities of 1,000 and 10,000. Production cost dominates in the case of the automobile for quantities of 1,000 or more. The quantity at which production cost plus marketing cost exceeds half the total cost occurs for most manufactured goods for lots of a few hundred or a few thousand. For many kinds of information units this critical quantity is of the order of millions. For the book in Fig. 3 the critical quantity is between 1,000 and 10,000.

A serious problem in the marketing of information arises out of the fact that, even with inefficient production techniques, a user can often produce a copy of the information at less cost than the price of the information that the creator must charge to cover his costs, if the quantity produced is less than this critical quantity. Under these conditions the marginal cost of production, including the cost of creation, is greater than the cost of copying the information and therefore such information will not ordinarily be produced. The copyright laws are, of course, intended to prevent unauthorized copying and therefore to create a market in information produced in smaller quantities than the critical quantity. The copyright laws are relatively easily enforced against violators who attempt to sell unauthorized copies of information on the open market. However, it is essentially impossible to protect against a user copying information for his own use. For example, print materials can be copied for a few cents a page and audio and video materials can be transferred to reusable magnetic tape at very modest cost.

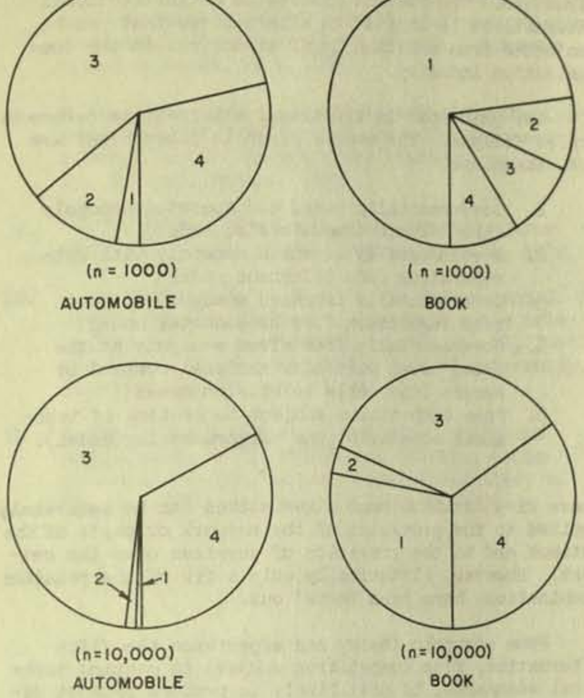


Fig. 3. Relative costs of (1) creation, (2) development of a reproducible unit, (3) production, and (4) marketing for a hypothetical case of a manufactured good such as an automobile and an information unit such as a book in quantities of 1,000 and 10,000.

One of the most important and significant features of a computer-communication system is the fact that it combines a technical capability for charging for each use of information stored in its files with a capability to assist the user in finding the information he wants. The interactive search capability and the potential for keeping stored information up to date make such a facility more like a library than a book. Therefore, such a system does not simply produce information which can be copied, but rather it provides a service for which the user may be willing to pay much more than the cost of making a copy of the final pages of hard copy obtained. For example, a user may spend half an hour searching the files for information in a given topic and finally end up with one page of information of interest to him. He may be quite willing to pay several dollars for this service, whereas he would be reluctant to pay more than a few cents for the page itself, if he has previously determined that this is the page he needs. Thus, a computer-communication system has the potential for allowing a market to be established through which an information creator can offer a service at a price which covers his costs, even to small user populations. Of course, complete services can be copied and competitive services will be created which will force the price down nearly to cost. What is important is the possibility that creation costs can be covered for many new services that previously could not exist.

4. Alternative Institutional Forms for Computer-Communication Systems

The same basic technology is likely to evolve under any of the institutional alternatives considered here, but rather widely different future markets for services can result from different choices of institutional structure. The present discussion of institutional alternatives is limited to alternatives that can be projected from existing legal structures in the communication industry.

The principal institutional alternatives before us are as follows. The models given in parentheses are U.S. examples.

1. Governmentally owned and operated monopoly (the defense communication model)
2. Governmentally licensed monopoly with rate regulation (the telephone model)
3. Governmentally licensed monopoly without rate regulation (the broadcaster model)
4. Governmentally franchised monopoly at the local level with some national control of access (the cable television model)
5. Free competition subject to meeting of technical standards (the teleprocessing model).

These five institutional alternatives can be separately applied to the provision of the network or parts of the network and to the provision of services over the network. However, historically only a few of the possible combinations have been tested out.

From economic theory and experience the fifth alternative, free competition subject to meeting technical standards, is most likely to produce product diversity and low cost. This alternative is most likely to be applied to the provision of information services through the telephone network in the U.S., because of the precedent that has been established in the field of teleprocessing [6], [7], [8]. Thus the most likely alternative for U.S. computer-communication systems which use the telephone network or a network provided by a competing specialized carrier is a combination of alternative 2 for the network itself and alternative 5 for the user terminals, central studio hardware, and the actual service itself including creation of software and programming.

It is by no means clear that this same alternative will be selected for services provided through the cable television network in the U.S. which now operate under alternative 4 in most U.S. cities. Since cable television as an industry has many historical and operational connections with the broadcast industry, there are many forces at work which cause cable television regulation to resemble broadcast regulation. Broadcasters have historically viewed themselves as having combined responsibility for both hardware and program content, unlike the telephone industry which leaves message content to its subscribers. If cable television evolves in a mode which is closer to the broadcaster model with respect to the provision of new information services, we can expect to see these services offered to users either directly by cable system operators or under exclusive agreements with these operators. On the other hand, if cable television evolves in a mode which assures entrepreneurs of new services non-discriminatory access to cable channels, in much the same way that telephone channels can be leased for data processing services, we can expect to see a much more competitive situation in the provision of services over cable channels [10]. Present FCC rules make provision for leased channels, but the FCC has postponed consideration of rate regulation until some experience has been gained [11].

In either case, an important question which has not yet been resolved is that of the rules governing competition between service offerings which can reach a user through either the telephone system or through the cable system. One of the first examples of this type of competition is a case in which Columbia Pictures Industries has leased lines from the New York Telephone Co. to allow it to offer pay-TV movies in five New York hotels. Sterling Communications, Inc. has the cable TV franchise in the part of Manhattan in which the hotels are located and has sought relief at the federal level from the FCC [9]. The New York-Sterling franchise agreement is involved as well as federal regulations, and the Sterling cable franchise is an exclusive one. Cities need not grant exclusive franchises to cable system operators and if non-exclusive franchises are granted, competition between cable systems operators as well as between cable systems and the telephone companies can occur if not precluded by regulatory action at a higher level.

A very anti-competitive alternative would be to preclude the offering of computer-communication services through telephone systems in order to protect the markets of the cable system operators. However, this alternative would be historically in keeping with the actions of some U.S. regulatory bodies in the past. For example, when the trucking industry began to compete with the railroads, rather than de-regulating the railroads and allowing competition to occur the ICC regulated the trucking industry.

The alternative futures that appear most plausible are thus as follows, in order of increasing competition in the provision of services.

1. Exclusive cable franchises are used as a basis for granting exclusive "sub-franchises" to certain providers of services within the the monopoly area of the cable franchisee. Telephone companies are prevented from competing with cable operators in the provision of network services. Many services are provided by subsidiaries or divisions of the large multiple-system cable operators, and providers of information services seeking to compete are kept out of these systems altogether. This alternative is closest to the present broadcasting industry.
2. Cable operators and the telephone companies are allowed to compete, but no special rules are adopted to force the cable operators to provide nondiscriminatory access to their channels. Telephone companies do offer nondiscriminatory access and a wide range of independent providers of services evolve who compete through telephone company cables with each other and the cable system operators and their licensees.
3. Cable systems are forced to provide nondiscriminatory access to their channels. Cable and telephone systems evolve as competing sources of channels with independent providers of services competing through both systems.

Without a more extensive analysis, not much can be said about the relative sizes of the markets that would be expected under each of these alternatives. However, it is not difficult to say from economic theory that prices will be lower and the total quantity and diversity of service offerings will be greater for the more competitive than for the less competitive alternatives. It is also fair to say that more users will be served under the more competitive alternatives and that the market for both hardware and software will be greater as we move from alternative 1 toward alternative 3.

References

1. W. S. Baer, "Interactive television: prospects for two-way services on cable," Report R-888-MF, Rand Corp., Santa Monica, Calif.; Nov., 1971.

2. P. Baran, "Potential market demand for two-way information services to the home 1970-1990," Report R-26, Institute for the Future, Menlo Park, Calif.; 1971.

3. Communications Technology for Urban Improvement, Report to the Dept. of Housing and Urban Development, National Academy of Engineering Committee on Telecommunications; June, 1971.

4. D. A. Dunn, "Cable television delivery of educational services," EASCON '71 Record, pp. 157-163; Oct., 1971.

5. J. M. Henderson and R. E. Quandt, Microeconomic Theory, McGraw Hill, N.Y., 1958.

6. Federal Communications Commission, "Final decision and order in the matter of regulatory and policy problems presented by the interdependence of computer and communication services and facilities," 28 FCC 2d 267-308, Mar. 18, 1971.

7. D. A. Dunn, et. al., "Policy issues presented by the interdependence of computer and communication services," SRI Report 7379B (2 vols.) Commerce Clearing House, Feb., 1969.

8. D. A. Dunn, "Policy issues presented by the interdependence of computer and communications services," Law and Contemporary Problems, v. 34, pp. 369-388; Spring, 1969.

9. "CATV concern files complaint with FCC against AT&T unit," Wall Street Journal, July 11, 1972.

10. J. H. Barton, D. A. Dunn, E. B. Parker, and J. N. Rosse, "Nondiscriminatory access to cable television channels," Discussion Paper No. 11, Inst. for Public Policy Analysis, Stanford University, Stanford, Calif., May, 1972.

11. Federal Communications Commission, Rules and Regulations, Cable Television Service, Cable Television Relay Service; Federal Register, v. 37, pp. 3252-3341; No. 30; Feb. 12, 1972.

MEASURING THE COMPUTER'S IMPACT ON ORGANIZATIONAL STRUCTURE

David W. Conrath
University of Waterloo
Waterloo, Ontario, Canada

ABSTRACT

An organization's structure is described as a set of relations among the persons who are members of that organization. These relations are typically expressed via communication behavior, and are affected by the environment in which communication can take place. Thus, communications technology is likely to have an effect on the relations among persons, and therefore on the structure of an organization.

The computer is becoming increasingly used as an integral part of an organization's communication system. Hence, its use, and more particularly, the manner in which it is used, may well have an effect on the structure of its host organization. Since the effects are not necessarily all beneficial, nor are they well known, we argue that one ought to understand the impact that the computer may have on organizational structure, particularly via its use as part of the communications system.

We present a scheme for measuring organizational structure via an analysis of intra-organization communication, and a methodology for obtaining the desired data. We suggest that their use is applicable for studying the impact of new communications technology, such as the communications role of the computer. Since the research program is ongoing, only preliminary empirical results of the use of the methodology to describe an organization are presently available. Some of these are noted briefly. We also mention some concerns based on current insights about possible impacts of the use of the computer as an integral part of an organizational communication system.

INTRODUCTION

The introduction of new communications technology typically has been undertaken with little regard for its impact on the structure of interpersonal relations. The advances made have often been engineering marvels, but their development and use have been fostered by the insights of engineers rather than by studies of the needs of potential users. When the new technology had little direct interface with the user, such as the transistor and micro-circuitry, the impact has been essentially economic in nature. The relations among people remained the same, for their perceptions of the instruments at their disposal did not appear to vary. Recently, however, technological developments have come along which directly affects how one user relates with another. Two common examples are the computer and two-way video communications.

Initially the computer was used to perform repetitive operations that involved little interpersonal contact, the goal being increased efficiency, speed and accuracy. Over time we have come to see that the computer is capable of much more, and in particular increased use is being made of the computer in the decision process of many organizations (Jones, 1970; Morton, 1970). This cannot help but redefine roles within the organizational structure, and as a consequence influence the manner in which these roles (the persons in them) relate to each other. In another context, as the computer has become more involved as an augmentation and intercessor device in interpersonal communication (see the papers presented in the Teleconferencing Session of this Conference), a direct impact on structural relationships will be observable. While these uses of the computer are going to be further developed, little consideration is being paid to the effects, both good and bad, that these uses will have on the operating structure of organizations - the networks of interpersonal relations that make up the structures.

The Picturephone is a good example of an engineering marvel that was developed without adequate attention being paid to its use for interpersonal relationships. The assumption upon which development work began was that a need existed for a device that would approximate the intimacy of face-to-face communication. If an audio-visual device could be developed that could do this without completely disrupting an existing narrow band telecommunications network, it would have a great potential. In particular, management expected that Picturephone would provide an appropriate substitute for much of the travel currently undertaken for face-to-face interactions (Molnar, 1969). What was not perceived was that most planned face-to-face interactions between persons located at disparate points in space involve the presentation of some alpha-numeric (printed) material. Picturephone has not proven to be suitable for that purpose. Nor does it allow for the chance encounters that one gets with an on-the-premise visit. While the business firm was to be the initial market, no evidence exists that A.T.&T. adequately understood the concepts and realities of organizational communication, especially at other than the senior executive level.[1]

The introduction of a new communications technology may engender one of two types of organizational structural responses. The one usually observed is the direct response of "reorganizing" to use the new technology more effectively, to provide a place within the structure to house the technology. The introduction of the computer has brought about departments of "data processing" and "information systems".(a recent phonomenon) and other manifestations of restructuring. This is the type of structural response with which most organization theorists and practitioners have been concerned. These might be referred to as the macro-structural adaptations; they concern how groupings relate to each other.

Secondly, the prime focus of this paper, is the change induced among interpersonal relations, micro-structural adaptations. To illustrate, if three parties relate to each other about equally over the telephone, and now two of them are given Picturephones, an increased impetus exists for these two to communicate with each other more than with the third party who does not have the device. Or, in another context, let us consider the introduction of a computer based information system into a firm that previously communicated via a well define hierarchical structure. The information system is likely to provide for direct access to data that previously could only be obtained from a particular source. The ability of that source to control the recipient of the information is now greatly diminished, a significant change in the organizational structure.[2]

ORGANIZATION AND COMMUNICATION

Organizational structure, as we define it, is a network of interpersonal relations. But a single network is unlikely to encompass the complexities of the concept organization. Consequently, we view An Organization[3] (a particular organization) as a set of interpersonal networks, each of which is based on a particular characteristic used to link the persons to form a given network. The networks are not independent (orthogonal to one another in a dimensionalized space); rather, each is generally seen as unique to a purpose and each has a place when comparing one composite structure with another.

The most common view of organization is that imbedded in the organization chart - the reporting relations that exist within An Organization. These relations reflect the "authority structure", the presumed basis for the exercise of direction and control. Taking a rather liberal interpretation of the authority structure, we find that it has been the major focus of attention for organization theorists from Weber (1947) to many of the present day researchers (eg. Blau, 1970; Pugh, _et al_, 1968; Lawrence and Lorsch, 1967). An industrial engineer, on the other hand, might view organization as the set of work flow or task relations needed to meet the output goals of The Organization. The concern here is with the dynamics of structure. Others, for example, office landscapers and architects, see the structure as a set of spatial relations that exists among the persons (Pile, 1969).

Each of the above views of structure are essentially normative. They specify a structure on an _a priori_ basis, given a purpose to be accomplished and a set of persons to accomplish it. They are also part of the description of An Organization, however, for they provide for our perceptions of what The Organization is. But they cannot provide for a complete description of An Organization. For example, they exclude any consideration of the informal organization which arises somewhat independently of the formal structure (Conrath, 1968; Davis, 1953). Furthermore, while the authority structure may indicate an expected ability to exercise control via rewards and sanctions, it ignores the factors of personal abilities and coalitions which can negate formally sanctioned power. Thus, we need to know something more about An Organization to properly and meaningfully specify its structure.

Looking at each of the dimensions of structure described above we note that each has implications for intra-organizational communication. The authority structure, to be exercised, must provide for communication, for without it there is no means to give direction nor execute control. The task structure or work flow is in itself a communications network. On a production line that which is transmitted from one person to another may be a physical object, a product. In administration, however, the work flow relations are likely to involve what we more readily see as communications - messages. Office landscapping has been sold on the basis that it provides for spatial relations that encourage desired communications and discourage undesirable ones. It provides for a physical network which is designed to foster the desired interpersonal network. Given the association between normative properties of structure and communications, it would seem appropriate to consider describing An Organization from a communication behavior or network point of view.

The idea of considering organization to be reflected in communications behavior is not novel, but it is suprising that the use of communications data to study organization has not become a more common practice. To date the use of such data usually has been confined to highly controlled laboratory studies of small group behavior (Shaw, 1964; Mackenzie, 1967), or to draw correlations between communications behavior and roles (Burns, 1954), attitudes (Lawler, _et al_, 1968) and process (Elion, 1968), each a rather restricted aspect of organizational structure.[4]

While consideration of communications behavior to study organization has been quite limited, its value has been recognized for some time. Thirty four years ago Chester Barnard, who might be regarded as the father of modern organization theory, stated (1938, p. 91): "In an exhaustive theory of organization, communication would occupy a central place, because the structure, extensiveness, and scope of the organization are almost entirely determined by communication techniques". In a similar vien, Karl Deutsch noted (1952, p. 367):

> "Communication and control are the decisive processes in organizations. Communication is what makes organizations cohere; control is what regulates their behavior. If we can map the pathways by which information is communicated between different parts of an organization and by which it is applied to the behavior of the organization is relation to the outside world, we will have gone far toward understanding that organization".

Recently, Brewer succinctly described the position of the organizational sociologist Peter Blau (1971, p 479): "communication flow in organizational hierarchies is the combined result (1) of the structurally induced communication needs of managers and operating personnel and (2) of the opportunities that the organization structure provides for communication between them". These three positions provide the _raison d'etre_ as well as some of the methodology of our total research effort directed toward the study of organizational communications and organizational structure.

STRUCTURE, ITS MEASUREMENT

The measurement of organizational structure by means of communication behavior rests on three principles central to the entire research effort. First, the data must be obtainable in the field - the real world. The theories and concepts of structure to be developed are to be application oriented. If they are not based on reliable real world data, problems of interpretation and implementation too frequently appear to be insurmountable. Secondly, the format in which the data are to be considered must be sufficiently general to represent most concepts of structure that might be relevant to the study of organization. Thirdly, the data and the format must be useful for conducting both longitudinal and comparative research. We want to be able to compare one form of structure with another.

The three criteria stress the importance of field study methodology. They also focus attention on the means by which field data are used to reveal descriptive properties of structure that can be related to each other and to normative concepts of structure as well. The processes used to obtain organizational communication data will be described in a later section. We now turn our attention to a discussion of a format suitable for representing a variety of structures.

A satisfactory format, one which is relatively general, must be able to encompass such concepts of structure as those involving authority and task or work flow relations. Likewise, the format should be able to handle the spatial relations of An Organization. All three of these essentially normative structures are based on a set of dyadic relations - relations between any and possibly every two persons whom we wish to consider as elements in the structure. The authority structure indicates who reports to whom, and each reporting relation can be expressed as being between just two persons, even if one has several subordinates, or even several superiors. Likewise the task structure can be seen to be a sequence of dyadic relations, a flow from x to y, y to z, and so on. And spatial relations, such as distances, can be seen to exist between any two persons and shown as such.

Communications behavior, our basis for a descriptive model of An Organization, can also be thought of as a relation between two persons. For many communication acts this holds strictly. Others, such as a one to many transmission, can be viewed as a set of simultaneous one to one transmissions, the originator being the same in each case.

Whether all concepts of structure can be represented as a collection or set of dyadic relations is another matter. One must be concerned, for example, with such things as the representation of coalitions. But a coalition of two against a third might be treated as one particular relation between the two coalescing persons, and two identical relations between each party of the coalition and the third party. So far we have been unable to find a concept of structure that cannot be represented by a set of dyadic relations, though it may take several dimenions in which to express all the possible relations that might exist. Thus, it appears feasible to consider a particular organizational structure as a set of dyadic relations. Whether or not all structures can be so represented is an empirical question that has yet to be answered.

If we agree that a structure can be appropriately represented by a set of relations between any dyad in An Organization, the remainder of the format problem is easy. The simplest and most useful format for the display of dyadic relations would appear to be a symmetric matrix, the rows and columns of which represent all of the persons considered to be members of The Organization. Data in a matrix format are comprehensible, and a matrix is a convenient form for data manipulation and comparison. This could be demonstrated by many examples, and the sections on the measurement of impact and some early empirical results will note several.

To give an example of a specific use of the matrix format, suppose we wish to indicate that "b" reports to "a". If the rows of the matrix represent superiors and the columns represent subordinates, then the appropriate relation will be at the intersection of row "a" and column "b", r_{ab}. If the authority structure is indicated by a 0,1 measure, the 1 describing a "reporting to" relation and the 0 indicating otherwise, then $r_{ab} = 1$.

Communication behavior can also be presented in a matrix format. We might, for example, have each r_{ij} indicate the frequency of interaction between i and j. In a more complex sense we might consider the content of interactions as matrix entries, and this might require an n dimensional matrix if we have n content categories. A multi-dimensional matrix might also be needed to describe interactions between groups (as opposed to individuals).

There are no limits to the values that a given r_{ij} might assume, except that most matrix manipulations require real number entries. We must decide, therefore, precisely how we want to represent, say, an authority structure, for we have an infinite number of measures available to us. At one time we might simply use a 0,1 measure, as we have done in the proceding example. On the other hand, we might wish to express degrees of authority, or problems of dual authority, or distances in the chain of command. Each of these could be expressed by a measure suitable for their expression. At this stage of our research we have not considered even a small fraction of the possible ways to represent even a given structure. The measures we eventually use will be those that provide us with the greatest insights on the properties and problems of structure. This will be an empirical determination.

IMPACT - CHANGE, ITS MEASUREMENTS

Assuming that we use a matrix format to represent the properties of organizational structure, how might we measure impact or change? The first step would be to determine what the structure was prior to the introduction of the impetus for change. This is the purpose of obtaining the overlay of matrices describing An Organization. These can also be used to represent the structure during and after its response to some exogeneous influence, such as the introduction of a computer based communications system. If all of the representations of structure are the same, apparently no change has taken place. If some of them differ, two aspects of impact must be considered: the extent of change and the nature of the change. The first aspect is likely to have a quantitative measure, the second is likely to be qualitative.

Extent of change measurement is to answer the question, how great is the deviation in The Organization's structure? Unfortunately, since organizational structure is not an unambiguous concept, no single measurement scheme is likely to be satisfactory. Several dimensions will need to be considered. Also, a given value will have meaning only in the context of a set of values likely to be obtained by the measurement technique applied to a large number of change situations. No measurement value can meaningfully stand alone. To tell you that my house has ten rooms means something only when you have some idea how many rooms most houses have. Furthermore, you would probably want to know how I define a room and what the approximate sizes of the rooms are. Even then the meaning of ten rooms is likely to depend upon whether I speak as a North American or as an African. What we must do, therefore, is build a data base to provide a context for the measurement of impact on organizational structure. This should be done in light of some of the ways in which we might consider extent of change.

One simple way to measure change in organizational structure is to consider each dyadic relation as independent. If we do so, we might sum the absolute differences between all of the r_{ij} entries, $\sum_{ij} \left| r_{ij}^{a} - r_{ij}^{b} \right|$ and call that our measurement of change. If the entries were normalized for each matrix, the maximum amount of change could be 2, and the minimum, of course, 0. If we obtained a number of data points using such a measure, we might get some idea as to the implication of various values along the 0,2 scale.

The sum of absolute differences could also be applied to structures represented by 0,1 entries, possible values ranging from 0 to n·(n-1). A ratio of the sum of the absolute differences to n·(n-1) would give us the proportion of all possibly dyadic relations that changed.

Both of the above measurement techniques assume relatively independent rij, but such independence might not hold. Rather, independence might relate to the individuals who are members of The Organization. Thus, we might wish to calculate the above sums, but only for each individual, and then consider each individual as a single data point in our measurement of organizational change. For example, it may be useful to know that 17 out of 116 persons had their role in the organizational structure change. Also, we might be interested in the change of an individual communication profile. For example, the new equipment led to a more even distribution of outgoing communications for Mr. X.

Many other measurements of impact could be calculated. One final example is based on a concept being developed which we call "relational clustering" (Conrath & Johnson, 1972). Basically, it involves grouping or clustering together those persons who have a high level of interaction (or other relation values) among themselves relative to their levels of interaction with others. We can obtain the sizes of these clusters, who is clustered with whom, how many times we have to cluster the clusters to obtain the entire organization, etc. Each of these measurements provides a basis for comparing The Organization before and after the impact of an exogenous force. For instance, we might find that the introduction of a comprehensive information system reduces the average cluster size, interpersonal dependence now being replaced by centralized processing capabilities.

AN EMPIRICAL STUDY, DATA COLLECTION AND EARLY RESULTS

Four primary objectives guided the first empirical study. First, the data collection instruments were to be tested. The intent was to discover how much and what kind of organizational communication data could be gotten without disrupting An Organization's activities. Secondly, the data was to be processed in a variety of ways to establish the types of insights that might be obtained about different properties of organizational structure. Thirdly, several simple conjectures were set forth about the relationship between a priori structures and those based on communication behavior. These were to be tested. Fourthly, a before and after study was to be undertaken to investigate the impact of office landscaping on organizational communication. Unfortunately, the administration of The Organization researched went through several major changes and the cooperation necessary to conduct the follow-up data collection disappeared.

Since reliable data about specific communication events were desired, both personal experience and that of others discouraged us from using recollections about past events. Hence, the data was to be recorded during or immediately after an interaction, and a method had to be devised for doing so. Several researchers (e.g. Lawler, et al, 1968; Wickesberg, 1968) have found a simple self-recording diary satisfactory. A particular use of this approach was developed by the Quickborner Team of Germany (Lorenzen, 1969), the first active proponents of office landscaping. They viewed the ideal office layout as one which fostered desirable interaction and inhibited undesirable communication. Therefore, they wished to determine existing communication patterns to understand current needs. Unfortunately they appear never to have conducted data collection ex post office landscaping.

A modification of the Quickborner "communications tally sheet" was used for the first field study. The actual form used appears in the Appendix. A detailed instruction sheet accompanied the form, and in addition we found that a face-to-face presentation of the instructions along with the reasons for the survey greatly enhanced data reliability. One can see from the Appendix that communications data were obtained on a from/to basis, and it included the mode of communication used (face-to-face, telephone, written), a rough estimation of the elapsed time involved, and the number of parties to the interaction. Initially we proposed to get data on communication content and a more accurate estimate of the elapsed time of a communication event. Pilot studies indicated, however, that anything more elaborate than the form finally used would reduce the likelihood of participation significantly, and those who continued to participate would supply less reliable data.

The existing sample of organizational communication data was obtained from approximately 350 managerial and senior staff people (all those who were not absent) from a division of a large Canadian manufacturing and sales firm. Each significant interaction (one involving more than just a personal greeting or the like) was to be recorded by the receiver of the communication[5] over a five-day period (nine days for some). While a longer period would have been preferred, as would data collected at random points in time, the pragmatics of data collection (we did not wish to strain our relations with our subjects any further) prevented us from these approaches.

The processing of the data bank is an ongoing activity. Since many different manipulations are to be tried, and since the data base is quite large (some 15 to 20 thousand events were recorded), most manipulations have to be done by computer. The writing of computer programs have proven to be more onerous than expected (which is typically the case), and several are just now nearing completion. In the meantime, some data analyses were done by hand on a sub-sample of 30 persons at one plant location. These reflect on the three conjectures postulated as a part of the field study methodology (we were not testing theories, but rather were looking for insights into structure that were interdependent with the data to be collected).

The first conjecture was: among various modes of communication, the pattern of written communication will most closely parallel the formal authority structure. We expected this to hold because of the use of written material to provide for both direction and control since it is hard copy, and therefore capable of being retained. Using both 0,1 and normalized measures of the authority structure and communications frequency patterns (for further information see Conrath, forthcoming) the conjecture held true.

The second conjecture was: among various modes of communication, the pattern of telephone communication will most closely parallel the task or work flow structure. The real-time feedback properties of the telephone system appear to be well suited to many task needs. Where face-to-face contact is costly from either an effort or economic standpoint, we expected that the telephone would be used. This conjecture also found support using both 0,1 and normalized measures of structure[6].

The third conjecture was the most obvious of all: among various modes of communication, the face-to-face mode will be most influenced by physical proximity (Barnlund & Harland, 1963). This was varified by our data, and at a high level of statistical significance.

? COMMUNICATION USES OF THE COMPUTER, IMPACT ?

We cannot state measured results of the impact of computer based communications systems on organizational structure. Furthermore, attempts to predict the impact cannot be based on such a simple logic as that used for the previous conjectures. For example, while the computer normally produces hard copy, a desirable attribute from an authority structure standpoint, the copy is seldom in a format suitable for filing in the standard letter sized filing system. While separate filing systems can be created, this adds to the cost of the use of such hard copy. Furthermore, the computer can provide the filing system internally, so there is less need for hard copy. But now the problem is one of search and retrieval, and these problems increase with an increase in the size of the files.

The computer is also capable of real-time feedback, but usually on a man-machine rather than a man-man basis. While the computer has been task oriented in its use, it has been used primarily on intra-individual rather than inter-individual tasks. Thus, the initial use of the computer has probably had little effect on existing organizational communications patterns. Nevertheless, its potential to bypass the more complex and indirect aspects of existing task patterns must be recognized. The computer, rather than reinforcing a particular dimension of structure may instread restructure the organization, especially from a communications standpoint, around its capabilities.

The organizational impact is a changing one for the capabilities of computer based systems are continuing to increase. A study of its impact on organizational structure, therefore, must be longitudinal. Furthermore, since the potential impact of the computer on the structure absorbing it may be dramatic, considerable research appears warrented if we are to predict and possibly control the path of organizational change which takes place.

It is not enough to say that the organizational impact of the computer will be a new data processing division, or a computer systems division, or what have you. That merely states where the control of the hardware lies. The important impact of the computer lies in the development of software that will allow it to assist and augment interpersonal interactions. How this will affect existing structural properties remains to be seen. At the one extreme the computer may do no more than reinforce the existing relations. At the other, the only structure that might exist is every man relating to the computer, no man relating directly to another. Since little has been done to research the organizational structure impact of such significant technologies, all we can do now is construct conjectures based on intuition. Hopefully, future organizational research will allow us to rely less on guesses and more on documented facts.

Footnotes

* Financial support from the Canada Council and Bell-Northern Research is gratefully acknowledged. Thanks are also due Gordon Johnson of Bell-Northern Research for his many contributions.

1. The market trials all involved senior executives, yet any large scale use obviously has to involve persons at lower levels in the organizational hierarchy. Virtually no research was undertaken to determine the implications of usage at these levels. At entire issue of the Bell Laboratories Record, May/June 1969, was devoted to the Picturephone. Only one article discussed in any detail the use of the Picturephone in a business context (Harris & Williams, 1969), and the bulk of that article had to do with the incorporation of the Picturephone into the firm's local communication network. The discussion assumed that the need for the Picturephone had been identified and understood and that the current design (Mark II) was satisfactory for that need. Organizational communication was viewed as a technical rather than as an interpersonal relations problem.

2. A more elaborate explanation of what might happen can be found in Chadler, et al, 1972.

3. To distinguish between the concept "organization" and a particular Organization which we are attempting to describe, we will capitalize the latter. Thus, when we write An Organization or The Organization, we are speaking of an entity such as General Motors, or Northern Electric, or a provincial Department of Education, or what have you.

4. For a review of organizational communication studies to 1965, see Guetzkow (1965).

5. This was done to reduce the work load of the subjects so that we might get maximum cooperation, and to reduce the bias that might be introduced if one had to record his own initiations. Hence, we are not able to verify data reliability (though personal observation indicated that most subjects were conscientious in their recording of interactions). Hopefully the process of data collection will be sufficiently improved in the future to obtain cross checks on each communication event without alienating our subjects.

6. Both of the above conjectures have also been tested on a larger sample, n = 116, and the results were highly significant. Recently, however, we have realized that the level of significance may be influenced by an artifact of the data base which appears to have biased the results in the direction of the conjecture.

BIBLIOGRAPHY

1. BARNARD, Chester I., The Functions of the Executive, Cambridge Mass.: Harvard Press, 1938.

2. BARNLUND, Dean C. and HARLAND, Carroll, "Propinquity and Prestige as Determinants of Communication Networks," Sociometry, 26 (1963) pp. 467 - 479.

3. BLAU, Peter M., "A Formal Theory of Differentiation in Organization," American Sociological Review, 35 (1970), pp. 201 - 218.

4. BREWER, John, "Flow of Communications, Expert Qualifications and Organizational Authority Structures," American Sociological Review, 36 (1971) pp. 475 - 484.

5. BURNS, Tom, "The Directions of Activity and Communication in a Departmental Executive Group," Human Relations, 7 (1954), pp. 73 - 97.

6. CHADLER, Ernest; DRUMMER, David; and NADER, Peter, "Integrated Information Systems - Dream or Reality?", Telesis, 2/5, (1972), pp. 16 - 22.

7. CONRATH, David W., "The Role of the Informal Organization in Decision Making on Research and Development," IEEE Transactions on Engineering Management, EM-15 (1968), pp. 109 - 119.

8. CONRATH, David W., "Communication Patterns, Organizational Structure and Man: Some Relationships," Human Factors, forthcoming.

9. CONRATH, David W. and JOHNSON, Gordon A., "Analyzing an Organization Via its Internal Communication Patterns," Paper presented to the XIX International Meeting of the Institute of Management Sciencies, Houston, Texas, April, 1972.

10. DAVIS, Keith, "Management Communication and the Grapevine," Harvard Business Review, 31, (1953), pp. 43 - 49.

11. DEUTSCH, Karl W,, "On Communication Models in the Social Sciences," Public Opinion Quarterly, 16 (1952), pp. 356 - 380.

12. EILON, Samuel, "Taxonomy of Communication", Administrative Science Quarterly, 13 (1968) pp. 266 - 288

13. GUETZKOW, Harold, "Communications in Organizations," in March (1965), pp. 534 - 573.'

14. HARRIS, James R., and WILLIAMS, Robert D., "Video Service for Business," Bell Laboratories Record, 47 (May/June 1969), pp. 149 - 153.

15. JONES, Curtis H., "At Last: Real Computer Power for Decision Makers," Harvard Business Review, 48 (Sept/Oct 1970), pp. 75 - 89.

16. LAWLER, Edward E. III; PORTER, Lyman W.; and TENNENBAUM, Allen, "Managers' Attitudes Toward Interaction Episodes," Journal of Applied Psychology, 52 (1968), pp. 432 - 439.

17. LAWRENCE, Paul R.; and LORSCH, Jay W., Organization and Environment: Managing Differentiation and Integration, Boston: Division of Research, Harvard University Graduate School of Business Administration, 1967.

18. MACKENZIE, Kenneth D., "Decomposition of Communication Networks," Journal of Mathematical Psychology, 4 (1967), pp. 162 - 174.

19. MARCH, James G., (ed.), Handbook of Organizations, Chicago: Rand McNally, 1965.

20. MOLNAR, Julius P., "Picturephone Service - A New Way of Communicating," Bell Laboratories Record, 47 (May/June 1969), pp. 134 - 135.

21 MORTON, Michael S., Management Decision Systems: Computer Based Support for Decision Making, Boston: Division of Research, Harvard University Graduate School of Business Administration, 1970.

22. PILE, John F., "The Nature of Office Landscaping," AIA Journal, July 1969.

23. PUGH, D.S.; HICKSON, D.J.; HININGS, C.R.; and TURNER, C. "Dimensions of Organization Structure," Administrative Science Quarterly, 13 (1968), pp. 65 - 105.

24. SHAW, Marvin E., "Communication Networks," in Berkowitz, Leonard (ed.) Advances in Experimental Social Psychology, New York: Academic Press, 1964, pp. 111 - 147.

25. WEBER, Max, The Theory of Social and Economic Organization, New York: Free Press, 1947.

26. WICKESBERG, A.K. "Communication Networks in the Business Organization Structure," Academy of Management Journal, II (1968), pp. 253 - 262.

APPENDIX

Communications Tally Sheet

							NAME					DATE		
IDENTIFY PERSON FROM WHOM YOU ARE RECEIVING COMMUNICATIONS	TELEPHONE CALLS			PERSONAL (FACE TO FACE) VISITORS							PAPER			
				1 VISITOR			2 OR 3 VISITORS			4 OR MORE VISITORS	PERSONALLY ADDRESSED	ALL OTHER MATTER		
	SHORTER THAN 3 MINUTES	3 TO 15 MINUTES	LONGER THAN 15 MINUTES	SHORTER THAN 3 MINUTES	3 TO 15 MINUTES	LONGER THAN 15 MINUTES	SHORTER THAN 3 MINUTES	3 TO 15 MINUTES	LONGER THAN 15 MINUTES			SHORTER THAN 3 MINUTES	3 TO 15 MINUTES	LONGER THAN 15 MINUTES
DISTRIBUTION SALES DIV.														
SWITCHING DIVISION														
TRANSMISSION DIVISION														
WIRE & CABLE DIVISION														
R & D DIVISION														
CORPORATE (ALL OTHERS N.E.)														
BELL CANADA														
OTHER CUSTOMERS/SUPPLIERS														
OTHERS (NOT N.E.)														

THE IMPACT OF LACES

N.D. Hill and T.F. Watling
International Computers Limited
London, England

Abstract

The London Airport Cargo Edp Scheme - LACES - has been in regular use every day since August 1971. The paper gives an outline history of LACES and the reasons why it was developed, and also describes the way in which the project was organised. The system is dedicated to providing customs clearance for goods imported through London Heathrow airport and to providing an inventory control system for these goods. Part of the interest of the LACES project lies in the large number of independent parties involved in the implementation and use of the system. The paper examines some of the wider implications of LACES and its impact upon the environment at Heathrow Airport.

The authors are both members of the Designated Systems Organisation of ICL.

LACES - the London Airport Cargo Edp Scheme - is a system dedicated to providing customs clearance for goods imported through London Heathrow airport and also providing an inventory control system for these goods. This is not a blue skies project. The system has been implemented and has been working successfully since August 1971. Since then all cargo imported through Heathrow has been handled by LACES. In due course LACES will be extended to handle exports.

This paper will first briefly outline the history of the advanced on line real time system which is LACES, describe the organisation of the project to bring it into existence and outline the system. Then we will examine the wider impact of LACES. Users become hooked onto the system in more senses than one. As time goes by they become vulnerable to interruption of service from whatever cause. They become tied to a system which is itself a vehicle of change; a system which has an impact on competition and on the entry of new firms of agents into the business of expediting cargo through the airport. Considerable power is placed in the hands of those who control dedicated and exclusive computer systems. This aspect warrants careful consideration before commitment to such a system is given. For LACES the problem was solved, so far as it can be solved, by putting the control and operation of the system into the hands of a government agency - the United Kingdom National Data Processing Service, which is not itself a user of the system.

Our own involvement in the system arose because ICL was the main contractor for computer equipment - hardware, terminals and software - for LACES. Within ICL the responsibility for the successful implementation of the project was shouldered by the Designated Systems Organisation to which we both belong.

In 1966 London's Heathrow airport handled more international air cargo by value than any other airport in the world. People concerned with moving freight through the airport were aware of a range of problems and the indications were that these problems would in time become worse.

While it was possible to fly cargo into Heathrow from most major airports in Western Europe in less than two hours, it could take several days to clear the goods through cargo sheds and customs. Some consignments were immobilized for days or even weeks. Normal handling of air cargo involved a great deal of paper work.

The air cargo delays even in 1966 had a number of unfortunate consequences. Although the volume and value of air cargo being handled through Heathrow was steadily increasing air handling of cargo was not as competitive against surface shipment as it should have been, particularly for short journeys to and from continental Europe. The indirect costs of transportation were expected to increase as a result of delays because of the huge value of cargo in transit at any moment. For someone, whether consignor or consignee, this represents "work in progress". The value of cargo handled through London airport in 1966 amounted to £865 million. If cargo on average was only delayed unnecessarily by a day the cost of this at the borrowing rate of 8% pa current in the United Kingdom in 1966 might be estimated at £188,880 a year. Apart from this every extra day that cargo remained on the airport represented an extra day of exposure to the risk of theft.

Shed space on the airport, as on all airports close to large markets, was expensive. Delays in cargo handling increased the amount of shed space needed. Delays meant unhappy customers and a reduction in staff morale. Both these factors called for a higher level of staffing than would have been necessary with a shorter stay for cargo on the airport.

On top of the existing unsatisfactory situation, the forecast was that traffic handled would increase by 14% a year, i.e. to double every five years or so. This was expected to compound the delays and ensuing problems. It would also directly increase the paper work load and the problems of customs clearance.

In the face of this situation, the foreign airlines operating into Heathrow, decided in the summer of 1966 to form an Edp Cargo Working Group to study the feasibility of creating an automated and integrated import cargo clearance system. It is perhaps of interest that of the many parties concerned with the movement of cargo through the airport it was the foreign airlines who took the initiative to solve the problem in a co-operative fashion. They, perhaps, appreciated more clearly than the other parties involved the extent of the problem and the need for co-operation.

Once the initiative had been taken, the project began to gather momentum. In September 1966 the foreign airlines agreed on the principles which were to govern the study. Representatives of the two British national airlines, BEA and BOAC joined the working group. H. M. Customs and Excise endorsed the principles, which had been agreed. In 1967, the Institute of Shipping and Forwarding Agents (now renamed the Institute of Freight Forwarders Ltd.) started to participate in the activity of the working group. By no means all of the 150 or so cargo agents operating at the airport were members of the Institute All agents participating in the scheme would of course have to pay for it. This latter day attempt to obtain payment without representation may have contributed to the subsequent problems in settling the user contracts. On the other hand, it would not have been practicable to deal with each agent individually.

In June 1967 the Airlines Edp Cargo working group issued its report on a proposed computer based system to operate inventory control and customs clearance procedures for goods imported into the United Kingdom through Heathrow. The system later became known as the London Airport Cargo Edp Scheme or LACES for short.

At this stage it was decided that someone was needed to carry the scheme through the various stages of implementation and then to operate it. The parties involved at the time were the 16 international airlines, including BEA and BOAC, the 150 or so freight agents and Her Majesty's Customs and Excise. A further party bound to be involved at some stage was the British Airports Authority, which provides the airport facilities.

Eventually it was decided that the task should be given to none of the parties directly involved. An invitation was extended to the National Data Processing Service (NDPS) to design and run a system to serve all the interested parties. The NDPS is a branch of the British Post Office. It was thought to be both sufficiently neutral and to have sufficient experience in the design, implementation and operation of large computer projects for it to be a suitable choice. NDPS accepted the invitation and started work on planning the project. At this stage LACES had been described in a single volume feasibility study of 114 pages.

The policy direction of the project was the responsibility of the LACES Steering Group whose chairman was the director of NDPS. The secretary of the Steering Group was the NDPS Project Manager for LACES, Mr. A. N. James. The other members of the Group were provided by the users. The airlines Edp Cargo Steering Committee had 3 members, including M. Etienne Dreyfous of Air France who had been a driving force behind the idea of LACES since its inception. The Institute of Shipping and Forwarding Agents, had two representatives and H. M. Customs and Excise one.

The Steering Group retained for itself the right to approve the LACES system and all changes to it, the right to approve all financial arrangements and the right to agree the contractual basis of the arrangements between NDPS and the users. All dealings between NDPS and their subcontractors were specifically excluded from the terms of reference of the Steering Group.

NDPS pressed forward work on four fronts; the design of the system; the selection of a building site; making arrangements for the construction of the computer centre; and the selection of a computer supplier.

Design of the system was undertaken by a combined team drawn from NDPS, H. M. Customs and Excise and the airlines.

A site was found off the airport and about 5 miles away from the airport cargo terminal. This had the advantage of being close to the airport, whilst being outside the jurisdiction of the airport's authority.

At this time the Post Office was a department of the British Government and was bound by a number of government rulings. One of the these was that all building for the Post Office must be executed under the direction of the Ministry of Public Building & Works (later absorbed into the Department of Environment). Arrangements were made with the Ministry to put in hand the contruction of a building, based on a standard design, on the selected site.

The consideration of where to place the order for the computer hardware, software and applications programs was a protracted one. Eventually on 30th June 1969, NDPS invited ICL to be the main contractor for all the computer hardware and software for LACES. ICL provided two 4-72 computers with fast drums, multiple spindle exchangeable disc stores and magnetic tapes as well as the usual input/output peripherals. The writing of a special LACES oriented executive program and the application programs was subcontracted from ICL to Computer Sciences International (UK) Ltd. The manufacture of video display units (VDUs) and controllers was subcontracted to Cossors. A number of other smaller subcontracts were let, for instance for the laying of cables to connect VDUs with controllers.

ICL specialises in managing the implementation of complex computer projects for users of ICL equipment. A special project management unit called the Designated Systems Organisation, provide project managers and staff for such projects. In this case a very experienced project team was rapidly assembled.

As in all large and complex projects, a balanced project management team was needed. It had to provide the main skills needed whilst being kept to a manageable size. The team consisted of a project manager, responsible for ensuring that ICL and its subcontractors met their commitments to the project, a programming manager responsible for monitoring the programming subcontracts, a communications manager responsible for monitoring progress of the VDU subcontractor and for setting up the communications network, a systems analyst concerned with change control, a progress manager responsible for the PERT analysis and progressing of ICL divisions and an operation manager responsible for operations support and advice for the project including the provision of a test cell.

In addition to the project team, ICL appointed a manager within the manufacturing organisation to supervise the engineering production and installation work. A site maintenance engineering team was provided for both the computer centre and the communications network. Dataskil, the ICL Software house, provided operating assistance to NDPS in the early months.

Within ICL a management committee reviewed the progress of the project every four weeks. Links between the ICL project team and the project teams of NDPS and

CSI(UK) were of course close. Apart from regular formal meetings, there were frequent informal and casual meetings between the people involved and their opposite numbers in NDPS and in the subcontractors' teams.

The subcontracts were on a fixed price basis. Control was exercised by monitoring progress against agreed PERT networks and baselines. Strict change control procedures were observed.

Any cargo consignment other than mail and diplomatic bags arriving at London airport (whether for import or trans-shipment) is covered by an air waybill, issued by the originating airline at the airport of departure. This document can usually be uniquely identified by the air waybill number, but where two or more air waybills have the same air waybill number, they are differentiated by the air waybill prefix. Each airline uses its own unique prefix.

The air waybill is passed to the relevant airline office in the cargo terminals, where data from the document is transmitted via the office's visual display to the LACES computer centre. The air waybill number, together with the air waybill prefix, is the basic reference for data held by the system and forms the key to one of the two principal files, the common basic reference (CBR) file.

When a consignment of cargo arrives at the cargo shed, an operator keys into a visual display unit the air waybill number, the number of packages received and the storage location of each package unloaded into the shed. The CBR record of expected goods created by the airline office is then updated by the computer system accordingly. (Should the airline office not have received the air waybill, the shed operator's entry creates a skeleton CBR record, which is completed when the air waybill is eventually received, and its contents transmitted to the computer system.)

The airline gives a copy of the air waybill to the forwarding agent responsible for receiving the cargo.

After assembling all documents relating to a consignment, the agent completes a coded customs entry. Using his visual display unit he keys data from the entry into the computer. By doing so, the agent is declaring goods to customs. The computer immediately calculates necessary revenues such as duty and purchase tax payable (charging these to the agent's account) and with a confidential customs program, selects the type of inspection, if any, to which the consignment should be submitted. Details of revenues and customs inspection required, together with the time entry was made, are shown on the visual display screen. The agent copies them onto the coded customs entry form. A record of the entry details is recorded on the computer system's other main file, the customs entry record (ER) file.

The agent can, if he wishes, make an information customs entry via his visual display. He may do this in order to establish the potential customs duty or purchase tax on a consignment. The response to an information entry, however, will not say which type of inspection will be adopted for the consignment. This will be displayed on the screen only when a formal entry is made, or when an information entry is converted to a formal entry. Conversion to a formal entry can be achieved by depressing a single key of the visual display.

The kind of customs inspection selected for the agent's consignment is shown on the visual display screen by a channel number. Channel 1 means that documents only are required for inspection. Channel 2 tells the agent that both documents and goods are required for inspection. Channel 3 means that the goods have provisionally been given automatic release, so that neither goods nor documents need be inspected.

In effect the computer gives automatic clearance after a certain period of time has elapsed. This period of time may be varied by H. M. Customs. During normal working hours it is set at one hour. At other times the "clock" may be set for a longer period.

Having written details of charges payable, clearance channel and time of entry, on his entry coding document the agent certifies that the entry particulars have been correctly input to the computer. He then sends the entry, together with the relevant supporting documents, to the customs office.

Entries must be presented to customs within the automatic clearance period. That is if the clock is set for one hour then entries must be presented within one hour of keying in entry details to the computer. After receiving an entry, customs can change the computer-selected channel number. They can, for instance, decide to over rule the computer program and re-assign a channel 3 consignment to channel 1 or channel 2 inspection. In other words they can decide to inspect the documents or both documents and goods of a consignment which the computer system had selected for automatic clearance. If the "clock" period for channel 3 clearance expires without customs intervention, the goods are automatically given customs clearance.

Customs are at liberty to carry out a channel 2 inspection in addition to, or instead of, a channel 1 examination. In this case the documents are sent to the customs office in the cargo shed, where they are inspected together with the associated goods.

All entries must be cleared by customs before goods can be released. But the computer will not issue a note authorising delivery until all packages covered by the CBR record have arrived, and the airline is satisfied that it will be paid by the forwarding agent for transport costs. To avoid unnecessary delays in releasing goods, a CBR record can be split into two or more parts. This may be done for instance in the case of house air waybills, when not all the goods covered by an air waybill are for one consignee, or in the event of an agent wishing to have a consignment split for entry or delivery purposes. Before a consignment can be released by the computer system three release conditions must be met.

The first release condition is met when the cargo operator has entered details of the final package of a consignment to arrive at the cargo shed. The second is met when the computer confirms that the airline is satisfied with the forwarding agent's credit. The third condition is satisfied when Customs give clearance to an entry.

When the computer is satisfied that all three conditions have been met a release note authorising removal of the goods, is printed by a terminal situated in the cargo shed.

Agents, the airlines and H. M. Customs, can interrogate the computer system at any time via their visual display units. Agents want information on the progress of goods and documentation. An agent might, for instance, wish to know the status of a consignment whose release authority has been delayed, so that he can remedy the delay, or explain it to the consignee.

An important aspect of the system is its security. The system is involved in the collection of £1 million a day of the Queen's revenue. The detailed customs procedures used are highly confidential. Information of a commercially secret nature is confided to the system by both airlines and agents. The basic hardware of the security system is the badge and badge reader. The badge itself is a piece of plastic 4½" x 3½", with 24 metal coding strips buried inside the plastic. The badge reader appears to the user as a slot in the keyboard. This contains sensing devices which read the badge code and associate it with every message sent from the VDU. The computer holds a file of badge numbers containing a list of all valid badges in use showing the organisation to which they have been issued. In the case of badges issued to H. M. Customs it also shows the security level of the badge.

Each time the computer receives a message from a VDU, it reads the badge number and checks that the message can validly be sent by that badge holder. To take an obvious example a customs clearance cannot be entered using an agents badge. Similarly the only person, other than a customs officer, who will be given access to the details of a customs declaration is the badge holder, who made the declaration in the first place. The customs entry itself can only be made from a Visual Display Unit designated for the transmission of messages from specified paying agents using the badge issued to that agent.

The real time system is only part of LACES. A considerable amount of batch work is done on a second processor to support and supplement the real time system. This work includes an extensive suite of programs, which checks the message log kept on magnetic tape by the real time system. This ensures the internal consistency of the computer records and ensures that there is a secure position, from which to restart the system in the event of catastrophic failure. There are also programs to handle H. M. Customs accounting and statistical requirements.

Each agent maintains a standing deposit with H. M. Customs. Agents' accounts are debited each day with total duty, purchase tax and other revenues payable. Statements of account are produced by the computer and submitted to the agents on the morning of the following day. Payment must be made to customs by the afternoon of that day. Agents are themselves responsible for obtaining reimbursement from their consignees. If the duty due on a consignment is more than the amount of the agents standing deposit then the real time system will not give full customs clearance until the duty is paid. However the size of the standing deposit is calculated so that in the majority of cases it is large enough to cover any day's transactions.

The system also produces various reports, including accounting outputs, system statistics, consignment analyses, management statistics, and record analyses. Reports fall into daily and monthly classes. Daily reports are produced not only for accounting purposes, but also on traffic analyses. They are sent to the relevant airlines and agents.

Management statistics are provided every four weeks. They analyse the number and size of consignment handled by a shed operator, and give detailed information on the operating performance of visual displays. The reports are of principal interest to the airlines, but a summary of visual display performance is also sent to agents. There are facilities to call off very detailed statistics indeed, should the need arise. For instance a detailed analysis of messages by badge number can be produced. However, the actual statistics produced are more limited.

It is interesting to note the way in which LACES was brought into existence. The need for improvement was seen by many parties. If LACES could be brought to successful fruition there were undoubted advantages to be gained. One might perhaps have expected the airports authority to see the need and provide the service. So far as we know there is no evidence that the authority either saw the need or was in any way interested in meeting it. The cynical might wonder whether the authority's interests were in conflict with the users. The greater the need for bonded warehousing on the airport, the greater the income that the authority would draw - so long as the throughput of cargo was not restricted by poor service.

The real drive to initiate the scheme came from the foreign airlines with support from H. M. Customs. The British national airlines were originally far from keen to participate, no doubt because they hoped at that time to solve their cargo handling problems by a joint computerised and automated cargo handling system. Although H. M. Customs were keen supporters of the initial idea of LACES, they did not wish to put themselves in the position of providing a service to the users. They felt there might be a conflict with their revenue raising functions. The cargo agents were at that time far too loosely organised and felt too competitive with each other to provide a driving force.

One of the consequences of LACES has been to draw the firms of agents much more closely together, largely because they felt the need to present a united front to NDPS in their contractual and financial negotiations.

The foreign airlines, operating a cargo service at Heathrow had the most to lose from the continuation of the existing situation and were best able to see clearly the benefits that would accrue to them from the proposed system. In the event although they have clearly gained the benefits foreseen, it is at least arguable that the main beneficiaries will in the long run prove to be the British national corporations and the cargo agents. One of the principal beneficiaries may prove to be the airport authority because of the increased usage of the airport for movement of freight into Europe from across the Atlantic reflecting the improvement in service at the airport.

The method of moving the scheme forward to successful implementation is interesting. Many of the users were experienced in the implementation of computer systems. Several of the airlines and H. M. Customs had their own successful large scale computer systems in operation. Yet for good reasons, as already mentioned, it was decided to ask an independent concern to implement the system. It might have been possible to go to a computer utility in the UK or overseas and ask them to create and operate the system. On balance it seemed better to go to a government backed concern, NDPS. Apart from being a subsidiary of a nationalised industry, they were one of the largest and most successful users of computers in Europe. NDPS were however only prepared to implement and operate the system as agents for the users. They were not prepared to go into LACES as commercial operators, charging a rental based on the risks involved. This led NDPS into serious difficulties when the system was almost ready to go live. Contractual arrangements with the users, despite the existence of the Steering Group were fairly loose and rested on a number of understandings. Many of the users did not really expect LACES to go live on time. After all they had their own experience of implementing computer projects and the whole United Kingdom press had been virtually unanimous in forecasting a disaster - or at the least a very delayed implementation. In practise the first user, KLM, came onto the system live on 23rd August 1971 - twenty three days later than the date set when the order for the computer system wasplaced on 30th June 1969. Furthermore all the users were on the live system by 27th September 1971, a truly remarkable rate of take on.

The principal reason for the successful implementation probably lies in the careful planning and systems work, which went into LACES. This was primarily the responsibility of NDPS, but they received significant assistance from the airlines and in particular from H. M. Customs. The Customs procedures which are followed at Heathrow, are hallowed by antiquity. H. M. Customs and Excise have been in existence for three centuries and although airports are a comparatively recent innovation, the customs procedures are derived from seaport procedures. This made the contribution from Customs staff invaluable. Where there are deficiencies in the system from the user viewpoint, and it is inevitable that there are some, these probably arise in the areas where there was least user contribution to the systems work. This, perhaps, underlines one of the prime rules for large systems. The actual users of the system must be intimately connected with the detailed design of the system. Every system differs in detail, particularly in the way the working level people short circuit it to provide better service. Only if these people are involved in detailed design work will the computerised system be better than the system it replaces. LACES is undoubtedly an improvement on what went before. It might provide an even better service had the agents been involved at a more detailed level in the systems design.

Linked with the sound systems work, a very tight control was exercised over the system changes. A firm systems base line was agreed by November 1969. This took the form of a printed document, running to nearly 2,500 pages. Thereafter, every proposed change was evaluated for its effect on system performance, cost and delivery date. No change was implemented until this evaluation had been completed. Some changes are inevitable and some 700 changes were approved from November 1969 to the time the system was handed over to NDPS for acceptance testing on 1st July 1971. Naturally many changes have been proposed since then. Some as a result of experience of live running and some due to changes in the environment in which the system operates. An example of the latter is the decision to introduce value added tax into the United Kingdom in 1973, which means major change to the system.

LACES is of course addictive. Because it provides an improved service and a reduction in paper work for all concerned with cargo handling through the airport, H. M. Customs have withdrawn some of the manual quick clearance facilities previously available. This represents no drawback to the users - so long as they remain on the system. It is probably true to say that any airline or agent, who withdraws from the system now, could only offer a degraded service by comparison with their competitors.

The operation of the system must also have an effect on the creation of new firms of cargo agents. The minimum cost of an agent's terminal on the system is about £6,000 a year. Many agents of course band together to share and so reduce the cost. However, sharing arrangements may be easier for firms already in operation than it would be for newcomers. Cargo agency has been a field in the past, where it has been relatively easy for new firms to be set up.

It has been possible for one enterprising man in the employment of an airline or large cargo agent to break away and set up a small new firm with virtually no capital. In future this will be just that much more difficult.

A feature of the system is that it is possible to connect remote terminals to it. The system provides facilities for importers, acting on their own behalf, to become parties to LACES. No importer has yet directly availed himself of this opportunity. However one large continental European firm has had discussions with NDPS. If these, or similar discussions come to fruition, the importer will have a visual display unit on his own premises and input directly into the system his own airway bill information and also his own customs entries. This facility can be particularly important for non United Kingdom manufacturers, who wish to be able to hold a base supply of spare parts outside the UK, yet be able to provide a swift spare part service to their UK subsidiaries and customers. Unfortunately with current costs (including line costs) the facility is only economic for firms who move a considerable volume of goods through the airport. Apart from this special importer facility, some firms have their own tied agency, whose sole purpose is to effect rapid clearance of their goods through the airport. Examples of this are Hewlett Packard and the British Broadcasting Corporation (BBC).

A significant proportion of the freight agency work at Heathrow airport is done directly by the airlines without intervention by a freight agent. The efficient working of the inventory control part of LACES may make it easier for airlines to handle agency work than it was in the past.

It will be interesting to see in five years or so how the economic balance at Heathrow has altered as a result of LACES. Some part of the justification for using an agent to clear goods through the airport has been the need to overcome the delays and frustrations to which the old system was liable. LACES undoubtedly provides significant benefits to agents. They can clear goods more quickly; they are able to provide a far better service to their customers as a result of the enquiry service; and they can minimise the size of their standing deposit. This last is of considerable importance, particularly to small firms for whom it may represent their major capital commitment. However, the ability of large firms to handle their own importation work from their own shipping office together with the possibility of more direct competition from the airlines may force the agents into competing more heavily in the other - off airfield - service which they offer to their customers.

One consequence of a centralised system of this sort is clearly that it is vulnerable to failure. The software is robust. The hardware is reliable and adequate redundancy of both central processor and peripherals is provided. The terminal network has been devised so that there is also a considerable degree of route redundancy. Power supplies come into the computer centre from two separate sources, though eventually both are supplied from the National Grid. Despite all this the system is inevitably vulnerable. The only serious interruptions to service so far have arisen from the effects of the power strike in early 1972. Initially these were serious until arrangements could be made to ensure a continuous priority supply to the computer centre. Fortunately most causes of serious interruptions to service are likely to hit other airport facilities just as hard as they hit LACES - if that is any consolation to the users.

The fact remains that after five years of hard effort by all the parties concerned - NDPS, Customs & Excise, the international airlines, the cargo agents, Department of Environment and its building contractors, ICL and its subcontractors, CSI and Cossor - London airport has a total system which gives a very much improved handling, movement, customs clearance and accounting system for goods moving through the airport. LACES puts Heathrow several years ahead of any other airport in the world in its facilities for international cargo traffic.

HOW THE PUBLIC PERCEIVES THE COMPUTER: SOME SOCIAL-PSYCHOLOGICAL DIMENSIONS

Thomas L. McPhail, Department of Communications, Ottawa, Canada

In industrialized societies, there is little doubt that the computer is becoming a major technological device in many of the prime areas of human interaction. From the "Little Red School-House" with its computer terminals and computerized learning carousels, to the corporate world with mini-computers on many of the desks, to the home situation where computerized billing services remind us almost daily of the ability of the computer to put names and numbers together on a monthly basis. The purpose of the paper is not to talk about the obvious fact that the computer is the technological innovation of the 20th Century, nor is the purpose to describe the intricate engineering make-up, but rather, the purpose of this paper is for the first time to the best of my knowledge to pull together the emerging empirical studies on how people perceive the computer. For years we have had subjective responses concerning individuals' unique impression of the computer. These, of course, range from the extremely favourable, represented by engineers designing the systems or shareholders holding stock's certificates in companies like IBM, to the extremely negative, represented by individuals who blame the computer for billing mistakes and as well as a host of other social ills. The result has been a rather distorted picture based upon selective perception and selective recollection. The purpose of this paper is to present a brief review of related attitudinal studies and then report a major Canadian study completed recently.

Let me begin by describing an interesting example of a group of people in an experimental situation reporting their perception of the computer. I came by way of this information at a presentation in Vienna last summer by Christopher Evans, an experimental psychologist in the Computer Science Division of the National Physical Laboratory in Britain. Briefly, his experiment is as follows. The experiment took place in a Scottish Hospital and the purpose was to utilize the computer to perform basic, repetitive tasks in order to release the medical profession to perform more professional duties rather than simple administrative duties. But to a sociologist the first problem encountered comes as no surprise, but to others it may be - both a surprise and somewhat frightening. The first task was to program the computer to take a straightforward medical history. The problem encountered after asking several doctors what information they collected for a medical history resulted in several different responses, most of them ambiguous and many of them vague. What was first thought to be a simple problem turned out to be a time consuming task in establishing a clear medical history format. Finally, this was done, patients were divided into an experimental group and a control group. The experimental group was interviewed by the computerized system and the control group was interviewed using the traditional doctor's approach. Dr. Evans programmed the computer to first of all explain how to use the computer and its very simple keyboard. Second, once the patients had learned to use the computer facility, the history taking information was then presented, and finally a computer printout of the information was made available. Here are some of the human reactions uncovered. First of all, the people, despite the anxiety of being both ill and in the hospital, enjoyed the computer interview. Second, at several points in the program the machine was programmed so as to tell the people that they were free to leave if they did not want to continue - none left. In fact, 50% said they preferred the computerized medical history taking procedure over the normal or traditional doctor-patient interaction. Dr. Evans in his presentation also mentioned that people felt more at ease with the computer and that they thought the machine was very polite and very patient, perhaps reflecting negatively on doctor's office manners.

Needless to say, after conducting the original study several members of the medical profession disputed the validity of computerized medical history format as being valid and reliable medical history information. In order to deal with this criticism, another experimental situation was contrived and it goes as follows. To overcome the criticism, Dr. Evans established three types of medical-history interviews and then compared the results of each. The first situation saw a patient being interviewed by a medical doctor and then being interviewed by a computer. The medical histories were essentially the same. The second situation saw a different group of patients being interviewed first by the computer and second by the doctor. Again there was no essential difference in the results.

The third situation saw a patient being interviewed first by a doctor and second by a different doctor. Then they compared the results and here there was a noticeable difference. The doctors disagreed more between themselves. From this point on the doctors were reluctant to challenge Dr. Evans and his computer.

Situations such as this raise the question, is the introduction of the computer in the future going to be mostly a technical question, an economic question, or a sociological question. Again using European data, we can see an indication of the direction of implications of decisions surrounding the computer. Using a sample of 1,500 people in a pre-post test design this question was asked: What is the most important issue in terms of the realization of new telecommunications facilities (of which the computer is obviously one). The pre-test showed 7% thought the technical problems were the most important issue, 46% thought it was an economic issue, and 40% a sociological issue. But after being exposed to a symposium discussing many aspects of telecommunications in the future there was a noticeable shift evident in the assignment of priorities. In the post-test, these were the results. Technical problems dominating decision-making fell to 2%, economic considerations fell to 25%, and the sociological issues and implications being the most important increased to 57%. The point I wish to make is that what happens in the computer-world in the future may no longer be as much a technical or economic issue, but rather a sociological-public acceptance one.

Two additional brief examples concerning perception or the psycho-social impact of computer technology are represented by, first, the findings of the Stanford Center for Research and Development in Teaching. I will not go into the details of this particular study (Hess and Tenezakis, 1971). The thrust of this study was to deal with how students feel about their experience with computer-assisted instruction. Over all, the students had a more favourable image of the computer than of the teacher. In addition, they found the computer learning situation provided more reliable and more objective evaluation than instruction monitored by traditional classroom teachers. The fact that they obtained almost simultaneous feedback on the quality of their performance was found to be rewarding. In fact, both the experimental and the control group of students tended to attribute even charismatic quality to a greater degree to the computer than to the teacher.

The second study that I want to mention briefly is the joint study sponsored by the American Federation of Information Processing Societies Inc. (AFIPS) and Time Magazine. This study involved a sample of 1,100 telephone interviews with a probability sample in the United States. Some of you are probably aware of the study and it represents one of the larger and insightful studies dealing with the computer. The data was collected during July and August 1971, and briefly the results are:

> "In analyzing the survey results, much of the data can be grouped into eight basic areas -- job involvement with computers, incidence of personal problems because of a computer, the role of computers as perceived by consumers, people's image of computer usage in business and in government, the general effect of computer usage as perceived by the individual, attitudes towards career opportunities in the computer field, and the relationship of computer usage to privacy.
>
> Each individual is urged to analyze all of the data and formulate his own conclusions. However, selected data from each area is indicative of the overall results. The following information is representative...
>
> Job Involvement: Those surveyed indicated that the public has a relatively high degree of job involvement with computers. Approximately 49% have had a job requiring either direct or indirect contact with a computer, with 30% currently having such a job. In addition, 15% feel their current job requires some knowledge of computing while 7% report their job requires working directly with computers.
>
> Computer Problems: Thirty-four per cent (34%) of those surveyed reported that they have had a problem "because of a computer." Billing problems were most frequent, accounting for almost half of the difficulties reported.
>
> About 75% of those surveyed reported they had never had problems in having a computerized bill corrected. Of the 24% who reported difficulties, 71% placed the blame on the personnel of the company involved while 12% felt the computer itself was at fault.
>
> Computers and the Consumer: The public view towards the use of

computers in providing consumer benefits is generally positive. Approximately 89% felt computers will provide many kinds of information and services to us in our homes; 65% felt computers are helping to raise the standard of living; and 68% believe computers have helped increase the quality of products and services.

However, in some areas attitudes were less positive. 48% felt computers make it easier to get credit versus 31% who disagreed. In addition, 48% felt the use of computers in teaching children in school should be increased, versus 25% who felt such use should be decreased.

Approximately 63% felt the use of computers should be decreased in sending mail advertisements to the home, and 55% felt the use of computers should be reduced in matching people for dating.

Computers in Business: The public's image of computers as used in business is somewhat mixed. 89% feel we can do many things today that would be impossible without computers vs. 33% who disagreed with this statement. Only 36% felt computers create more jobs than they eliminate vs. 51% who disagreed.

There was also mixed feeling on the apparent reliability of computers. 81% felt "computer mistakes" are really mistakes made by people who use computers; 47% felt computers often make mistakes in processing bills and 37% believe computer systems break down frequently.

Computers in Government: In general, the public believes government should make increased usage of computers in a number of areas, that such usage will make government more effective, and that there will, and should be, increasing governmental involvement in the way in which computers are used.

Of those surveyed, 63% believe large computerized information files will make government more efficient. 70% feel use of computers will not increase the chance of war.

Approximately 5% feel that the government will determine what computers can or cannot be used for and 61% believe the government is presently concerned about regulating the use of computers. In addition, 84% believe government should be concerned about such regulation.

Areas where the public felt the use of computers should be increased include keeping track of criminals, 78%; guidance of missiles for national defense, 71%; vote counting, 66%; surveillance of activist or radical groups, 56%; projection of election results, 50%; and compiling information files on U.S. citizens, 50%.

With respect to accuracy of data, 53% felt that safeguards are used by government to make sure that personal information stored in computers is accurate vs.28% who disagreed.

Computers and Privacy: There is major concern about the use of large computerized information files. 38% of those surveyed believe computers represent a real threat to people's privacy vs. 54% who disagreed. 62% are concerned that some large organizations keep information about millions of people. In addition, 53% believe computerized information files might be used to destroy individual freedoms; 58% feel computers will be used in the future to keep people under surveillance; and 42% believe there is no way to find out if information about you stored in a computer is accurate.

Computers and Life: The general effect of computer usage as perceived by the individual drew a divided response.

Approximately 91% agreed that the uses of computers are affecting the lives of all of us. On the positive side, 75% felt computers will improve our lives and 86% agreed that they will create more leisure time.

However, 55% felt people are becoming too dependent on computers; 54% believe computers are dehumanizing people and turning them into numbers, and 33% felt computers will decrease our freedom vs. 59% who disagreed.

With respect to the computer's ability to "take over," 12% felt computers can think for themselves, while 23% believed computers of the future might disobey the instructions of the people who run them.

TABLE 1

COMPUTERS WILL ENABLE GOVERNMENT AND BUSINESS TO MAKE BETTER DECISIONS

	TOTAL RESPONDENTS (1030)	TOTAL URBAN (780)	TOTAL RURAL (250)	TOTAL MALE (489)	TOTAL FEMALE (541)	UNDER $5,000 (223)	$5,000-$7,499 (258)	$7,500-$9,999 (229)	$10,000-$11,999 (129)	$12,000 OR MORE (126)
	%	%	%	%	%	%	%	%	%	%
Strongly Agree	7	7	5	9	4	5	6	5	10	8
	53	55	44	59	47	39	50	55	61	72
Agree	46	48	39	50	43	34	44	50	51	64
Disagree	28	27	30	24	31	26	30	29	26	21
	31	30	34	29	33	31	33	31	28	23
Strongly Disagree	3	3	4	5	2	5	3	2	2	2
No Opinion	16	15	22	12	20	30	17	14	11	5

COMPUTERS WILL IMPROVE QUALITY OF EDUCATION

Strongly Agree	8	10	4	11	5	7	6	11	8	11
	58	60	55	64	53	52	51	66	65	63
Agree	50	50	51	53	48	45	45	55	57	52
Disagree	26	24	29	23	27	25	29	19	23	30
	28	27	30	25	30	26	33	21	27	30
Strongly Disagree	2	3	1	2	3	1	4	2	4	-
No Opinion	14	13	15	11	17	22	16	13	7	7
Not Stated	-	-	-	-	-	-	-	-	1	-

Career Opportunities: Of those surveyed, 76% were in favor of a young person entering the computer field as a career vs. 5% who would be opposed.

Most saw the jobs in the computer field as interesting (80%), challenging (75%), secure (73%), having high salaries (72%), requiring lots of training (71%), and offering rapid advancement (67%).

To contrast with the preceding materials, I would like to now present the results of a very recent Canadian study done for the Federal Department of Communications (Government of Canada) by the Social Survey Research Center, Toronto Canada. This study commenced with a qualitative study in October, 1970, where adult groups were monitored in order to elucidate the issues concerning the public's perception of computers and computer services. As a result of the qualitative phase, a quantitative survey was initiated in order to evaluate the feelings and reactions of a cross-section of Canadians on the basis of a 73 item questionnaire. A total of 1,030 inhome interviews with adults were conducted in a number of communities throughout Canada. Without going into either the methodology or the statistics employed, I would like to move directly to a consideration of the highlights of the study.

In terms of the study, 85% of the total number of respondents indicated that they had either a direct or indirect (e.g. computerized bills) contact with a comput-

TABLE 2

COMPUTERS WILL CAUSE UNEMPLOYMENT

	TOTAL RESPONDENTS (1030) %	TOTAL URBAN (780) %	TOTAL RURAL (250) %	TOTAL MALE (489) %	TOTAL FEMALE (541) %	PROFESSIONAL /MANAGERIAL (213) %	OTHER WHITE COLLAR (143) %	BLUE COLLAR (404) %	OTHER (270) %
Strongly Agree	22	22	20	17	25	20	24	23	20
	71	68	76	62	78	60	68	74	75
Agree	49	46	56	45	53	40	44	51	55
Disagree	22	24	15	29	16	32	25	20	16
	24	27	16	32	17	35	28	22	17
Strongly Disagree	2	3	1	3	1	3	3	2	1
No Opinion	5	5	8	6	5	5	4	4	8

er. Generally, awareness is positively correlated with urban living rather than rural, males rather than females, higher income rather than lower, being under 50 rather over, and professional or white collar rather than blue collar.

In terms of the perception or image of the computer, 60% thought of it as a very efficient mathematical machine, 19% as just another appliance, and 16% as an intelligent machine. Of this last group, 47% thought of it as being more intelligent than an average person. 54% indicated that there is almost no limit on what it can do.

Concerning the effects of computerization on one's life or on society, the following reflect values concerning benefits, fears, and the impact of change. Table 1 illustrates that slightly over half of the respondents perceive the computer in a beneficial way on the two items presented in the table. A more clear cut distribution was found in terms of 73% agreeing that computers will provide for more leisure time.

Table 2 illustrates an important point -- the effect on employment. Almost three-fourths of the respondents thought that the computer would cause unemployment. Two points are relevant here. First, although the perception vis-a-vis reality may not be real, i.e. the computer may not cause unemployment, the consequence of the perception becomes real, i.e. people fear the introduction of the computer, particularly in the work environment. The second point is that positive reactions are negatively correlated with occupation. The blue collar working class was very much (74%) threatened by the computer.

In terms of some of dysfunctional aspects of increased usage of computers, the following Tables 3-7 point out some of the more pressing concerns. And finally, the impact of computers in terms of changing lives and life-styles is shown in Table 8.

TABLE 3

COMPUTERS WILL TAKE OVER OUR PERSONAL LIVES

	Total Respondents (1030) %
Strongly Agree	4
	-28
Agree	24
Disagree	55
	-60
Strongly Disagree	5
No Opinion	11
Not Stated	1

TABLE 4

PEOPLE ARE GOING TOO FAR IN USING COMPUTERS

	Total Respondents (1030) %	
Strongly Agree	5	-43
Agree	38	
Disagree	35	-38
Strongly Disagree	3	
No Opinion	19	

TABLE 5

COMPUTERS MAKE YOU THINK INDIVIDUALS ARE JUST BECOMING NUMBERS

	Total Respondents (1030) %	
Strongly Agree	13	-62
Agree	49	
Disagree	25	-29
Strongly Disagree	4	
No Opinion	9	

TABLE 6

COMPUTERS WILL MAKE PEOPLE THINK LESS

	Total Respondents (1030) %	
Strongly Agree	10	-55
Agree	45	
Disagree	35	-38
Strongly Disagree	3	
No Opinion	7	

TABLE 7

COMPUTERS THREATEN FAMILY LIFE

	Total Respondents (1030) %	
Strongly Agree	5	-28
Agree	23	
Disagree	51	-57
Strongly Disagree	6	
No Opinion	15	

COMPUTERS THREATEN OUR PERSONAL PRIVACY

	Total Respondents (1030) %	
Strongly Agree	5	-37
Agree	32	
Disagree	44	-48
Strongly Disagree	4	
No Opinion	15	

TABLE 8

CHANGE THROUGH COMPUTERIZATION

DEGREE OF CHANGE:	Own Life (1030) %		Future Lives (1030) %		Society (1030) %	
Entirely	3	11	11	50	6	36
A Great Deal	8		39		30	
Somewhat	39	84	36	43	46	57
Not at All	45		7		11	
No Opinion	5		7		7	

In summary, the research points to three major items. First, attitudes toward and response to the computer elicit from people distinct, specific, and selective reactions to the questions asked. To most people the computer is becoming the overriding technical symbol of the 20th Century. Yet, there probably is very little awareness of the technical complexity of the device. But there is a considerable amount of perceived awareness of both its existence and its impact.

In addition, one can state after reviewing the data that many of the people express two sets of conflicting attitudinal clusters surrounding computer technology. On the cognitive level many people appreciate the benefits derived from increased computerized system in terms of medical, scientific, and technical progress; yet, at the same time, these same people express considerable affective hesitation in terms of what the computer will actually do to them in terms of interpersonal relationships and their day-to-day activities. There is a latent fear of de-personalization and loss of control over decisions affecting one's behavior, and the computer somehow represents the frustration experienced in this situation.

The final point is this research is limited by sample-size, composition, subjects, potential biasness of the original questionnaire, as well as other traditional limitations on survey research data. But even given the limitations, one can still safely conclude that there is a growing body of literature with an empirical data base upon which more objective decisions can be made concerning the psychological and sociological dimensions of computer technology. What is needed is further research of a survey nature as well as laboratory experimentation where more specific issues can be pursued in greater depth and detail to increase our understanding of the human impact of computer technology.

COMPUTER-ASSISTED INSTRUCTION—A SYSTEM AND ITS ASSESSMENT IN JAPAN

Shigeru Watanabe, The University of Tokyo
Hoichi Itaya, Masayuki Shimada, and Kazuo Yagi
Japan Society for the Promotion of Machinery Industry, Tokyo, Japan

Summary

This paper describes the effectiveness of the instruction programs and the characteristics of the biggest CAI system in Japan, developed by the Society for the Promotion of Machine Industry.

The major characteristics of this CAI system is that the computer system centering 32KW mini-computer and 2.5KW magnetic disc memory device was designed to control 30 study terminal devices simultaneously.

At the same time, the CAI language and its compiler were developed to realized the control over the proceedings of the learnings as a software suitable to the instruction. This CAI language is easily understood not only by a programmer, but also by an author of the instruction program.

The assessment group was organized to achieve the best revision of the instruction program on the basis of their results.

Eight instruction programs are prepared for the use of high school graduates and it takes some 9 to 28 hours to finish them. The students had the impressions that the CAI study was more pleasant and easier to understand, although it got them more tired than the ordinary lectures.

In consequence of the comparison from the educational points of view, it was found that the CAI study was superior to the conventional one and that the study hours were remarkably reduced.

Besides, this development was completed under the assistance of the governmental fund. And it has been widely utilized in business and industrial fields.

Introduction

In November 1968, Japan Society for the Promotion of Machine Industry decided on development of the CAI system as one of its projects for diffusion and promotion of new machines, organizing a committee (chairman: Dr. Shigeru Watanabe, Professor of the University of Tokyo). The CAI system was completed in March 1972 by a joint development of Japanese major electrical companies. Expenses paid for its completion amount to about $10 mil.

1. Hardware of the CAI system

(1) Composition

The composition of the CAI system is as shown in Figure 1. This system is designed for individuals. It is composed of a computer system equipped with a 2.5MW magnetic disc memory device and 30 terminal divices. The terminal devices comprise two kinds. The society has 8 I-type terminal devices and 22 II-type terminal devices. The former type of terminal consists of a random access slide device (RAS), a CRT display device with a light pen, a random access audio device (RAA), a keyboard (KB) and a light pen (LPN), and the later, a random access slide device, a printer (PR) and a key board. These devices were manufactured by the following companies:

Hitachi, Ltd. (the electronic computer system), Nippon Electric Co. (keyboards, printers and booths), Tokyo Shibaura Electric Co. (RAS and RAA) and Matsushita Electric Industrial Co. (CRT display devices).

Specifications of Component Devices

o Electronic computer system
Central processing unit (memory capacity: 32KW)
Tape reader
Tape punch
Line printer
Card reader
Card punch
Magnetic disc driving gear (memory capacity; 2,000KW)
Disc pack
Off-line card punch, verifier
Collective controller
Communication controller
Peripheral device controller
Software
System program & CAI language compiler
Terminal devices
Key board
Key format
Data key (white) 65 (two-stage shift)
Function key (blue) 10, Shift key (blue) 1
Symbol: 8 bits
Standard (ISO) cord
Connection type: Multiwire type
Printer
Printing type: Type-wheel type
Fixed platen
Printing speed: 1,000 letters per minute
Number of letters on one line: 128 letters
Printing pitch: 10 letters per 25.4 mm
Line space: two-stage switchover of 6 lines per 25.4mm and 3-lines per 25.4mm
Ink ribbon: black
Symbol: Eight bits
Standard (ISO) cord
Random access slide device
Type of original picture: 16mm movie film
Number of frames of original pictures: Maximum 3000 frames
Volume of information contained in one frame:
About 250 to 300 letters
Film moving speed: 50 frames per second (in both right and opposite directions)
Screen: See-through type
300 (breathwise) x 225 (lenghwise)mm
Random access audio device
Recording medium type: Type of magnetic sheet attachable and detachable around a drum
Capacity of information contained in one magnetic sheet: Maximum 384 message
Length of one message: Maximum 15 seconds
Average access time: 3 seconds
Type of speaker: Headphones
CRT display (including a figure controller)
Form of display: Dot
Letters, English figures, symbols, katakana, (a kind of Japanese letter)
5 x 7 dots 7-color display (red, green, blue, yellow, cyan, magenta and white)
640 letters per screen (32 letters x 20 lines)
Figure: Circle, circular arc, straight line, letter (7 x 7 dots) are composed within the limits of 176 (breathwise) x 128 (lengthwise) dots.

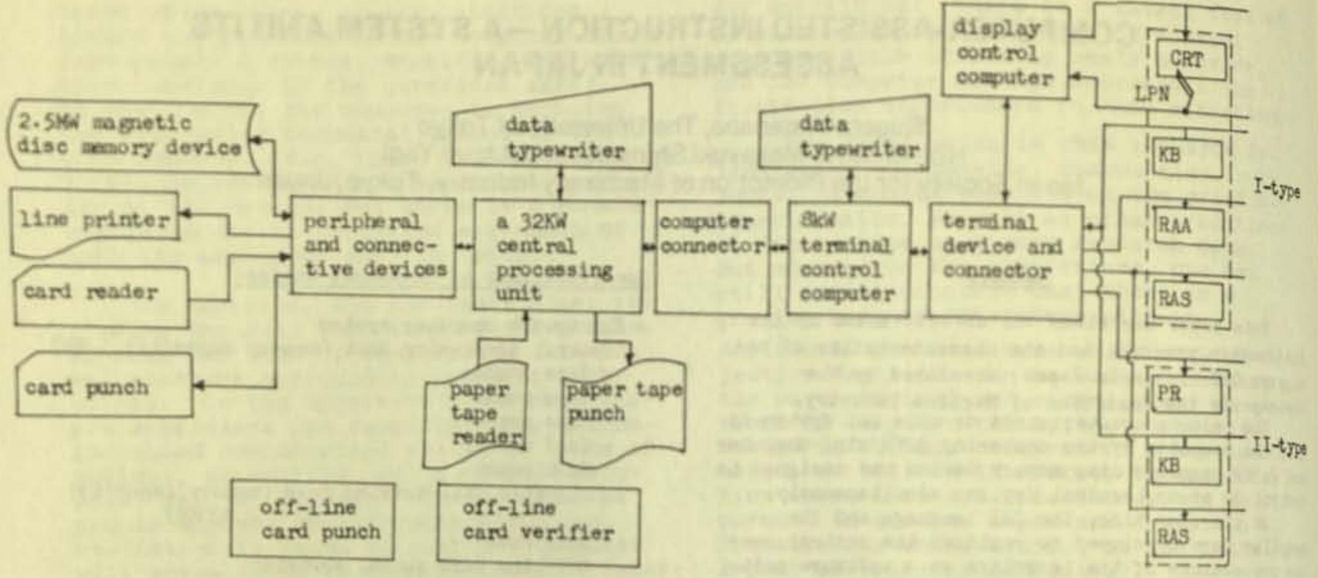


Figure 1 Composition of the CAI system

Figure buffer memory: Magnetic core 1kW/22 bits
Braun tube: 16-inch color Braun tube, control mini-computer (8kW)
Number of terminal devices connected: 8 (with light pens)

(2) Features

1) All students are able to have a conversation within waiting time of three seconds.
2) Subject of instruction program is limitless.
3) As the form of response to questions, not only multiple choice, but also constructed response, such as answering by use of English letters, Kana letters and figures can be available. Answers also can be made by detecting a point on the screen with a light pen.
4) The system is designed so that students can not only answer questiones and explanations presented by the instruction program, but also that they can make requests for hints, correct answers and reference materials as they want.
5) When a student answers a question presented, the computer records his results, the process of the program he passed and the time he needed. It can judge the ability of the students and present the next question according to his ability. That is, branching is employed for instruction programs as the methods of effective learning.
6) The process of study is recorded and displayed. Contents of the record include the name of the lesson course, the name of the student, the question presented, the answer and the time spent.

In the following, we will explain Feature 3), 4) and 5) in detail.

In the CAI system, learning proceeds through repetition of dialogues between the computer and the student as shown in Figure 2.

Presentation of instructional materials is made by audio-visual devices; the random access slide (RAS), the printer, the CRT display and the random access audio (RAA). To this, students make answers by the key board (KB) and the light pen (LPN). The key board is composed of English capital letters, figures, special symbols and Kana letters. It is used for information input. Concerning the function keys, REV, ANS, CA, HINT, OK and CALL keys are prepared for study, and DEL CAN keys for correction of information. There is also a key arranged for debugging. (See Table 1)

To the answers by the student, the computer makes a judgement and a decision as shown Figure 3.

Table 1

REV	This key is used to require for return to the previous step
HINT	This key is used to require hints
OK	This key is used to request for return to the main sequence from HINT sequence
CALL	This key is used to call instructional reference materials
ANS	This key is required both for proceeding to the next step and for making the system recognise the input massages as the student's answer
DEL	This key is used to eliminate letters of input messages from the last one by one
CAN	This key is used to eliminate the whole input messages at a stroke

2. Software (instruction program) of the CAI system

The Japan Society for the Promotion of Machine Industry has prepared 8 instruction programs for the CAI system. In the following, we will describe the features of the instruction program, the method used for its development and the effects of the CAI.

(1) The level of the student suitable to the instruction program and its goal

"NC machine tools"

Qualified students: Those who have scholarship equal to high school graduates and two or three years' experience of working at machinery plants.

Goals of instruction: To enable them to understand handling of NC machine tools and handle them, and to enable them to work out a program type for simple

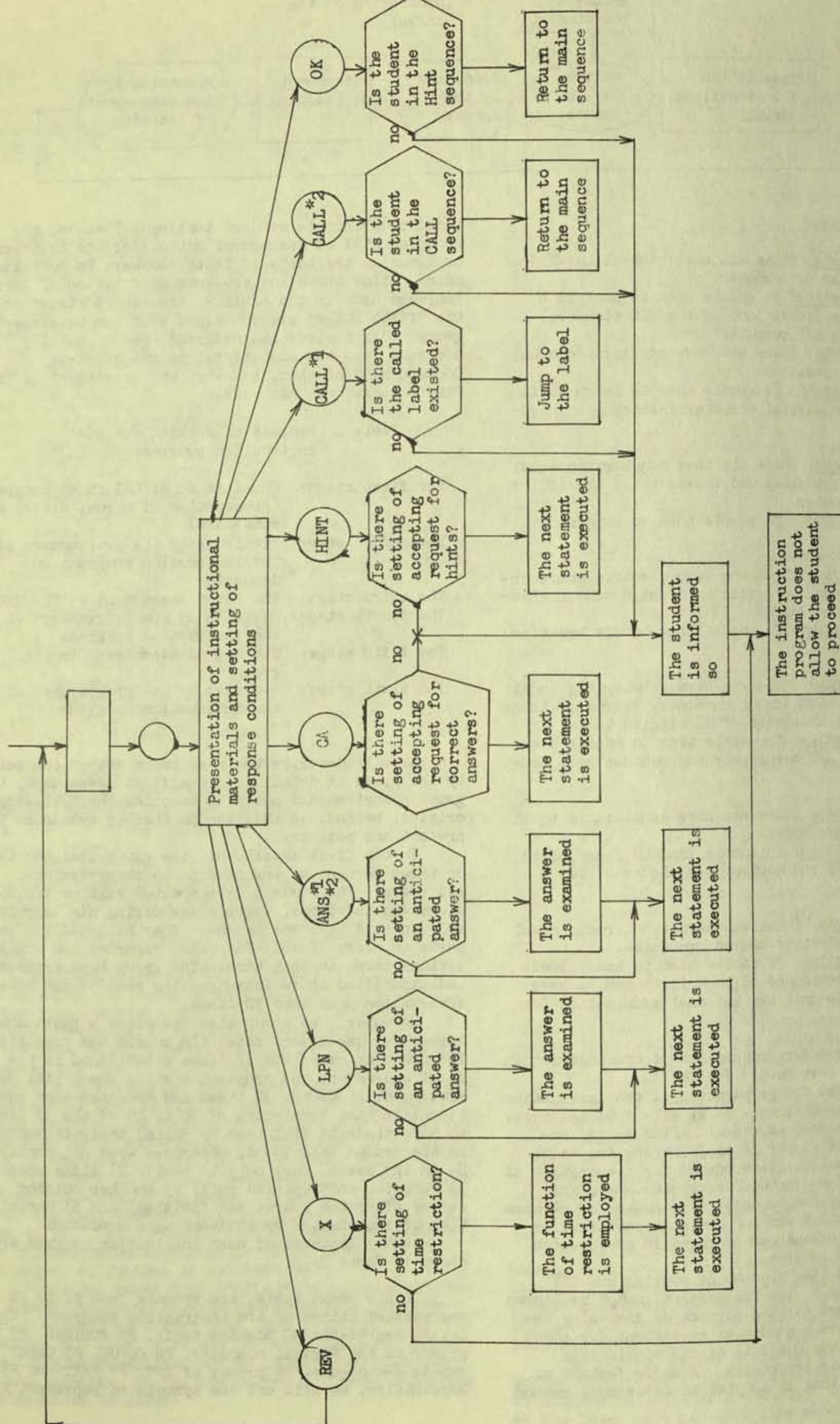
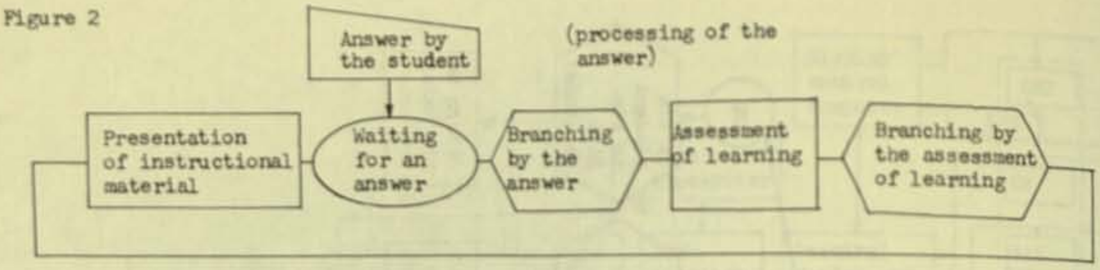


Figure 3 Algorithm of judgement and decision

Figure 2



works by themselves.

"FORTRAN"

Qualified students: Those who have general knowledge about the outline of computer.

Goals of instruction: To enable them to program by FORTRAN language (level 5000 of the JIS standard)

"Primer of the digital computer"

Qualified students: Those who have scholarship equal to high school graduates.

Goals of instruction: To enable them to learn basic knowledge of the digital computer centering on hardware.

"COBOL"

Qualified students: Those who have scholarship equal to high school graduates and knowledge of the ABC of the digital computer.

Goals of instruction: To enable them to program by COBOL language.

"Color TV repair techniques"

Qualified students: Those who have electric engineering knowledge equal to graduates of high school electric or electronic course.

Goals of instruction: To enable them to fix troubles of color TV sets.

"Primer of APT"

Qualified students: Those who have scholarship equal to graduates of high school machinery course or those who have general knowledge of NC machine tools.

Goals of instruction: To enable them to master the middle class of APT completely and conduct part programming of APT

"PERT"

Qualified students: Those who have scholarship equal to high school graduates.

Goals of instruction: To enable them to learn basic knowledge of PERT and control techniques of its application.

"ZD"

Qualified students: Those who have scholarship equal to high school electric course graduates.

Goals of instruction: To enable them to learn basic knowledge and techniques of lectricity and to have ability enough to pass the national examination for electricians.

(2) Method employed for preparation of the instruction programs and cost needed for it

The cost needed for preparation of the instruction program amounted to $1/4 mil. to $1 mil. Of the cost, 60 to 70% was personnel expenses. Slide making expenses were about $1.5 per frame for black and white and about $4 for color. Time needed for preparation of the program was eight months for shorter ones, and about two years for longer ones. The number of persons engaged in preparation was for the most part four or five. While, 10 persons were engaged in preparation for an instruction program.

Table 2 shows the "process of working out CAI instruction program." The "assessment group" mentioned is a group organized for the purpose of assessing the CAI system. The group conducted assessment for about 10 months. The group consists of six specialists including university professors. A major feature of the assessment is that it adopted an assessment method based on principle student testing. That is, the group adopted an assessment method for the instruction programs on the basis of tests, answers to questionnaire, slide check and students description of their impressions. In the table 2, the "third stage" and the "fifth stage" are concerned with the assessment group. Six instruction programs were especially assessed from the educational points of view. Concerning 2 of them, however, the assessment of the "fifth stage" was not done.

Table 2 Process of working out "CAI instruction programs"

The first stage:	Compilation of CAI study programs (preparation of cards, slides and materials)
The second stage:	Trial-use by the programmers of the instruction programs and debugging
The third stage:	The first assessment by the group
The fourth stage:	Revision by the programmers of the instruction programs
The fifth stage:	The second assessment by the group
The sixth stage:	Revision by the programmers of the instruction programs

We will describe the assessment method conducted by the group in detail. In the first assessment, the group contrived and put into practice the following original method, in addition to that by readiness test, pretest and post-test.

In the first assessment, the group assigned two or three persons to one student as observers and recorders. The student introspected various impressions and questions, which arose during his learning and reported them one by one. They recorded the reports on every frame. They also recorded the students actions taken at that time. The group then arranged and classified those introspection records and action observation ones into 32 items.

Moreover, the group also conducted a quationnaire survey for assessment after the course was completed. The group examined the 32 items established in the first assessment and prepared a slide check list consisting of 11 of the 32 items and "chapter-end" questionnaire list consisting of 20 to 29 of the 32 items to have students fill them up during their learning. In the second assessment, observers were not employed. So in the second one, only 1) tests, 2) slide check, 3) "chapter-end" opinion poll, 4) another opinion survey after the course was finished, and 5) description of impressions by students were conducted.

Lastly features of each instruction program are shown in Table 3.

In the above table, the symbols from P1 to P8 represented titles of the instruction programs. That is, P1 represents NC machine tools, P2 FORTRAN, P3 primer on the digital computer, P4 COBOL, P5 Color TV repair techniques, P6 primer of APT, P7 PERT and P8 ZD. Hereinafter, we will use the symbols to represent the

Table 3 Features of each instruction program

	P1	P2	P3	P4	P5	P6	P7	P8
Number of chapters	16	14	19	18	31	14	7	9
Number of slides	1030	1230	2210	2040	1310	1000	1600	1260
Kind of slide	Color	Black&White	Black&White	Color	Color	Color	Color	Black&White
Auxiliary material	1 copy	2 copies	1 copy	3 copies	2 copies	1 copy	1 copy	Null
Terminal device	II-type	II-type	II-type	II-type	I-type	I-type	II-type	II-type

titles.

(3) Results of tests

Results of learning recorded during execution of the instruction programs from P1 to P6 are shown in Table 4. Those for P8 are omitted, because a study for check under the program has not yet been formally conducted. The students, who took part in the tests, were for the most part university students. Some of them did not satisfy preconditiones for the study, and others had already attained the goals of instruction. Both of those students were excluded from the assessment. To be concrete, those students who got less than 70 marks in the readiness test and more than 80 in the pretest were excluded from the table. As is known in the Table 4, it can be said that the initial goals of the instruction programs were attained to some extent, except P1 and P5. The results of the post-test for P5 were not good. This is because practical exercises, which are necessary for learning, were not conducted. The results of learning, in which exercises were conducted, will be mentioned later in (6).

In this treatise, in assessing the results of the study, the precondition was established uniformly for all of the instruction programs that the students should get more than 70 marks in the readiness test. But further examination is required. In the instruction program P1, if the marks that must be obtained in the readiness test are set at 90, the average in the post-test will become 82.6 marks.

(4) Study hours

Each instruction program was prepared on the assumption that it will take about 20 hours. But it took 11 to 15 hours for students to complete the instruction program.

In order to see variation of study hours, we calculated coefficient of variation. The coefficient was the largest for P4 with 0.462, and the smallest for P7 with 0.09. That of P7 is small, because all the students chose the same linear type sequence of the course.

Table 5-2 shows the time of thinking, reading and key striking and their percentages per chapter for each instruction program respectively. Concerning P1 and P2, the percentage of thinking time are the largest. In the case of P1, about half the study hours of the chapters was spent for thinking. But, as shown in Table 5-3, the thinking time per frame of P1 is smaller than that of P2. It can be said, therefore, that frames required for thinking are larger in number in P1 than P2. The percentage of thinking time of P3 is the same as that of P2. But the study hours of any chapter in P3 is short in general, and therefore, the time required for thinking per frame is short. It is obvious from this that students will have more fatigue in P2 than in P3. It is very interesting when the results of the questionnaire mentioned in (5) are referred to. The degree of fatigue is smallest in P1.

Table 5-4 shows that in P1, P3 and P5, the number of frames requiring answers is larger than that requiring answers is larger than that requiring only reading, and that in P2 and P4, the number of frames requiring only explanations is larger than that requiring answers. In P4, the number of frames requiring explanations is large.

(5) Impressions of students (analysis of the questionnair)

Table 6-1 and 6-2 show the results of polls gathered after the courses. The figures of ψ_{it} were calculated by subtracting the ratio of negative opinions from that of affirmative ones. The figures of ψ_t are averages of ψ_{it}. Therefore, both indices are values between +1.00 and -1.00. They show that the nearer they are to 1.00, the better the impressions of the students, and that the nearer they are to -1.00, the worse those of the students.

Table 6-1 shows the results of the poll for II-type terminal device. In the table the negative assessment is pointed out as a general tendency that "they are tired by the CAI study" and that "they feel like quitting the study halfway." On the other hand, it can be cited as affirmative opinions that "typed letters are readable," that "they can understand the contents of the study well" and that "the study by the CAI system is more pleasant than ordinary studies."

In the results of the poll for the I-type terminal device, that "the sound is hard to catch" (-0.78), that "timing of sound is not good" (-0.78), and that "we get tired more than ordinary studies" (-0.70) can be cited as negative opinions. On the other hand, as affirmative opinions, "colors of the letters on CRT are good" (0.78), that "the place where the letters appear on CRT is appropriate" (0.45), that "we could study at our own pace" (0.40) and that "we never felt uneasy during the study" (0.40) can be cited.

The indices ψ_t for the instruction programs for the II-type terminal device show that all the opinions are affirmative, while those for the I-type one negative. It is considered that this is partly because the instruction program for the II-type terminal device were checked twice by the assessment group except P7. As to the assessment of the instruction programs for the I-type terminal device, it is thought the assessment is bad, because the device is more complicated than the II-type one, and because of tone obscurity.

(6) Comparison of effects by the CAI study with those by lectures

Students' impressions of lectures and the CAI system are shown clearly by the results of polls mentioned in (6). The greater part of them answered that the CAI study was more pleasant and easier to understand than by lectures. This is "significant" at the level of 1 per cent by χ^2 test.

If the CAI study and ordinary lectures are conducted with the same contents of study, what difference will be expected? In this connection, an experimental study was conducted with the P6 and P7 instruction programs. Table 7-1 shows results of tests in the measurement of the study effect by lectures and by the CAI study taking P7 for example. The procedure of the tests is as follows:

First, learners were divided into two groups on the basis of results of the readiness test and the pretest conducted prior to the study. Then one of the groups studied by lectures, and the other by the CAI system. All members of the two groups took the post-test when they finished the study.

The results of the post-test of the CAI group were better than those of the lecture group on an average of 67.2 marks and S.D. of 6.2 marks. This is "significant" at the level of 5 per cent (one sided test) by means of

Table 4 Results of tests

		P1	P2	P3	P4	P5	P6
Post-test	Average	73.8	91.3	86.2	80.2	62.5	86.0
	S. D.	11.2	37.7	8.9	14.6	15.7	4.3
	Range	88-51	97-86	100-61	94-43	76-37	90-80
Pretest	Average	16.6	41.6	23.6	16.6	14.0	3.3
	S. D.	8.2	26.3	21.7	19.4	16.9	4.2
	Range	33- 0	79- 0	86- 0	67- 0	43- 0	10- 0
Readiness	Average	81.1	82.5	95.9	77.8	78.3	81.3
	S. D.	7.2	6.2	4.6	16.3	6.0	5.2
	Range	91-70	96-73	100-80	100-40	88-72	85-74
Number of students		10	12	36	14	4	3

Table 5-1 Study hours of each instruction program

	P1	P2	P3	P4	P5	P6	P7
Average	13hr.41min.	10h 59m	14h 31m	8h 31m	28h 21m	6h 36m	14h 20m
S. D.	3h 48m	2h 46m	5h 9m	3h 56m	8h 24m	2h 6m	1h 17m
Maximum	23h 5m	15h 42m	27h 15m	12h 25m	38h 48m	8h 25m	16h 40m
Minimum	9h 51m	5h 35m	5h 57m	7h 30m	17h 25m	3h 39m	11h 50m
Coefficient of variation*	0.278	0.252	0.352	0.462	0.308	0.318	0.090
Number of students	9	8	36	10	3	3	10

* Coefficient of variation was calculated by dividing S.D. by average.

Table 5-2 Analysis of study hours within a chapter

	Thinking	Reading	Key striking	Average study hours
P1	1235 seconds(49%)	357 seconds(14%)	938 seconds(37%)	2530 seconds
P2	1795 " (41%)	1304 " (30%)	1283 seconds(29%)	4415 "
P3	483 " (41%)	152 " (13%)	532 " (46%)	1167 "
P4	1079 " (36%)	1678 " (55%)	281 " (9%)	3038 "
P5	2391 " (40%)	467 " (8%)	311 " (52%)	5969 "

Table 5-3 Time spent for a frame

	Thinking	Reading	Key striking
P1	13 seconds	6 seconds	9 seconds
P2	34 "	21 "	21 "
P3	14 "	15 "	6 "
P4	14 "	8 "	4 "
P5	28 "	21 "	6 "

Table 5-4 Analysis of frames within a chapter

	Number of frame requiring answers	Number of frames requiring reading
P1	95 (61%)	60 (39%)
P2	52 (46%)	62 (54%)
P3	35 (78%)	10 (22%)
P4	77 (27%)	210 (73%)
P5	85 (79%)	22 (21%)

a statistical test. (We adopted the non-parametric method and the Mann-Whitney U test).

Concerning the study hours, the lecture group took 20 hours. On the other hand, the CAI group took 14 hours and 20 minutes on an average which is shorter than that by 5 hours and 40 minutes.

We conducted the experimental study with the instruction program P6 under three different conditions. The results are shown in Table 7-2. The "CAI group" mentioned in the table learned only by the CAI system without the practical exercises. Table 7-2 shows that the three groups are different from one another in respect of readiness. If the lecture-plus-practise group is reorganized so that it may become equal to other groups in terms of readiness, it will be as shown in Table 7-3.

The averages of the readiness test of the two groups- the lecture-plus-practise group and the CAI-plus-practise group- are the same. So is the S.D. But the post-test results of the former and the latter group wer 79.2 and 90.4 marks, respectively. So the CAI-plus-practical group is concluded superior to the lecture-plus-practical group in educational effectiveness. This difference is "significant" at the level of 5 per cent (one sided test), when it is examined by means of the U test.

Comparing the effects only by the CAI study with those by lectures and practise, the post-test of the lecture-plus-practise group are about 10 marks superior to the CAI group on an average as shown in Table 7-4. But when it is examined by the U test, the significant difference cannot be recognized, even at the level of 10 per cent.

Finally, the study hours of the CAI group were about half those of the lecture group.

Table 7-4

CAI Group

Readiness	Post-test
88	62
72	37
78	76
75	75
Average 78.3	62.5
S.D. 6.0	15.7

Lecture-plus-practise Group

Readiness	Post-test
84	85
82	62
79	76
73	53
74	86
Average 78.4	72.4
S. D. 4.3	12.9

Table 6-1 Results of a poll gathered after the courses (II-type terminal device)

			Total	ψ_{it}
1.	Was your study by the CAI system pleasant?	Yes Not quite pleasant No	45 (45%) 55 (54%) 1 (1%)	0.46
2.	Was the study more pleasant than ordinary lessons?	Yes Not quite pleasant No	57 (57%) 36 (36%) 7 (7%)	0.50
3.	Could you understand the contents of the study well?	Yes Not quite well No	57 (56%) 42 (41%) 3 (3%)	0.53
4.	Was the study by the CAI system easy to understand compared with that by lectures?	Yes Not quite easy No	52 (53%) 34 (34%) 13 (13%)	0.40
5.	Did you get tired by the CAI study?	Yes Not very much No	25 (25%) 61 (61%) 13 (13%)	-0.12
6.	Did you get tired more than by ordinary lessons?	Yes Not very much No	38 (38%) 43 (43%) 19 (19%)	-0.19
7.	Could you study at your own pace?	Yes Not quite so No	54 (53%) 39 (39%) 8 (8%)	0.45
8,	Generally was the study difficult?	Yes Not so difficult No	17 (17%) 47 (48%) 35 (35%)	0.18
9.	Did you feel uneasy during the study?	Yes A little No	22 (22%) 36 (36%) 41 (42%)	0.20
10.	How long do you think a proper learning hour by the CAI system is?	Average	92.5 min.	
11.	How long do you think you can study by the CAI system continuously?	Average	174.2 min.	
12.	Generally could you see the slides well?	Yes Not quite satisfactory No	41 (41%) 45 (45%) 14 (14%)	0.27
13.	Could you understand figures of the slides at once?	Yes Not quite satisfactory No	40 (40%) 50 (50%) 11 (11%)	0.29
14.	Were the contents of the slides easy to understand?	Yes Not quite easy No	31 (31%) 56 (56%) 14 (14%)	0.17
15.	Were the questions in the slides clear?	Yes Not quite clear No	29 (29%) 62 (62%) 10 (10%)	0.19
16.	Could you operate type keys as you wish?	Yes No	67 (67%) 34 (34%)	0.33
17.	Did the sound of the printer annoy you?	Yes No	26 (26%) 72 (74%)	0.48
18.	Could you read the type letters well?	Yes No	89 (90%) 10 (10%)	0.80
19.	Did you feel like quitting the study halfway?	Yes No	62 (64%) 35 (36%)	-0.28
20.	Do you think the function "time-up" is useful?	Yes No	34 (49%) 36 (51%)	
21.	Was the study note easy to use?	Yes Not quite easy No	3 (27%) 7 (64%) 1 (9%)	0.18

	P1	P2	P3	P4	P7	P5	P6
ψ_t	0.13	0.28	0.36	0.34	0.14	-0.110	-0.113

Table 6-2 Results of a poll gathered after the courses (I-type terminal device)

		ψ_{it}		
		P5	P6	Total
1.	Was your study by the CAI system pleasant?	0.50	-0.50	0.10
2.	Was the study more pleasant than ordinary lectures?	0.50	0	0.30
3.	Could you understand the contents of the study well?	0.17	0.25	0.20
4.	Was the study by the CAI system easy to understand compared with that by lectures?	0.33	0	0.20
5.	Did you get tired by the CAI study?	-0.83	-0.25	-0.60
6.	Did you get tired more than ordinary lessons?	-0.83	-0.50	-0.70
7.	Could you study at your own pace?	0.33	0.50	0.40
8.	Generally was the study difficult?	0	0	0
9.	Did you feel uneasy during the study?	0.33	+0.50	0.40
10.	How long do you think a paper learning hour by the CAI system is?	116-133 min.	100 min.	
11.	How long do you think you can study by the CAI system continuously?	195 min.	165 min.	
12.	Generally could you see the slides well?	-0.50	0.50	-0.10
13.	Could you understand figures of the slides at once?	-0.50	-0.25	-0.40
14.	Were the contents of the slides easy to understand?	-0.33	0.25	-0.10
15.	Were the questions in the slide clear?	-0.33	-0.75	-0.50
16.	Could you operate type keys as you wish?	-0.60	-0.50	-0.58
17.	Did you feel like quitting the study halfway?	-0.66	-0.50	-0.60
18.	Were the sounds easy to hear?	-0.20	-1.00	-0.78
19.	Was the timing of sounds good?	-1.00	-0.50	-0.78
20.	Did the sounds and slides fit well?	-0.60	-0.25	-0.45
21.	Was the color of the letters on CRT good?	1.00	0.50	0.78
22.	Was the position where the letters on CRT appeared proper?	0.40	0.50	0.45
23.	Was the timing of appearance and disappearance of letters on CRT good?	0.40	-0.50	0
24.	Was the light pen easy to use?	0	0	0
	Ψ_t	-0.110	-0.113	

Table 7-1 Comparison of effects of the CAI study with those by lectures

CAI Group

Student	Readiness test	Pretest	Post-test	Study hours
1	80	20	63	1,000 minutes
2	60	35	68	797 "
3	60	25	73	825 "
4	40	25	77	825 "
5	80	15	63	905 "
6	60	15	70	935 "
7	60	15	58	865 "
8	80	5	68	710 "
9	60	5	61	832 "
10	40	5	77	904 "
Average	62.0	16.5	67.8	860 " (14hr. 20min.)
S. D.	13.9	9.5	6.2	76.7 min. (1 hr. 16.7min.)
Range	80 - 40	35 - 5	77 - 58	16hr. 40min. - 11 hr. 50 min.

Lecture Group

Student	Readiness test	Pretest	Post-test	Study hours
1	80	40	73	20 hours
2	40	35	67	"
3	80	20	63	"
4	80	10	53	"
5	60	20	63	"
6	80	10	52	"
7	60	15	61	"
8	40	15	39	"
9	40	20	57	"
10	60	0	50	"
11	40	5	73	"
Average	60.0	17.4	59.2	"
S. D.	17.1	11.2	9.8	0
Range	80 - 40	40 - 0	73 - 50	0

Table 7-2 Results of tests with the instruction program P7

Lecture-plus-practise Group

Student	Readiness test	Pretest	Post-test	Study hours
1	68	24	53	42 hr.
2	90	57	83	"
3	90	46	79	"
4	81	36	58	"
5	89	33	74	"
6	94	10	87	"
7	79	36	76	"
8	90	26	54	"
9	89	6	65	"
10	86	46	86	"
11	73	32	53	"
12	86	17	80	"
13	80	42	60	"
14	94	24	75	"
15	94	29	89	"
16	82	4	62	"
17	83	43	75	"
18	94	50	82	"
19	56	31	33	"
20	80	27	65	"
21	90	30	98	"
22	88	20	60	"
23	89	25	74	"
24	84	32	85	"
25	93	14	68	"
26	74	6	86	"
27	67	12	68	"
28	86	20	64	"
29	91	12	70	"
30	95	23	85	"
31	87	54	83	"
Average	83.9	27.9	71.9	"
S. D.	8.9	14.1	13.6	0
Range	94 - 56	57 - 4	98 - 33	0

CAI-plus-practise Group

Student	Readiness test	Pretest	Post-test	Study hours
1	97	42	95	1,421 min.
2	93	50	92	1,224 "
3	79	32	92	1,063 "
4	90	24	82	1,289 "
Average	89.8	37.0	90.3	20hr. 49min.
S. D.	6.7	9.8	4.9	2hr. 9min.
Range	97 - 79	50 - 24	95 - 82	

CAI Group (CAI system only)

Student	Readiness test	Pretest	Post-test	Study hours
1	49	2	71	
2	88	0	62	2,328 min.
3	72	8	37	1,730 "
4	66	13	36	864 "
5	78	4	76	1,045 "
6	75	43	75	
Average	71.3	11.6	59.5	24hr. 52min.
S. D.	11.9	14.5	16.87	9hr. 41min.
Range	88 - 49	43 - 0	76 - 36	

Table 7-3 Effect of study of the two groups, which had practical exercises

Lecture-plus-practice Group

Readiness	Post-test
95	85
94	82
94	89
94	87
94	75
93	68
91	68
90	83
90	79
90	54
90	98
79	76
74	86
Average 89.8	79.2
S.D. 6.0	10.8

CAI-plus-practise Group

Readiness	Post-test
97	95
93	92
90	82
73	92
Average 89.8	90.3
S.D. 6.7	4.9

IMPACT OF COMPUTER COMMUNICATIONS AS TOOLS IN THE SOCIAL AND ECONOMIC DEVELOPMENT OF CANADA

H.J. von Baeyer
The Canadian Computer-Communication Task Force, Ottawa, Canada

ABSTRACT

The Canadian Government decided in 1970, to set up a Task Force on computer communications with the objective of developing a set of policy recommendations for public discussion and subsequent formulation of government policies. The Task Force report has now been published in two volumes under the title: "Branching Out".

The Task Force investigations concentrated on the peculiarities of the Canadian situation and the special problems posed by these peculiarities concerning public policy issues in the broad field of computer communications.

Dr. von Baeyer, who was head of the Computer/Communications Task Force, and continues his work for the Canadian Government in an advisory capacity, will concentrate on the impact of computer/communication policies on the Canadian economic and social development, as described in the Task Force Report, and will discuss the highlights of the Task Force recommendations which are at present the subject of public review.

IMPACT OF COMPUTER COMMUNICATIONS AS TOOLS IN THE SOCIAL AND ECONOMIC DEVELOPMENT OF CANADA

[illegible]

The Canadian Computer-Communication Task Force, Ottawa, Canada

[illegible]

The Canadian Government [illegible] set up a Task Force [illegible] "Branching Out".

[illegible]

[illegible]

DATA BANKS AND INDIVIDUAL PRIVACY

CHAIRMAN:

Dr. Alan F. Westin, Professor of Public Law
Department of Political Science, Columbia University
New York, New York, U.S.A.

PROGRAM COORDINATOR:

Philip M. Walker
Georgetown University Law Center
Washington, D.C., U.S.A.

PAPERS:

Dr. Bryan Niblett, Chairman, Law Specialist Group, British Computer Society, London, England, *"Data Banks and Individual Privacy: Developments in the United Kingdom"*

Dr. Hans P. Gassmann, Directorate for Scientific Affairs, Organization for Economic Cooperation & Development, Paris, France, *"Data Banks and Individual Privacy: The Situation in the German Federal Republic"*

SPEAKERS:

Dr. Calvin C. Gotlieb, Professor, Department of Computer Sciences, University of Toronto, Toronto, Ontario, Canada, *"Data Banks and Individual Privacy: Developments in Canada"*

Dr. Alan F. Westin, *"Data Banks and Individual Privacy: Implications for the U.S. from Foreign Developments"*

THEME: This session will attempt to compare the experience of the U.S., Canada and several European countries in developing, using and controlling computer-based data banks which contain information on the individual citizen. The current state of data bank development in each country will be described and the legal measure–legislation, regulatory agency actions, and court decisions–taken to prevent infringement of the individual's civil liberties will be explained. Any steps taken by the computer industry or professional organizations in these countries to protect privacy in data banks will also be explored. Finally, each speaker will discuss those government and private actions which he feels should be taken in his country to safeguard the right of privacy in the computer data bank.

DATA BANKS AND INDIVIDUAL PRIVACY: DEVELOPMENTS IN THE UNITED KINGDOM

Bryan Niblett
British Computer Society
London, England

INTRODUCTION

The United Kingdom, like most other highly developed Western countries, is undergoing a period of intense planning and implementation of data banks storing in computer-accessible form information about the personal affairs of private individuals. These developments are accompanied by public awareness of the benefits which such data banks may bring. To give but one example, the success of computerized air-line reservation methods, to which the British State air-lines BOAC and BEA have made noteworthy contributions, has brought the public at large to realize the convenience and utility of such systems. It is becoming apparent that integrated data files with on-line up-dating and interrogation, coupled with the power of the computer to analyze the data in an endless variety of relationships, can introduce a new dimension in social planning, and help solve the many problems that arise at the growing points of our Society.

At the same time there is increasing public concern, expressed sometimes calmly, sometimes vociferously, with the potential hazards to individual liberty that data banks may engender, and a growing apprehension that data bank technology, like many another technology, is accompanied in its application by threats to the quality of life. One of the most significant events in the recent British debate on computers and personal freedom was a Workshop on the Data Bank Society organized by the (United Kingdom) National Council for Civil Liberties in November 1970 (1)*. A remarkable feature of this meeting was the heterogeneity of the participants: computer specialists, lawyers, civil libertarians, government officials, members of both Houses of Parliament, bankers, journalists, University dons, not to mention maverick members of the public - all bore witness to the universality of concern in the United Kingdom with the hazards that data banks may bring.

A major study of intrusions into privacy in the United Kingdom and the need for safeguards has recently been undertaken by a Committee on Privacy (the Younger Committee) appointed by the previous government in 1970. This Committee has devoted a considerable part of its Report to the privacy problem posed by computers. The present paper takes advantage of the recent publication of this study (2) to provide a brief survey of some aspects of the growth of data banks in the United Kingdom, to examine selected elements of the accompanying public discussion, and to consider some of the current proposals for regulation and control.

DATA BANK DEVELOPMENTS IN THE UNITED KINGDOM

In discussing selected examples of data bank development in the United Kingdom it is natural to turn first to central and local government since these are the areas in which personal files are most likely to be held. In central government there has been a rapid growth in the use of computers for administrative data processing since about 1964. As from the 1st April 1972, a new Central Computer Agency has been established as part of the Civil Service Department, with a staff of some 600 and an annual budget of 30 million, whose task is the formulation of policy and planning for computers in central government and procurement of the necessary hardware and software. The policy document "Computers in Central Government: Ten Years Ahead" (3) prepared in 1970 by three consultants is not an official statement of Government policy, but is nevertheless a helpful account of the likely direction in which central government computer plans are likely to develop in the next decade.

An example of the sort of data bank development underway is the Personnel Record Information System for Management (PRISM) a comprehensive data base management system suitable for on-line processing. The system is designed to integrate across the Civil Service through a network of computers the calculation of salaries and wages with the maintenance of personnel records and a register of posts. It will give instant access to data files containing information on all non-industrial civil servants. It is expected that this scheme will take about ten years to implement fully.

---

*Numbers in parenthesis refer to footnotes at the end of the paper.

The Ministry of Transport is responsible for the development and implementation at Swansea of a major system for the registration of vehicles and driving licences. This comprehensive scheme covering the whole of England and Wales is expected to be operational by about 1976. Under the auspices of the Home Office a police records data bank is being set up at Hendon to serve public authorities throughout the country. It will include records of stolen vehicles, finger prints and wanted and missing persons. It is intended to be a real-time system providing access to police authorities through a total of some 700 remote terminals by 1975. With further massive direct access store the system could be expanded to include further data such as person's previous history.

Integration of central government plans for data bases, and of the data bases themselves, is a particularly important issue since it is integration that brings efficiency accompanied by hazards for personal liberties. It has already been announced that the police data bank at Hendon will receive as direct input the vehicle registration information from the Swansea MOT vehicle licensing system. There is similar scope for integration of data between the Inland Revenue and the Department of Health and Social Security. These two departments have much in common in that the taxpayer and the recipient of benefit declare separately details of their personal circumstances, and there is obviously opportunity here for improved efficiency.

It is perhaps in local government even more than central government that the benefits of computer-based information systems and the associated hazards are likely to be most marked. Local government in England and Wales is currently undergoing a basic reorganization. The Local Government Act 1972 creates a new structure in England and Wales allocating the functions of local government among new authorities as from the 1st April 1974. This reorganization results in fewer and larger local government units and will, as an incidental consequence,make easier and more effective the computerization of local data files.

A number of local authorities have made substantial progress in the direction of computer operations, notably West Sussex known for its imaginative approach to computer based systems and Haringey Authority, which has planned in great detail the application of computers to the activities of its three constituent London Boroughs. Plans have recently been announced for the development in Leeds of a pilot Local Authority Management Information System (LAMIS). The LAMIS project is being undertaken by Leeds Corporation in collaboration with ICL and with the assistance of LAMSAC - the Local Authority Management Services and Computer Committee which has the responsibility of co-ordinating computer applications in the local government area. The data bank will at first hold land and property data, but it is intended that all types of local authority information will be added later.

In the commercial field the most sensitive personal information is that held by the Credit Bureaus and the Banks. So far Credit Bureaus in the UK have not transferred their data files to computers but the Banks have been particularly active and successful in their computing operations. The problems for Banks in the UK are perhaps simpler and more tractable than in the USA because the British 'Big Four' clearing Banks are national organizations with branches throughout the country and yet with relatively short lines of communication. Of the large clearing Banks the Lloyds branch accounting system has been particularly successful. Lloyds had all its branch accounting computerized by October 1970, well before decimalization day and is scheduled to have all branches on-line by 1973.

Another field in which the UK can claim to be making a major contribution is in data banks for the storage of medical information. Of the experimental systems, one of the most advanced is the Kings College (London) Hospital Computer Project concerned with the computerization of the medical record (4). The ultimate aim is a comprehensive one: to store and access on-line by computer the complete medical data for the patient, from whatever source, throughout his life. The first phase of the project dealing with the clinical record is now complete and covers three wards with some 60 beds. It is intended by the end of 1972 to link all wards with the computer system and to include not only the clinical record, but also laboratory and other investigative data. The nursing record has been developed alongside the medical record so that data may be shared between them. The system has proved its advantages in allowing nurses to spend more time with patients and less time in finding out what treatment has been prescribed.

A feature of the Kings College system is the care that has been taken to maintain security and privacy. Only those who are authorized, doctors and nurses, may interrogate the system, and they use special codes which allow access only to specific parts of the record.

All information is coded with the doctor's name so that it is known by which doctor, and at what time, the information was entered. Experience shows that the computer system is more reliable than the manual system for storing and retrieving records and the patients show a sympathetic interest. A valuable incidental advantage is that much more information on the statistics of illness and disease is available, which is an important aid to medical management and administration.

The computerized data banks described briefly above are, of course, only a small part of those planned or in operation in the UK. Nonetheless, it should be noted that there are as yet very few integrated on-line systems holding personal information in the UK that present a menace now to personal privacy.. It seems that those who have predicted an early realization of large integrated data bases have overlooked or underestimated the immense practical problems, particularly in software design, that have first to be solved. But it is not only software and hardware problems that impede the development. Perhaps even more difficult to overcome is the resistance to organizational changes. In order to establish a practical system a revolution in administrative method is demanded and it is clear that this may take many years to accomplish.

PRIVACY AND THE LAW IN THE UK

Whether or not the growth of computer data banks is the source of the anxiety or merely an adjunct of it, during the last few years there has been an active public concern in the UK with the problem of individual privacy and the law. An independent commentary on the relevant UK law which is particularly helpful is that prepared by 'Justice' the British Section of the International Commission of Jurists (5).

It is generally agreed that there is no general right to privacy provided by common law or statute law, civil or criminal, in England and Wales (6). If an individual suffers and invasion of privacy he has a remedy only if he can bring it within the ambit of one of the existing laws, for example a head of liability in tort, whose primary function is to provide a remedy for some other purpose. The possible causes of action are scattered and varied forming for the lawyer, let alone the private citizen, a confused patchwork of largely inadequate and incoherent remedies. The Younger Committee, whose Report we shall consider in detail later, found cause to remark: "One general observation may be made with confidence: that Great Britain has less in its law aimed specifically at invasion of privacy than any other country whose law we have examined."

As an example of the common law remedies which are available, the tort of trespass to land or to chattels may protect an infringement of privacy accompanied by unlawful interference with premises or property. The law of defamation may provide a remedy for those invasions of privacy which attend publication (in the legal sense) of defamatory statements about an individual. The action for negligence may be available where the act leading to a privacy infringement involves breach of a duty of care to the individual. Similarly, actions for nuisance, for infringement of copyright, for passing-off may play a part as remedies for certain types of privacy invasion. The action for breach of confidence is perhaps the most important in the privacy area and might well be developed to have the greatest relevance to protect the individual from disclosure of information held in data banks.

Of the statutory provisions which are available the Justice of the Peace Act 1361 (highly effective in the right circumstances) is distinguished by its honorable antiquity. The Act may be used to bind over to be of good behavior those who eavesdrop or offend in some similar fashion. The disclosure of private information by public officials is substantially protected by the criminal law. The Official Secrets Acts provide penalties for improper or unauthorized disclosure of official information or information entrusted in confidence to an official. Under the provisions of the Taxes Management Act 1970 (which consolidated previous legislation) Inland Revenue officials are required to declare that they will not disclose information received except in the execution of their duties, or the prosecution of an offence relating to Inland Revenue, or in such cases as may be required by law. The probity of public officials in the United Kingdom is acknowledged to be outstanding and a great deal of public confidence resides in their discretion and integrity. There have been few if any scandals in which private information in the hands of civil servants has been used or disclosed improperly. Indeed it is the reluctance of public departments to make information available which is worthy of comment. It is probably fair to say that public debate in the UK is hampered and impeded by the rigidity and stringency of the rules dealing with release of information to the public.

One channel for dealing with infringements of privacy by the central government is provided by the Parliamentary Commissioner for Administration (the Ombudsman) whose responsibility it is to investigate complaints made by the public who claim to have suffered from inefficient or improper administration. Though the Commissioner's powers are restricted, his intervention can be effective. A recent investigation into a complaint made about the administration of the 1971 decennial census illustrates the manner in which the Ombudsman may act (7).

There is no legislation in the United Kingdom specifically concerned with infringement of privacy by the storing, processing, manipulating or transmission of information in computerized form except that enacted to deal with the computer activities of the Post Office which establishes criminal penalties for the disclosure of private information to authorized persons. These penalties were first included in the Post Office (Data Processing Service) Act 1967, and similar provisions were incorporated in S.65 of the Post Office Act 1969 which imposes an obligation of secrecy in respect of information obtained in the course of the provision of data processing services (8).

LITERATURE AND CODES OF PRACTICE

Though there is little legislation concerned with computers there is a burgeoning literature on the problems of data banks and privacy, some of it tending to the science-fiction in its attitudes, but much of it a serious and worthwhile attempt to present solutions to an acknowledged problem. The Proceedings of the NCCL Workshop on the Data Bank Society have already been mentioned (1). Two further books are those by Malcolm Warner and Michael Stone (9) and Adrian Norman and James Martin (10). The Society of Conservative Lawyers has published two pamphlets (11)(12) the first of which is a detailed discussion of data banks and privacy and was the first publication to recommend statutory provision of print-out for defined categories of information held in public and private stores. Perhaps most interesting of all the publicly available documents are the verbatim reports in Hansard of the debates in the House of Commons and the House of Lords on the occasion of the introduction of the various private members' Bills. The debates in the Upper House (usually better reading than similar debates in the Commons) are particularly worthwhile and demonstrate a remarkable degree of technical understanding on the part of their Lordships.

Active in the privacy arena has been the British Computer Society, a young professional society with some 16000 members, which has played a vigorous part in public discussion of the ethical and legal problems. Recognizing that a fundamental safeguard for protection of computer-stored information must be the professional standards and integrity of those responsible for their design and operation, the Society has taken the initiative in establishing a Code of Conduct for its professional members (13). The principles of the Code are briefly stated: -

"A professional member of the British Computer Society

1. Will behave at all times with integrity. He will not <u>knowingly</u> lay claims to a level of competence that he does not possess and he will at all times exercise competence at least to the level he claims.
2. Will act with complete discretion when entrusted with confidential information.
3. Will act with strict impartiality when purporting to give independent advice and must disclose any relevant interest.
4. Will accept full responsibility for any work which he undertakes and will construct and deliver that which he purports to deliver.
5. Will not seek personal advantage to the detriment of the Society."

The Code is enforced by an investigatory and disciplinary procedure which includes as the ultimate sanction exclusion from the Society.

In the local government area Notes Guidance have been prepared by the computer panel of the Local Authorities Management Services and Computer Committee (14). The Notes are designed to serve as guidelines for local authorities in respect of the confidential data in their computer systems. They include recommendations on such topics as data collection and preparation, file organization, access by remote terminals, computer room and terminal usage procedures. The Notes pay particular attention to the placing of responsibility for security on the shoulders of a designated senior local government official (15).

PROPOSALS FOR NEW LEGISLATION

Accompanying the public concern with problems of privacy in general and data banks in particular, there have been a number of detailed proposals for new legislation introduced in Parliament as private members' Bills in the last few years. None of these Bills has become law, but they have had the effect of generating informed parliamentary discussion. The three most relevant to the data bank problem are the Right of Privacy Bill of Bryan Walden M.P. (16) a Bill based with one small exception on a draft Bill appended to the report in 1970 of the 'Justice' Committee (17), the Data Surveillance Bill of Kenneth Baker M.P. (18), and the Control of Personal Information Bill of Leslie Huckfield M.P. (19).

The Right of Privacy Bill is designed to provide a general right of privacy with a civil remedy for any "substantial and unreasonable infringement.... actionable at the suit of any person whose right of privacy has been so infringed." The right of privacy is defined as

> "the right of any person to be protected from intrusion upon himself, his home, his family, his relationships and communications with others, his property and his business affairs, including intrusion by -

(a) spying, prying, watching, or besetting;

(b) the unauthorized overhearing or recording of spoken words;

(c) the unauthorized making of visual images;

(d) the unauthorized reading or copying of documents;

(e) the unauthorized use or disclosure of confidential information, or of facts (including his name, identity or likeness) calculated to cause him distress, annoyance or embarrassment, or to place him in a false light;

(f) the unauthorized appropriation of his name, identity, or likeness for another's gain."

The Bill provides specific defenses for 'legitimate' intrusions, including (inter alia) those by consent, those in which public interest can be established, and where the defendent having exercised all reasonable care neither knew nor intended there to be any infringement of privacy.

There was much support for the Bill from all sides of the House of Commons but also disquiet concerning the effect of the Bill on freedom of speech particularly as represented by the activities of the Press. It was as a result of the Bill that the Government appointed the Younger Committee to undertake detailed consideration of the law relating to privacy.

The primary purpose of the Data Surveillance Bill is to establish a register of data banks operated on behalf of certain defined categories of individual or organization, for example persons offering to supply information about another person's credit. It would be an offense to fail to register a data bank to which the Bill applies. A weakness of the Bill is its wide definition of data bank -- "a computer which records and stores information." Any computer however small would fall within this definition and if registration were to be practicable the definition would have to be much narrower. The Bill proposes that the register should be available for examination by the public, except for those data banks operated by the police, the armed forces and security services, details of which would be kept in a separate part of the register not generally available for inspection. A print-out is to be supplied to an individual when information about him is first stored, and thereafter at his request on payment of a fee. A person receiving a print-out may apply for an order that any or all of the data should be expunged or amended on the ground that it is incorrect, unfair or out of date.

The Control of Personal Information Bill defines a data bank as "any store of information containing details of individuals", so it covers both computerized and non-computerized data banks. The Bill provides for a registering agency, which it terms a Data Bank Tribunal, a body with quasi-judicial powers and authority to "grant licenses on such terms and conditions, if any, as it shall think fit." The Tribunal would be backed up by a Data Bank Inspectorate with substantial powers of inspection and examination; the Chief Inspector being required to report to the Tribunal on the compliance of persons with the provisions of the Bill. It is proposed that initially operating licenses for data banks should be required only where they store records on 100,000 persons or more. There are powers to reduce this number later under specified conditions. An interesting and necessary provision is that which requires a person who regularly uses a store of greater than the minimum size wholly or partly outside the United Kingdom to obtain a license. The Bill also provides the Tribunal with powers to order print-outs, to exclude certain classes of information from the data bank, and certain categories of persons from access to it, to order erasure or correction of information on the grounds of inaccuracy, incompleteness, or irrelevance for the purposes for which it is stored, and to order notification of such erasure or correction to previous recipients of the information.

Both the Data Surveillance and Control of Personal Information Bills are helpful attempts to frame in statutory form the basic requirements that are likely in future to be incorporated in data bank legislation: the provision of a register, a supervisory body with substantial powers and wide discretion, creation of new civil remedies and new criminal sanctions, the right to know that information is stored and to verify its accuracy. Neither Bill is likely to become law but both contribute valuable elements to the public debate.

THE YOUNGER REPORT

It is worthwhile to consider fully the recently published Report of the Younger Committee on Privacy (20) since the Committee gave detailed study to the problems of data banks, and their recommendations are likely to determine the government attitude to the regulation of data banks in the next few years. As has already been mentioned the Committee was established in May 1970 in response to views expressed in the Second Reading debate in the House of Commons on the Right of Privacy Bill introduced by Mr. Brian Walden M.P. The terms of reference of the Committee was a follows:-

> "To consider whether legislation is necessary in order to give further protection to the individual citizen and to commercial and industrial interests against intrusion into privacy by private persons and organizations, or by companies, and to make recommendations."

A serious restriction on the Committee's work was the limitation of these terms of reference to the private sector, particularly since the Committee itself recognized that many of the public anxieties were prompted by the activities of Government departments. It is therefore important to note that the present Government and the previous administration both gave general undertakings to the effect that the recommendations made would be considered in relation to the activities of central and local government. It will be necessary to ensure that these undertakings are met in full. During the same period that the Younger Committee was meeting, an intergovernmental group under the leadership of the Home Office was conducting a comprehensive survey into the categories of information held or likely to be held by the computer systems of Government departments and the formulation of roles governing its storage and use. This survey, coupled with the Younger Report, should enable a policy to be prepared for the privacy of information held in Government data banks.

On the major question for which the Younger Committee was established - should there be legislation to introduce a general right of privacy in English and Scottish law? - the Committee were divided. The majority thought not, but two members in separately prepared minority reports argued forcefully that such legislation is desirable.

The Committee recommended that the law relating to breach of confidence be referred to the Law Commissions for England and Scotland with a view to its clarification and statement in legislative form, and that there should be a new tort of disclosure or other use of information unlawfully acquired. Those two recommendations may well prove to be the most important the Committee made. Strengthening the law of confidence, particularly if it provides remedies against a third party, and a new form of civil action against a person who discloses or uses information obtained by illegal means may not only be of assistance to those whose privacy is infringed by computer-stored information, but may also assist in providing a solution to another problem faced by the computer industry - the protection of private property in computer software.

An interesting feature of the Younger study is the evidence it provides about public concern with the data bank's threat to privacy. On the one hand the Committee says " of all the forms of invasion of privacy which have been cited in evidence to us that involving the use or misuse of computers has been the least supported in concrete terms." On the other hand submissions concerning computer information stores were received from a wide range of organizations and individuals, suggesting that it is fear and apprehension of the computer's power rather than actual present hazard that is the problem. Of these fears about the computer, the Committee identified the three uppermost in the public mind as: the computer's capacity to compile detailed personal profiles on large numbers of people; its power to correlate data, so extracting new and more dangerous information from traditional sources; the new opportunities for unauthorized access to personal information. As the Report says: "Almost all the more credible of these apprehensions, however, relate to the use or possible use of computers by central or local government, in particular to those in the hands of the police, the Inland Revenue and the health and social services, all of which are outside our terms of reference."

As a suggested means of control the Younger Committee formulated basic principles which in their view should apply to the handling of personal information by computers. The principles, ten in number, are as follows:

- Information should be regarded as held for specific purposes and not be used, without appropriate authorization, for other purposes.

- Access to information should be confined to those authorized to have it for the purpose for which it was supplied.

- The amount of information collected and held should be the minimum necessary for the achievement of the specified purpose.

- In computerized systems handling information for statistical purposes, adequate provision should be made in their design and programs for separating identities from the rest of the data.

- There should be arrangements whereby the subject could be told about the information concerning him.

- The level of security to be achieved by a system should be specified in advance by the user and should include precautions against the deliberate abuse or misuse of information.

- A monitoring system should be provided to facilitate the detection of any violation of the security system.

- In the design of information systems, periods should be specified beyond which the information should not be retained.

- Data held should be accurate. There should be machinery for the correction of inaccuracy and the updating of information.

- Care should be taken in coding value judgments.

The Committee recommended the voluntary acceptance of these simple and fundamental principles by computer manufacturers, operators and users whilst recognizing that such self-imposed restriction would require "that those concerned, whose primary loyalties are at present to the profitability of their commercial enterprises, should be aware of and accept the safeguards to privacy that society desired and that they should have the sense of moral responsibility to observe them at all times."

But voluntary acceptance of these principles needs to be strengthened by the professional discipline of those responsible for programming and operating computers, and the theme of professional responsibility recurs throughout the chapter of the Report dealing with computers. Whilst applauding the Code of Conduct introduced by the British Computer Society, the Committee have concluded that it would be premature to expect the establishment in the near future of effective voluntary professional discipline which could properly be endorsed by legislation. An original proposal is that each user of a computer should appoint a "responsible person" charged with ensuring that use conforms with the ten basic principles. A similar concept has been used with success in the United Kingdom in the Mines and Quarries legislation which requires the owner of a mine to appoint a mine manager as "responsible person" with responsibility for ensuring safe operation of the mine. An Appendix to the Younger Report examines the choice of person who might be designated as "responsible person" and his duties and powers of delegation.

The Committee considered that the time was not yet ripe for the detailed legislative provisions contained in the Data Surveillance and Control of Personal Information Bills through some scheme of registration, licensing and inspection on their lines may be appropriate at a future date. As an alternative at present the Committee recommend that the Government should legislate to provide itself with a Standing Commission, an independent body with members drawn from the computer world and outside, to keep under review the growth in and techniques of gathering personal information and processing it by computers. This Standing Commission should collect information about computerized information stores, review the principles of handling personal information, receive complaints about invasion of privacy by the users of computerized information stores, and make recommendations from time to time as it saw fit for legislative or other controls. The Commission should have the duty of making an annual report to Parliament thereby providing an opportunity for regular debate. Finally, and of great importance, the Younger Committee recommends that the Government should consider the desirability of including both the public and private sectors within the purview of this Standing Commission.

CONCLUSION

What conclusions may be drawn from United Kingdom experience so far in preparing safeguards for the right of privacy from computer data banks? The first point that emerges is that the rate of development of computerized systems is by no means as rapid as many predicted, or would have predicted, five years ago. The effect of exposing plans for data banks to the real world has been to restrain and moderate the development of these sophisticated storage and retrieval systems. For sheer practical reasons, the time when on-line multi-access integrated computer-stored file of sensitive information will present an imminent hazard is likely to be the 1980's rather than the 1970's. So it is tempting to suppose that there is five or ten years for adequate safeguards and remedies to be prepared. And the awakened public awareness - the apprehensions and fears of which the Younger Committee spoke - have had the salutary effect of stimulating appreciation of the issues amongst those responsible for the design and operation of data banks. So this is perhaps one problem of technology which can be avoided by preliminary planning.

Many will feel that the most effective safeguard that could be provided is the one rejected by a majority of the Younger Committee, the provision by legislation of a general right of privacy. Such a civil right of action would give the ordinary man a legal remedy for infringement of his privacy by computer once he knew that it had happened. And it would give teeth to the other recommendations of the Younger Committee. The ten basic principles are much more likely to be respected if their neglect might result in an action for damages.

The proposal to create a Standing Commission to review the techniques of handling personal information and to consider in detail the need for licensing and registration systems will meet with general sympathy. Faced with a rapidly evolving technology and a multitude of inchoate plans for data banks, many of which are several years away from implementation, the need is for a flexible approach which a Standing Commission would provide rather than detailed legislation which might impede the establishment of data banks rather than ensure their effective control. Once legislation is on the Statute Book the inertia of the parliamentary process means it may be a long time before amendment or repeal is likely. It becomes of prime importance to get the right legislation in the first place.

In the meantime the voluntary adoption by computer users, actively encouraged by public concern, of the basic principles for handling information by computers provides an interim solution. Application of those principles, together with the choice of a "responsible person", would vary from installation to installation depending on the size, the nature, and the purpose of the information which is held. A separately prepared, administrative code of practice incorporating the principles relevant to each installation, voluntarily adopted by each computer user with the active assistance of the professional societies would represent a valuable step forward.

But voluntary safeguards can be only a temporary measure. In the end legislation will no doubt be necessary to impose some form of statutory inspection and control. Though the hazards of data banks appear at present to be some way off, we are dealing with a technology that could surprise us all by the speed of its advance. Data bank technology is after all based on the binary digit which knows only two extreme values and which switches instantaneously from zero to one, from the incomplete to the complete. We may wake up one morning and find the last bug is removed and the hazards we fear have arrived. Perhaps we should add as the first of all principles the Chinese adage - "it is later than you think."

FOOTNOTES

(1) The Proceedings have since been published under the title "Privacy, Computers and You" by the National Computing Centre Limited.

(2) Report of the Committee on Privacy, (Cmnd. 5012) HMSO July 1972.

(3) Available from HMSO (1971).

(4) The author is indebted to Professor J. Anderson, Head of the Department of Medicine at Kings College Hospital Medical School, for an up-to-date account of the project.

(5) PRIVACY AND THE LAW,'Justice", Stevens and Sons, 1970.

(6) The United Kingdom comprises three separate jurisdictions, England and Wales, Scotland, and Northern Ireland. This paper is concerned primarily with the law of England and Wales.

(7) Third Report of the Parliamentary Commissioner for Administration, HMSO, June 1972.

(8) Section 65 of the Post Office Act 1969 is expressed as follows: -

> "(1) information obtained by a person in the course of the provision for another, by virtue of this Part of the Act, of data processing services or services connected therewith shall not without the consent of that other, be disclosed by the first-mentioned person except for the purpose of performing his duties in relation to those services or in such cases as may be required by law.
>
> (2) A person who discloses information in contravention of the foregoing sub-section shall be liable
>
> > (a) on conviction on indictment, to imprisonment for a term not exceeding two years or to a fine, or to both;
> > (b) on summary conviction, to a fine not exceeding 400."

(9) Malcolm Warner and Michael Stone, THE DATA BANK SOCIETY, George Allen and Unwin, 1970.

(10) Adrian R. Norman and James Martin, THE COMPUTERIZED SOCIETY, Prentice Hall

(11) 'Computers and Freedom', Conservative Research Department, 1968.

(12) 'The Price of Privacy', Conservative Research Department, 1971.

(13) This Code of Conduct was approved by the Council of the British Computer Society on 17th February 1971.

(14) "Computer Privacy: Notes of Guidance for Local Authorities", Local Authorities Management Services and Computer Committee, (March 1972).

(15) "One positive step which it is recommended that every local authority must take is formally to assign responsibility for security to a designated Chief Officer who in turn should designate appropriate officers to be responsible for the security of specific terminals........The Chief Officer may call upon specialist staff to advise him, but this will in no way relieve him of his overall responsibility."

(16) The Right of Privacy Bill, (November 1969), a Bill to establish a right of privacy, to make consequential amendments to the law of evidence and for connected purposes.

(17) See note (5) supra.

(18) The Data Surveillance Bill, (May 1969) a Bill to prevent the invasion of privacy through the misuse of computer information. A Bill with identical wording was introduced in the House of Lords by Lord Windlesham.

(19) The Control of Personal Information Bill, (February 1971), a Bill to establish a data bank tribunal and inspectorate; to provide for the licensing of data banks containing personal information, and to make other provisions to prevent the misuse of information stored in data banks.

(20) Report of the Committee on Privacy, (Cmnd. 5012), HMSO, July 1972.

DATA BANKS AND INDIVIDUAL PRIVACY: THE SITUATION IN THE GERMAN FEDERAL REPUBLIC

H.P. Gassmann

Organization for Economic Co-Operation & Development, Paris, France

General

Computerization of many private and public administrative functions is relatively advanced in the German Federal Republic. After the U.S., Germany ranks second in the number of computer systems installed, which by now is well over 12,000. While private enterprises in industry and banking have been the most dynamic to introduce sophisticated ADP techniques, departments of public administration also have been eager to use ADP for partial automation of their information processing. One reason for the relatively fast introduction of advanced ADP has been the chronic shortage of office manpower, which reflects the general situation in this country, having only 0.2% of the active population unemployed.

Nevertheless, no spectacular, large systems which could attract international attention do exist at present, although some are in the planning stage. This is due to the fact that West Germany is a Federal Republic, where the various states ("Laender") have a rather large autonomy, and most of the responsibility for the day-to-day administration rests with these Laender. Every Land has, to a varying degree, set up its own government data centre, and much routine ADP is performed there.

New Developments

Public attention to the Privacy Issues in West Germany has been so far mainly focussed on problems concerning data banks in the public sector. Indeed, much to the difference of what is happening in North America, the main concern in the field of infringement of data confidentiality seems to be centred not so much on private personal data records, but rather more on what could happen in the public sector. These concerns are prompted by two facts:

(1) the existence for many years of a compulsory population registration system (much on the lines with that in existence in Scandinavian countries), and

(2) the plans to introduce so-called "personal identification numbers".

The Laender are now moving to a more sophisticated use of ADP, and many have plans for administrative "information systems." It is here where the notion of "data Bank" begins to play an important role, and inevitably also the notion of protection of privacy, its counterpart. For instance, the Land Hessen has its own government data centre (HZD), and recently a plan has been submitted for a "planning and analysis information system" (HEPAS). Other Laender are also moving along these lines. For instance the Land Bavaria will convert the large computer systems which have been installed for the Olympic games in Munich for use as part of the Bavarian Land Information System. Common to all these systems is their orientation towards regional planning. It is hoped that with the existence of these systems, much better information for spatial, regional and urban planning will be available so that in the long run the expenses for these systems will be recovered through a better utilization and distribution of resources.

Private credit information systems have not been a large source of concern so far in West Germany. This may also be due to the fact that credit cards are not as widely used as in the U.S. German banks are planning to participate in a European-wide banking data communications network (Project SWIFT), but it is to be anticipated that most of the information on this network will not have personal record characteristics.

Legal Measures For Protection Of Sensitive Personal Records

Three concepts must be distinguished before entering this discussion:

1. Privacy: is a concept of a moral nature, varying widely from country to country, which depends directly on the value system of a particular society.

2. Data confidentiality: this relates to the legal and organizational measures taken by a country to protect sensitive confidential data - it is a concept of how well and efficiently personal data are protected.

3. Data Security: this concept relates to technical measures (hard and software) in the computer system and its terminals to make sure that no information can be accessed by unauthorized persons.

Most of the following relates to data protection and confidentiality. Privacy issues, such as whether a compulsory population registration system is good or not are not considered.

Prior to 1970, legislation did not specifically aim at privacy or data confidentiality issues. However, some laws implicitly protect the citizen in this respect: (1)*

Constitution (Basic Law)

Art. 1 - Man's dignity must not be violated

Art. 19, para.2 - Scope of personality or human freedom should not be limited to the point that may endanger their very nature

Art. 19, para.4 - Comprehensive protection of constitutional rights guaranteed by the courts against possible infringements of individual rights by the public authorities

Penal Code

Art. 298 - infringement of the confidential nature of the spoken word by unlawful recording, on any sound recording medium, of the words of others spoken otherwise than in public, or by putting such a recording to unlawful uses

Art. 299 - infringement of the privacy of correspondence

Art. 300 - breach of professional secrecy by doctors, dentists, chemists, barristers and solicitors, patent agents, notaries, advocates in criminal cases, qualified accountants, auditors or tax advisers, or by their professional staff or other persons engaged in such an activity with a view to entering the profession

Art. 353 - infringement of the confidential nature of the spoken word by officials

Art. 354 - infringement of the privacy of correspondence by Post Office officials

Art. 355 - infringement of the secrecy of telegrams by Post Office officials, or persons appointed as supervisors or staff of establishments designed for the public despatch of telegrams.

Statistics Act

Section 12 - Guarantees that administrative authorities observe the principle of secrecy in regard to statistical data

Fiscal Code

Section 22 - Secrecy in tax matters is inviolable.

Since 1970, the creation of the aforementioned Land government data centres has stimulated legislative action specifically aimed at data protection at State level. Four groups of Laender can be distinguished according to the degree of development of data protection measures:

1) A Data Protection Act has been enacted: Land Hessen

2) A Data Protection Act is being implemented: Rheinland-Pfalz, Hamburg, Nordrhein-Westfalen, or planned: Baden-Württemberg, Berlin

3) Data Protection rules exist in the framework of ADP organization laws: Baden-Württemberg, Bayern

4) No Data Protection measures in force or planned: four Laender, (Some of these do, however, have some internal administrative rules).

The proliferation of legislative initiatives at State level has prompted the preparation of a Data Protection Act at the Federal level, and consequently, legislative action at State level is at present being slowed down.

The most interesting case obviously is that of the Land of Hessen which proposed a Data Protection Act on 25th June, 1970, which has become effective on 7th October 1970. A copy of this Act is available from the author. The Act has two parts. Part I refers to Data Protection and Part II to the Data Protection Commissioner. In Part I the Act

---

*Numbers in parenthesis refer to footnotes at end of paper.

defines the scope of data protection which covers all records prepared for the purposes of Automatic Data Processing, all stored data and the results of processing such records. The Act clearly limits itself to data processed in public institutions and establishments under the jurisdiction of the Land. It specifies that persons responsible for the preparation, transmission and storage of data shall be prohibited from communicating or making available to other persons any information concerning the records, data and results gained during their activities. It also formulates a claim to data protection in the sense that if stored data is incorrect an aggrieved party may demand rectification. Another section of the Act deals specifically with data banks and information systems. It specifies some rules under which data contained in these information systems may be communicated and published when there is no legal prohibition of disclosure. The rights of the land parliament and local representative bodies to information are also stated in the sense that the various public ADP centres shall be bound to give to parliament, its president and parliamentary parties such information as they are entitled to receive.

In Part II of the Act the legal status, the secrecy requirements and the duties of the Data Protection Commissioner are specified. The independent position of this "ombudsman" is defined in the sense that he shall be exempt from direction. Even after completion of his term of office, he shall be bound to secrecy about the facts with which he has become acquainted during his activities. It will be the duty of the Commissioner to insure that the provisions of the act and other regulations governing the confidential handling of information provided by citizens are observed in the ADP centres. He has to inform the responsible authorities of any infringement committed and has to initiate measures for improving data protection. He also is supposed to observe the effects of ADP on the operation and powers of decision of the Land ADP centres and it is his duty to note whether they lead to any displacement in the distribution of powers among the Land's Constitutional bodies, among local administrations and between Land and local administration. He shall be entitled to initiate any measures he thinks fit to prevent such an effect. No doubt this "preventive legislation" has been necessary to insure that close cooperation is achieved between local and State government as far as automatic data processing is concerned.

This Act also provides for a right for every person to apply to the Commissioner is he considers that his rights have been infringed by ADP activities.

EXPERIENCE WITH THE HESSEN ACT

The Data Protection Commissioner is required to submit a yearly report on the results of his activity to the Land parliament and the Prime Minister. In the meantime, the first Commissioner, Mr. Birkelbach, has presented his first report (2). He has been appointed for the length of one legislation period, i.e., four years. His report gives a first comprehensive picture of the situation concerning data protection not only in the Land Hessen but also generally in the Federal Republic. On the suggestion of the Commissioner, the Land government has started, on July 6, 1971, a survey of all institutions in the public sector which produce records for automated data processing or which store and process data. These institutions have also been requested to report on the organizational, personal and technical procedures which are in force concerning data protection. It is interesting to note that of the 800 questionnaires, punch card types and other data carrier formats which have been reported to exist, more than 50% contain data related to persons with about 1500 different information items. In the various departments of administration these data are needed several times, so that they have gathered in multiple forms: for instance the family name, the first name, the date of birth and the address have been requested up to thirty times. Considering this multiple data collection, there remain about 300 different information items on persons, which need protection if they are combined in different ways. This protection is also needed for other information items relevant to physical objects without identification characteristics, which, however, become relevant for data protection if they are combined with personal records.

As was to be expected in such an innovation as the institutionalization of the Data Protection Commissioner, the first survey has not given a very clear picture of the situation. It is necessary that the institutions and organizations become more acquainted with this innovation. Many of the questionnaires have not been answered at all and some of them in a wrong way. Also, some misunderstandings have occurred; for instance, the institutions responsible for public health care found that the concept of the right to information of the Data Protection Commissioner was incompatible with the duty of the physician to keep all medical information confidential.

Clearly the legal interest of the commissioner to obtain an overview on all those branches of public administration which prepare records for data automation without processing these data themselves has been underestimated. The scope of the Data Protection Act does not cover the traditional or manual collection of data; however, data protection has to start with the preparation of records for ADP.

Nevertheless the report states that a strong concentration of data processing in the public sector of the Land Hessen in the government service bureau (HZD) and the communal and local service bureaux could be observed.

So far traditional organizational and personal safeguard measures for data protection have been employed. It seems that up to now the privacy of the citizen and the balance of power between legislative and executive has not yet been threatened by automation of administrative data systems. Despite a considerable degree of automation, the integration of data bases of various branches of public administration is not very much advanced yet. The traditional control procedures of local or state administrations are therefore sufficient against a misuse of these data. However, the survey also showed that the necessity of data protection for future developments has not yet been perceived and agreed to by everybody. The report continues with a detailed description of the tasks and the activities of the Data Commissioner, and at the end makes some suggestions as to improvement of data protection. Some highlights of these suggestions are as follows:

1. Avoidance of identification characteristics

In order to preclude the illegal identification of individuals from data contained in administrative information systems, it should be considered already at the stage of collection of data whether it is possible to avoid identification characteristics, such as name, birth date, and address, or if this is not possible, whether the identification of the individual could be made more difficult through a coding or other anonymization of data collected.

2. Separate storage

It is suggested that with especially sensitive information, the identification characteristics should be stored separately from the other data already at the source. The identification characteristics should be given to other branches of administration only if this is absolutely necessary on explicitly stated grounds. In the interest of data protection a loss of efficiency through this separate storage should be considered as justified.

3. Statistical data

It is suggested that for statistical purposes the principle should be maintained that no individual data be stored but only statistical results.

4. Right to access

The question of who is entitled to access to data has no been clearly defined yet. This question is especially relevant for data concerning the private sphere of individuals. The present experience on this is still limited and it is not yet possible to regulate this by law. On the other hand organizational procedures could be institutionalized, such as the need for a specific decision by the director of the department supplying a personal record to another administrative branch.

5. Public-private interface

The interface between public administration or governmental service bureaux and private institutions is ill-defined. If information from the public sector is given to private institutions, or private persons are allowed to access some data stored by public administration, special confidentiality measures are necessary.

6. The responsibility of the administrations

Contrary to many expectations it must be remembered that the full responsibility for the implementation of data protection lies with the branches of administration which are concerned with ADP. The special control function of the data protection commissioner should not narrow down this responsibility, on the contrary it should reinforce it.

7. Data security

The security of data is rather underdeveloped and sometimes existent only in parts. Especially in the following areas, there is a great need to improve the situation:

(a) storage and transport of record carriers
(b) erasure of obsolete data in a planned way
(c) the organization in the computer rooms.

In this context it is also mentioned that soon data communications (for instance in personal registration and police applications) will raise additional problems of data security. These must be considered in the present planning phase.

8. The citizen's desire for information

The act does not provide a right to information of the citizen. However, since the administration is interested in the storage of correct data, the government should give instructions that the desire of the citizen to inspect the data concerning him should be fulfilled if possible or whenever this is possible.

Finally, attention is drawn to the relationship between information technology and democracy. In the industrially advanced countries this will rapidly become an issue of the political and societal development. If acceptable solutions are to be found to this issue, it is necessary that scientific analysis and research of the issues involved be made. The promotion by the government of information technology should, therefore, not stop at ADP technique and usage, but should, as rapidily as possible, comprise also the social sciences.

DEVELOPMENTS AT THE FEDERAL LEVEL

Three initiatives exist at present at the federal level.

(a) The Federal Government has presented a bill concerning population registration on October 4th, 1971 to the Parliament (Bundestag) with the main objective to create federal rules for the planned population information systems. A salient feature of this bill is the proposal of the attribution of a personal identification number to all German citizens and all persons living in the Federal Republic. The personal identifier should consist of 12 digits:

2 for day)
2 for month) of birth
2 for year)
1 for century and sex
4 serial numbers for persons of same sex born the same day
1 check digit

Para. 18 of this bill contains a right of a person to inspect all data collected about him, and a right for improvement if these data are incorrect.

Para. 16 states that transmittal of personal data to other departments of administration is only permitted if the need for knowledge of this information for their own administrative purposes can be demonstrated. Data security measures are requested in case of use of ADP and telecommunications; a record stating the receiver of the information, the type of data given and the time of transmittal is also imposed. Private persons are only allowed to request the name, address and date of moving into a dwelling of a given individual. There will exist a possibility for a person to prevent that information on him being given to third persons.

(b) An interparliamentary working group has introduced a bill on December 2nd, 1971, concerning data protection on the federal level. This bill contains almost all the requirements which have been suggested for a federal law and is aimed at protecting personal records stored in both public and private data systems.

A person should be informed when a personal record is entered the first time into the data system. There should also be a right of information of a person on all his personal data, as well as on the record of transmittal. All transmittal of personal data should be regulated.

(c) The Federal Ministry of the Interior has prepared a federal data protection bill which in some respects follows the bill described in (b). It concerns protection of personal records stored:

(i) in the public sector
(ii) in the private sector for internal purposes, and
(iii) in the private sector for transmittal to 3rd persons.

Its scope extends to ADP as well as manual record handling. A right to correction of wrong data is provided for as well as a request for erasure of personal information if it has been obtained against stated rules.

In the case of private data bases which have an objective of collecting and storing personal records for sale or transmittal to third persons, the bill contains more stringent provisions. A persons on whom a record is being collected has to be notified, and he has a right to see his record, on payment of costs incurred by the data collecting institution (computer printout, for example). Information has to be erased if not obtained according to stated rules. Upon request of the interested person, it must be erased after 5 years of collection. Collection of personal information is permitted only through direct questioning of a given person, or through processing of documents provided by this person (questionnaires, etc.), unless he explicitly permits data collection in a different form.

A private organization with an objective of collecting and storing individual data for sale to 3rd persons must register with supervising authorities who are nominated by Land governments. The registration should comprise:

1. Name of enterprise of person
2. Address
3. Objective of organization
4. Specification of technical means
5. " of ADP software and programs
6. Specification of data security measures

All these legal initiatives at federal level have not yet come to a conclusion so far, but it is to be expected that they will eventually form a comprehensive framework for effective data protection in the Federal Republic.

CONCLUSIONS

The Federal Republic is among the countries where file automation is well advanced. Public concern with the protection of individual records has prompted several legal initiatives, both at state and federal level to ensure that advanced ADP utilization and the introduction of personal identifiers will not result in a loss of privacy of German citizens and an increase of information pollution. Given the particular legal and social situation in the Federal Republic, it is not to be expected that the planned or enacted laws could be adopted by other countries. Rather, the concepts behind the legal framework for data protection should be carefully studied and might provide useful guideposts in tackling this extremely complex field of the protection of personal privacy in a mass society, where demands on the individual to provide personal information constantly increase.

FOOTNOTES

(1) See German reply to OECD questionnaire (DAS/SPR/70.10/01) 25-9-70.

(2) Document 7/1495, 29 March, 1972 of Hessischer Landtag

SOCIAL CONCERNS

CHAIRMAN:

William E. Hanna, Jr., Director of the Bureau of Data Processing,
Social Security Administration
Baltimore, Maryland, U.S.A.

PROGRAM COORDINATOR:

Dr. Peter E. Jackson
Bell Telephone Laboratories
Holmdel, New Jersey, U.S.A.

PAPERS:

Dr. Robert H. Kupperman and Richard H. Wilcox, Office of Emergency Preparedness, Washington, D.C., U.S.A., *"EMISARI – An On-Line Management System in a Dynamic Environment"*

Donn B. Parker, Stanford Research Institute, Menlo Park, California, U.S.A., *"The Nature of Computer-Related Crime"*

Gerald A. Petersen, National Oceanic & Atmospheric Administration, Silver Spring, Maryland U.S.A. *"AFOS: A Program for National Weather Service Field Automation"*

James W. Evans, Program Analyst, and Robert A. Knisely, Director, Division of Community Management Systems, U.S. Department of Housing and Urban Development, Washington, D.C., U.S.A., *"Integrated Municipal Information Systems: Some Potential Impacts"*

DISCUSSANTS:

John F. Grady, Consultant
Medford, Oregon, U.S.A.

Dr. Roy Herrmann, President, The Institute for Socio-Economic Studies, Ltd.
Washington, D.C., U.S.A.

Professor Mischa Schwartz, Department of Electrical Engineering, Polytechnic Institute of Brooklyn,
Brooklyn, New York, U.S.A.

THEME: The diverse and virtually unlimited potential for application, good or bad, of computers and data communications to man in his role as a social being.

EMISARI—AN ON-LINE MANAGEMENT SYSTEM IN A DYNAMIC ENVIRONMENT

Robert H. Kupperman and Richard H. Wilcox
Office of Emergency Preparedness
Washington, D.C., U.S.A.

Abstract: When President Nixon declared a 90-day wage-price freeze in August, 1971, he assigned administrative responsibilities for it to the Office of Emergency Preparedness (OEP). Resulting urgent dynamic requirements for exchanging transient status information between Washington and a nationwide field organization led OEP to implement an unusual computer-based, on-line interactive management information system called EMISARI. Design philosophy emphasized communication among managers at multiple levels; variations in data urgency; system flexibility to permit operational evolution; and immediate usability by non-computer people without needing extensive training or documentation. The system was placed in operation quite rapidly, evolved successfully throughout the freeze, and was heavily used by both regional and headquarters personnel. Specific capabilities and characteristics of the system are described in the full paper, and results of operating experience are related.

On August 15, 1971, President Nixon announced a 90-day general freeze on wages and prices in the United States. He also stated that implementation and administration of the freeze would be managed by the Office of Emergency Preparedness (OEP), an arm of his Executive Office.

Although OEP has only about 350 permanent staff members, they represent a wide variety of expertise and a small but effective field organization distributed throughout the country. In addition, a long-standing responsibility for coordination of all Federal aid during and following natural disasters has given OEP strong capabilities for (a) rapid response to the unexpected, (b) smooth interaction with other agencies, and (c) fluent communications with the public. All of these were of critical importance to effective freeze administration.

On very short notice OEP modified its field organization from eight to ten regions, and augmented both Regional and National Office staffs by borrowing personnel from other agencies. New organizational responsibilities were assigned to recommend and interpret Cost of Living Council rulings, to consider requests for exceptions and exemptions, to monitor public reaction and compliance, and to look into complaints of violations. Obviously, a major requirement developed for more and faster communications, particularly between the National Office and the ten Regional Centers.

First of all, it was important that all citizens receive timely and valid information concerning details of the freeze and its impact. It was equally important that the requests and complaints of citizen receive prompt and appropriate attention from the government. The Regional Centers bore the primary burden of achieving this effective communication with the public. But to accomplish it successfully, they had to receive consistent and coordinated policy guidance from the Cost of Living Council, as augmented and interpreted by OEP Headquarters, and OEP and the Council in turn had to receive current status information and problem statements from the Centers. Some inter-Center communications were also desirable to ensure consistent handling of comparable problems.

The initial communications load between the OEP National Office and Regional Centers was handled by additional but conventional telephone, teletype, and facsimile equipment. These media served their purpose well and continued in heavy use throughout the freeze. However, problems of information organization, updating, retrieval, and analysis generated by the sheer magnitude of the operation led to introduction of two complementary computer-based information systems to assist in meeting the urgent requirements of OEP managers, both national and regional.

The primary computer-based system was conceived, designed, implemented, and operated by OEP personnel, using the OEP Univac 1108 computer. In recognition of its functions and its role, it was named EMISARI, an acronym for Emergency Management Information System and Reference Index. It represented contemporary technology by being not only "on-line" but "interactive," so that a user could determine what pertinent items were on file and then retrieve them in a display tailored to his needs or convenience. Furthermore, he could obtain information whenever he needed it, without depending upon availability of some person at the other end of a phone line, or waiting while a teletype message was prepared, transmitted, delivered, considered, and answered.

While EMISARI handled the immediate operational reporting and guidance aspects of freeze administration, a separate and independent system was also developed by OEP to handle the detailed content of violations and exemption activities. This separate system, which was essential for complete freeze documentation and as a basis for subsequent analysis, was operated in very fine manner by the Census Bureau. However, because it was a completely batch-oriented system of non-trivial but conventional design, it will not be treated further in this paper.

The specific form of EMISARI was strongly conditioned by our personal philosophies (biases?) concerning management information systems. First, a true management information system is much more than a mere reporting system to top management. Reporting systems only obtain information at lower organizational levels, collect and perhaps organize it, and forward it to potential users at higher levels. There are two serious shortcomings of reporting systems: (1) the people who are data sources seldom receive any feedback to find out whether their

---

*The views and opinions expressed in this paper are those of the authors and do not necessarily reflect official policy of the U. S. Government.

submissions are appropriate--or indeed, even used--and (ii) failing such feedback or any other useful (to them) output from the system, the source personnel tend to become haphazard in the quality of data which they submit. In contrast, an effective management information system is a communications vehicle among managers at all levels. Thus, while it includes data reporting "up the line," it also provides for dissemination of feedback and policy guidance "down the line," and furthermore for coordination and cooperation among managers at parallel levels.

Second, information appropriate for support of management decisions involves a wide spectrum of requirements for timeliness. On the one extreme is the true emergency situation which demands immediate action, such as a fire alarm. In some management quarters a telephone call from a Congressman is an emergency of similar magnitude, requiring similar immediate response of some kind. At the other extreme of timeliness, permanent documentation is required not only for historical purposes but for careful analysis upon which subsequent long-range improvements can be based. This is the justification for many records and filing cabinets. In between these two extremes there exists a wide variety of intermediate degrees of urgency, and any effective management system must take them into account appropriately.

Finally, survival of any operational system demands flexibility. Thus the capability for relatively simple modification must be a fundamental design criterion for a management information system. This requirement comes about for three reasons. For one thing, no matter how experienced and wise the designer is, he simply cannot anticipate all problems that are going to occur. Bottlenecks will develop where they were not anticipated; conflicts among various components of the system will inevitably show up; some parts just will not work the way they were intended to. Thus it must be possible to "de-bug" the system to correct such problems. In addition, introduction of the system will inevitably perturb the activity into which it is placed. As the users gain experience, they will change their ways of using the system and, in fact, of doing their jobs. Thus user requirements will change as experience with the system is gained, and it must be possible to modify the system in order to track these changes in user requirements. Third, the overall environment in which the users operate will inevitably change with time, and thus the supra-system consisting of the management information system per se together with its users must be able to adapt, or it must perish.

In addition to reflecting the effects of these philosophies, the design of EMISARI was influenced by some relatively unusual conditions surrounding the wage-price freeze itself. For example, the operational environment to be expected was almost completely unknown, so that user requirements were impossible to define in the usual sense. In fact, one major purpose of the system was to help the Cost of Living Council find out what kind of problems would occur and what kind of information would be needed subsequently. For example, they were interested in learning what questions the public would ask, where significant violations would occur, what inequities were coming about, whether or not the populace would cooperate, and generally what was happening and what needed to be done about it. Another unconventional feature introduced by the wage-price freeze was the exceedingly tight time constraint. Each week that was allowed to elapse represented passage of 8% of the total freeze duration. The Cost of Living Council could not wait for conventional systems analyses to be performed, or for thorough system design and pre-testing to take place. Furthermore, there was no time to train users by conventional means or to prepare extensive user manuals for their assistance. Clearly, conventional reporting and communications systems were inadequate to the task presented.

The lack of opportunity for conventional training dictated a "self-teaching" approach for use of whatever system was developed. This was the reason behind initial selection of an on-line, interactive computer-based system, so that inexperienced users could get immediate feedback to their efforts at system use and thus correct their errors or modify their techniques accordingly. In particular, an approach was adopted whereby at each step the user was presented with a list of the choices available to him, from which he selected the one he preferred. A logical diagram of the complete set of system choices is shown in Figure 1; at any given step the user was presented automatically with the list of choices shown in the appropriate block. However, this listing of available choices could be suppressed by the experienced user in order to save time.

To provide the all important flexibility for rapid programming and convenient modifications, a greatly extended version of the Dartmouth Basic language, developed by Language Systems Development, Inc. (LSD) and called XBASIC, was chosen as the medium for programming. This proved subsequently to have been a very wise choice and the language lived up completely to its expectations for both power and simplicity of use and change.

Reporting "up-the-line" was provided in several forms. Quantitative values (called "estimates") could be provided for individual items of interest. Collections of related estimates (called "programs") could be called as units, and--for highly organized data--extensive two dimensional tables consisting of related programs were available. To state this in technical terms, quantitative reports could be organized as scalers, vectors, or matrices. Textual messages could be inserted also to provide qualitative commentary on the reported data, or they could be used independently of reports to provide for exchange of information among users of the system.

"Down-the-line" guidance was introduced in the form of several ad hoc files which had some special features. One file contained brief summaries, in question-and-answer form, of the several hundred rulings and interpretations issued by the Cost of Living Council. Another file contained brief summaries of the wide variety of requests for exceptions and exemptions which were denied. Later on a listing of pertinent news items, as prepared for a daily briefing to the Director of OEP, was introduced for the benefit of Regional Directors. Each of these several policy guidance files could be searched on the basis of "key words." By entering any word or phrase concerning a subject of interest, such as "tuition" or "rent," a user could obtain a listing of all entries which contained his stated interest. The key word entries were roots; that is, the key word "rent" would successfully locate also all entries containing such words as "rents," "rental," and also--sometimes disconcertingly --"apprentice." The user could employ a pair of independent key words to define his interest better (logical "and"), and he had the option of searching the full text of all items in the file, or only their short titles. This key word search feature proved to be both popular and effective. The system also recorded all key words tried, and whether or not the attempts were successful, in order to suggest subject

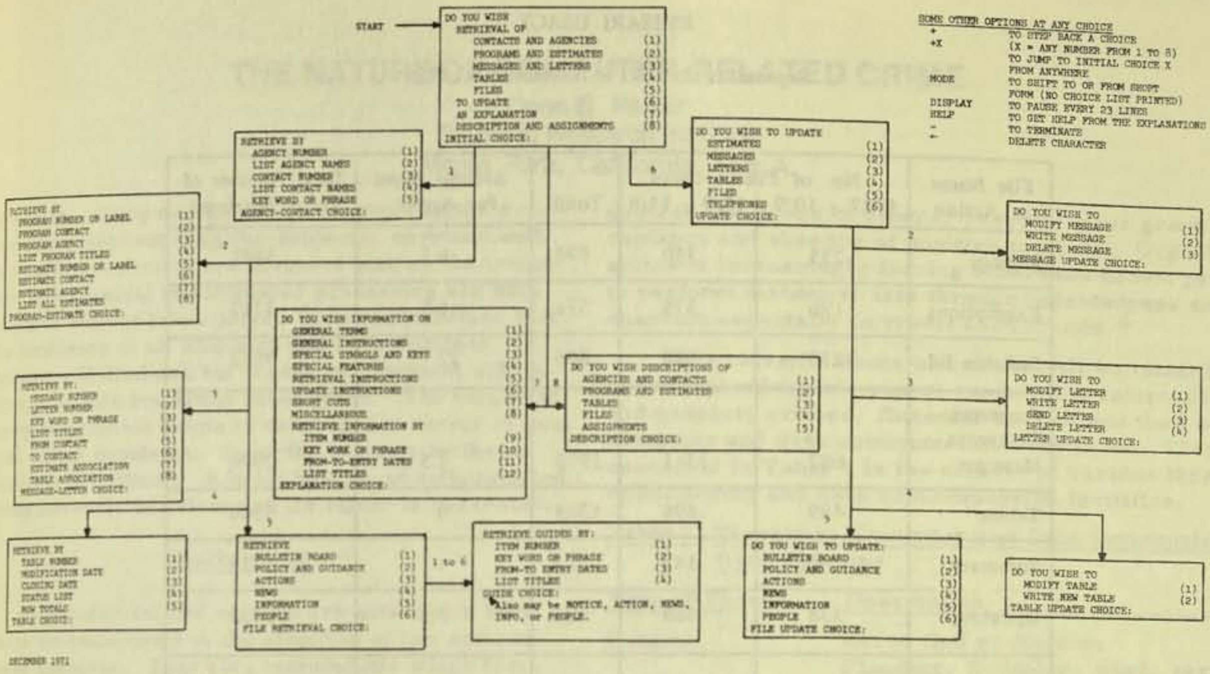


Figure 1 - Logic Diagram of EMISARI Access

areas for which guidance from the Council was desired (i.e., users sought information) but was apparently unavailable (or was inadequately listed).

Additional special "system reference" files made the various data and information exchange features of EMISARI more useful and usable. These files listed all formally designated system participants ("contacts"), the organizational elements ("agencies") with which they were associated, their office and home telephone numbers, and--particularly important--the portions of the system for which they were responsible. Each estimate, program, or table had some specific person named as being responsible for its status. Each critical location (e.g., a Regional Center), had a primary system contact listed. People with pertinent general responsibilities were also included, such as the System Project Manager, Regional Directors, and agency top management. These features provided much more than a convenient "classified directory"--they helped to pinpoint current responsibilities in a dynamic environment, and they identified a human being to call when questions or trouble developed. Furthermore, to facilitate such "non-standard" information flow, messages up to about 50 words in length could be placed in the system by any contact listed, addressed to any other contact or to all participants.

Many of the intra-system messages (and some desperate extra-system phone calls) were "addressed" to the system operations Monitor. This Monitor had some standard functions to perform, such as checking incoming weekly reports and calling for corrections when necessary, periodically copying primary data on to off-line "history tapes" for permanent retention, and keeping the file of system participants (and their private access codes) up to date. But the most important Monitor function was that of serving as central point of contact for all questions and procedures concerning use of the system and of the on-line terminals which provided the human-computer interface. The wage price freeze represented a highly dynamic environment, involving many non-computer personnel suddenly "drafted" from other agencies, and with system operators spread from Atlanta to Seattle. Thus the stable presence of an informed, sympathetic, and accessible human Monitor to answer questions and provide a clearinghouse for system status information was clearly one of the essential ingredients in successful system operation.

Once the basic system philosophy and requirements were established, the initial form of EMISARI was "on-the-air" within one week, thanks to some prior OEP experiments with computer conferencing and Delphi studies[1,2] This first version of EMISARI contained about 2500 lines of XBASIC programming. It provided reporting of individual items, a single guidelines file of initial Council rulings with key word search, a listing of system participants and their agencies, and a capability for associated messages. This initial system continued to evolve throughout the remainder of the freeze, with addition of such features as the data tables, the additional policy files, and some extensive reprogramming for improved efficiency as the size grew larger. It was still evolving when the freeze ended--a tribute to the flexibility of design and to the learning capability and imagination of all the system users.

In the eyes of many observers, the basic use of EMISARI throughout the freeze period was as the primary source of information for the weekly OEP report to the Cost of Living Council. However, system records collected automatically after the first month showed that even more use was made for other purposes, as shown in Table I. Overall, during the 10 week period recorded, the system was accessed about 900 times for purposes of entering data for reports. But the three policy guidance files, for rulings, exemptions, and the bulletin board (news), were accessed a total of more than 1900 times, and the individual estimates and textual messages were looked at an additional more-than-1900 times.

Discussions with various managers revealed several forms of usage. For example, both Regional

EMISARI USAGE

September 2, 1971 - November 8, 1971

File Name or Action	No. of File Accesses 9/27 - 10/9	10/9 - 11/8	Total	Average Items Per Access	Total Number of Items Retrieved
Rulings	235	359	594	6.5	3841
Exemptions	196	375	571	6.5	3712
Bulletin Bd.	270	536	806	12	9672
Programs, Estimates, Messages	662	1311	1973	1.5	2960
Tables	489	896	1385	4	5540
Summary	70	142	212	1	212
Updating	338	563	901	3.5	3153
Totals	2260	4182	6442	—	29,090

[1] Measured automatically by computer

[2] Estimated from Samples

TABLE I

and National Office personnel used the system contents as a basis for getting the latest numbers for inclusion in frequent public speeches. One Regional Director indicated that when he had to meet a special interest group, such as insurance executives, he would get a computer terminal printout of all responses to the appropriate key word (e.g., "insurance") to carry along in his briefcase; it not only helped him prepare for the meeting, but its very existence illustrated impressively to the group how well he was acquainted with the situation. The system was used both to answer difficult or unusual inquiries from the public, and to respond to specific management queries. During the later part of the freeze it was adopted in a secondary reporting mode as the sole basis for monitoring of the administrative situation. For example, Regions entered daily numbers for such items as quantities of correspondence received, disposed of, and remaining in backlog, so that National Office personnel could simply observe these numbers whenever they needed the appropriate information.

The overall system received its biggest genuine compliment when the Internal Revenue Service, in preparing to assume responsibility for administration of the post-freeze stabilization program, concluded that it could not satisfy Cost of Living Council requirements without adopting a system like EMISARI. Accordingly, OEP designed and provided to them a specially tailored version called IRMIS (Internal Revenue Management Information Systems).

Based on experience gained with EMISARI during the wage-price freeze, OEP Regional Directors suggested a variety of applications for "standard" OEP activities, particularly those pertaining to disaster assistance and crisis management. Thus EMISARI represents an extensive new OEP capability, developed and tested under fire, and ready to assist in new applications and benefits for the country.

---

1. M. Turoff, "Delphi and Its Potential Impact on Information Systems," Proc. 1971 FJCC, Vol. 39, AFIPS Press.

2. M. Turoff, "'Party-Line' and 'Discussion': Computerized Conference Systems," Proc. ICCC, 1972.

THE NATURE OF COMPUTER-RELATED CRIME

Donn B. Parker
Stanford Research Institute
Menlo Park, California, U.S.A.

Abstract. Computer and data communications systems are fast becoming the subjects, environments and tools in criminal and disputed acts as the transition from manual to automated processing and data communications takes place. It is hypothesized that the incidence of all financial and informational crimes will diminish but losses per incident will increase as this transition takes place. The nature of computer-related crime is described in terms of general crime problems. Security strategy by threat analysis is outlined. A brief history of authenticated computer-related crime and its future is presented.

Background

Computer-related crimes are acts which have come or could come to the attention of law enforcement agencies. They also include acts which form the basis for civil disputes. Computer and data communication systems can be the subjects of dishonest or disputed acts. They can form an environment in which an act occurs, or they can be used as tools in the perpetration of dishonest or disputed acts. Precise, uniform definitions of crime types are nonexistent in the U. S. A. Each state has its own laws and definitions of crime. Some other countries are in better shape such as the United Kingdom with the Theft Act of 1968 and Criminal Damage Act of 1971. In the U. S. A., computer-related crime can be included under fraud, theft, larceny, forgery, counterfeiting, defalcation, peculation, extortion, espionage, embezzlement, robbery, burglary, conspiracy or vandalism.

Computer-related crimes exclude acts involving data which is not meant for computer processing, such as most credit card fraud and check forgery--the exception is MICR coding forgery. A separate but related subject is privacy rights. They are not discussed here insofar as the current debates over constitutional rights are concerned. That is a subject for constitutional law, not information processing technology.[1] Invasion of privacy is considered to be an information crime within the computer-related crime context only when specific laws exist.

Many types of computer-related crimes are included in the context of white-collar crime. White-collar crime has been the subject of considerable study by social scientists since the term was coined by Edwin H. Sutherland in 1939.[2] White-collar crime has been defined for the science of criminalistics as follows: "Any endeavor or practice involving the stifling of free enterprise or the promoting of unfair competition; a breach of trust against an individual or an institution; a violation of occupational conduct; or the jeopardizing of consumers and clientele."[3]

Motivation to perpetrate computer crimes is an important but elusive concern. Motives include revenge, competition, politics, challenge to ability, power, wealth, avoidance of harm, sympathy to the desires and needs of others, respect, peer group acceptance and absence of positive motives. Organized crime is increasingly forcing otherwise honest people to perform dishonest acts through indebtedness and coercion especially in credit card frauds.[4]

Four classifications of crimes will be identified and discussed in this paper: vandalism, information and property crimes, financial crimes and theft of computer and data communication services. They are described in Table 1 in the context of various threats to computing and data communication facilities.

Table 1 Threats to Computer and Data Communications Systems

Type of Threat	Description
Natural	Act of God or Nature; Flooding, lightning, wind, earthquake, acts of animals, extraterrestrial radiation
Human errors & incompetence	Any acts imaginable; Many injurious results
Vandalism	Intentional damage, destruction, defacement of property; loss of capacity or ability of a system to function properly or at all
Financial crime	Embezzlement, theft, fraud, robbery, burglary, conspiracy, forgery, extortion; Direct financial losses in the form of credit or negotiable instruments
Informational and property crime	Theft, fraud, robbery, burglary, conspiracy, forgery, extortion, espionage, invasion of privacy; Loss of property and information including hardware, software, data, communication lines, documents, materials and supplies
Theft of computer and communication services	Theft, fraud, conspiracy, forgery; Unauthorized use of computer time, computer or communication facilities, software or data
Violence on people	Murder, bodily injury, mental anguish; Intentional use or control of computer process-control systems. Intentional alteration of product and process design.

Protection against Computer-Related Crime

The study of computer-related crime is important for the security of computer and data communications systems. Adequate and cost-effective security should be developed by starting with a knowledge of the threats to these systems. Threats should not be confused with methods of penetration of systems such as computer facility accesses, software trapdoors, masquerading, piggybacking, password detection and wiretapping.[5, 6] These are actually the technical methods of penetration and infiltration. Threats are the potential and actual actions of people and acts of God (Nature). The nature of people-produced threats includes circumstances of people's actions, their ability to act, the procedures they use and the technical methods they can employ. It is important to take the posture of looking outward through the external interfaces of the systems to the surrounding environment to successfully discover and analyze threats. Poking about in the system itself to find weaknesses and theorizing points of unintended penetration with little knowledge of the threats will only lead to serious discontinuities and inconsistencies in security.

For example, a technical study of security needs based on unintended methods of system penetration might suggest that magnetometers should be installed to detect magnets being carried in entrances to areas where data is stored and used in magnetically recorded form on tapes and disks. A threat analysis of this situation would show that there is only one verified, published case of the intentional use of magnets to destroy data done by a group of political radicals at a time when the computer center attacked was not in operation. The only few other suspected cases appear to be accidents. It might be concluded in this example case that external radical elements don't represent a threat, and the facility has no history of vandalism from internal employees. A person entering the areas could use much simpler and surer methods of damaging the tapes and disks beyond data recovery. Accidental creation of magnetic fields near enough to tapes and disks is a remote possibility. Simple signs warning of the danger and periodic reminders to workers in the controlled areas as well as maintaining the areas under access control is sufficient. The cost of installing magnetometers could be found to be unwarranted in this case relative to the threat, chances of loss, extent of other security measures and value of the data being protected.[7]

Limiting the strategy for developing security to penetration analysis and evaluation of assets to be protected is not adequate. For example, it may not take into account that multiple targets and combinations of penetration methods including bribery of employees may be involved. The owner's evaluation of the various assets to be protected may not coincide at all with the values placed on them by potentially dishonest people.

Security Strategy by Threat Analysis

1. Assets. Identify assets to be protected by who would want to destroy, change, steal or use them. Assets include staff, reputation, credit, negotiable instruments, software, data, hardware, use of facilities, materials, supplies and documents. Evaluate assets in terms of their value in potential threat situations:

   a. Perpetrator takes the asset for his own or others' exclusive use.
   b. Perpetrator takes the asset for his own or others' nonexclusive use (owner retains a copy).
   c. Perpetrator makes the asset of no further value to anybody temporarily or permanently.

2. Deterrence. Identify and reduce the population of potential perpetrators and their desire to perpetrate dishonest and injurious acts.

   a. Create a high moral climate and dedication to protection of assets.
   b. Minimize visibility of assets consistent with advertising needs.
   c. Maximize visibility of security short of creating interesting penetration challenges.
   d. Create an image of greater security/assets ratio than other computer systems. In other words, reduce your target profile relative to adjacent targets.

3. Statutory protection. Be aware of pertinent laws relating to identified threats and assets in your particular jurisdictions.

4. Prevention. Create an integrated set of barriers consistent with: potential threats, assets to be protected, operational efficiency and ability to maintain the barriers on a continuing basis.

   a. Allocate and fix responsibilities for security at all times and all points of penetration.
   b. Isolate assets to be protected.
   c. Maintain and test by auditing the integrity of the system identified as:
      - Hardware, software, data, materials and supplies
      - Physical environment
      - Staff, operating procedures, documentation
      - Communications

5. Detection. Increase the probability that any anomalies will be discovered and causes identified in documented form with legal-evidential quality where applicable.

6. Minimize potential losses through insurance and bonding.

History of Computer-Related Crime

The first Federal criminal case involving the use of a computer occurred in 1966.[8] A young programmer working for a software firm was responsible for programming and occasional operation of an IBM 1401 computer at a bank where he also had his checking account. He placed a change in the program to ignore his account when checking for overdrawn accounts in exception reporting. His plan was to leave the change in the program for only three days during which he knew his account was going to be overdrawn by $300. Four months later his account was still overdrawn by $1,352. The change was still in the program, and the computer failed one day, resulting in manual processing which uncovered the overdraft. The FBI was called to investigate and confronted the programmer with the evidence. He readily admitted his guilt, made restitution and was given a suspended sentence on two counts of altering bank records.

It is indicative of the way computer technology is changing the methods of crime to note that this was the first case in banking history that a nonemployee of a bank was ever indicted on charges of altering bank records. The programmer in this case fits the white-collar criminal mold by never having been in trouble with the law before, only intending to borrow for a short time, being highly apologetic over the trouble he caused and unable to resist temptations caused by lack of proper separation of sensitive job functions. The normal control functions of the bank failed to verify that the daily account postings showing overdrafts actually appeared in the overdraft exception report. It was only the unexpected change in processing methods which caused the discovery of wrongdoing.

In 1964 a programmer was in the habit of taking listings of programs he was maintaining home with him to work on at night, with the consent of his supervisor. He and another person attempted to sell copies of 75 to 100 programs to another company, a client of his employer, for $5 million.[9] The other company went along with the negotiations until there was enough evidence to indicate that the programmer was performing an illegal act. He was arrested, convicted of felony theft in the state of Texas, and sentenced to five years in prison. He appealed the conviction on the basis that the trial was unfair because the programs had never been introduced into evidence and did not constitute corporeal personal property and as such could not be the subject of theft. The Court for Criminal Appeals of Texas affirmed the conviction on the basis that the programs were available to all officers of the court and jury and were adequately described and identified by witnesses and that the law encompasses computer programs as property in relation to the crime of theft in stating, "... all writings of every description, provided such property possess any ascertainable value." Appeal was also made to the United States Court of Appeals, Fifth District and was again denied. In the second appeal the convicted programmer argued that no original documents were removed from the employer's premises, and he did no more than steal $35 worth of paper. He further contended that the computer programs were trade secrets not property covered by the law applied. It is not clear whether he would have been convicted in other states where theft laws are not as all-inclusive.

In California a felony trial[10] is currently in progress where a programmer is accused of taking a program by causing it to be printed at his remote batch terminal over a telephone circuit from a competitor company's computer. This poses a different problem under law compared to the Texas case where copies of the programs were physically removed from the owner's facilities. The case is being tried under a 1967 California theft of trade secrets law where the principal question is whether the program stored in the computer was sufficiently protected to be a trade secret. The defense attorneys are questioning the legality of the new, as yet untested, law.[11] This case also produced the first search warrant ever issued to search the memory of a computer for evidence. The validity of the search was questioned when it was learned that the police sergeant making the search admitted on the witness stand that he was unfamiliar with the items in the warrant. He required a programmer from the complainant company to carry out the search by identifying listings, card decks, documentation, magnetic tapes, and by dumping the computer memory onto tape.

The computer has been used as the environment for criminal actions such as the first case described of changing a program. Computer technology has resulted in the creation of computer programs as new valuable properties subject to theft by traditional means and also by remote terminals over telephone circuits. The computer has also been used as a tool in perpetrating a crime.

In 1968 the head accountant for a company was apprehended after six years of embezzlement of $1,000,880.[12] He was convicted of grand theft and forgery and is now serving a ten-year prison sentence. He was in an excellent position to commit his embezzlement with wide-ranging accounting responsibility. He also operated a data center business which performed all computer data processing for the company. He used the classical method of increasing the recorded amounts paid for raw resources and used the difference of the recorded amount less what was actually paid. The difference was made payable to his own dummy companies which presumably had performed services for his employer. He used a computer in his activities to decide when, how much and from which accounts he should extract money to avoid taking any single amount which might be noticed. He developed a model of the company using his computer and would experiment with his changed data and the true data. He also was successful in assuming that his management readily accepted any reports coming from the line printer of a computer as being a priori true. The significance of this case lies in the fact that he was successful for six years without detection. It was only when he broke his own rules and in the last two years became greedy in extracting money at the rate of $250,000 per year, that officers in the local branch bank where he had the accounts of his dummy companies became suspicious and reported his activities. This resulted in his arrest.

Computers have become well known targets of vandalism. Two major cases occurred in 1969, four in 1970 but no major ones have been verified in 1971. One threat of vandalism occurred in 1969 and one in 1970. Major cases have involved the destruction of hardware and data. Seven have been publicized in which data and programs have been destroyed or made useless by changing tape labels or changing programs to erase files. All major acts were for national political purposes. Most of the others were done by disgruntled employees taking revenge on their employers.

Few publicized cases have involved data communications, but data communications have caused and will continue to cause considerable legal problems when involved in criminal cases. The previously cited case of alleged program theft is one example. Another vandalism case[13] involved the central computer of a large company which automatically polled tape stations in branch offices. Disgruntled employees of a service company masqueraded as the computer sending instructions to the tape stations and made them inoperable. The perpetrators were threatened with criminal charges of aggravated harassment which is used primarily to prosecute cases of obscene telephone calls. One Federal case[14] involved interstate telephone calls to gain unauthorized access to a time-sharing service, but the U. S. Attorney dropped the case for lack of evidence. It is interesting to note that in England unauthorized use of a computer and communications facilities can be prosecuted under Section 13 of the Theft Act. This section defines the offense of dishonest use of electricity, the property of another. A penalty of up to five years imprisonment is possible.

At least 82 cases since 1964 have been nationally publicized, but a number of them have not been verified beyond newspaper or trade journal articles. A number of these are known or suspected to be untrue or not concerned with computers. All credit card frauds and other cases involving negotiable instruments have been omitted when such instruments were not in machine readable form, or the acts did not rely principally on machine reading. Table 2 shows the incidence of computer-related crimes.

Table 2
Incidence of Published Computer-Related Crimes

Year	Vandalism	Info. & Property Crime	Financial Crimes
1964		1	2
1965			
1966			1
1967			
1968	1	2	3
1969	3	1	1
1970	6	7	6
1971	5	20	10
1972 (1st qtr.)	1	3	3
Totals	16	34	26

Year	Theft of Computer & Communication Svcs.	Numbers Not Completely Verified	Total Crimes
1964		2	3
1965			0
1966			1
1967			0
1968		3	6
1969	1	3	6
1970	1	12	20
1971	2	27	37
1972 (1st qtr.)	2	5	9
Totals	6	52	82

The reporting of unethical, criminal and civil cases involving the use of computers has frequently been inaccurate and sensationalized. Frequently reputable authors of articles in public media, trade magazines and professional journals have reported cases based only on newspaper reports or have just repeated stories told in other articles. For the most part, these stories are told in articles on computer security wherein several brief anecdotes are told to illustrate the problems of computer security. If references are given for the stories at all, they quote news items or other security articles. Actual legal documents or personal investigations are hardly ever cited. One apocryphal story concerns a group of Boy Scouts who destroyed magnetically recorded data with magnets while on a tour of a computer center.[15] Another is the often repeated story of changing a program processing mutual fund accounts to collect the leftover tenths of cents in round-downs of arithmetic operations from many accounts and depositing the surplus in a favored account. There are also versions of this story concerning payroll programs. The mutual fund story was traced to a national newspaper article written by a staff member no longer employed there. The article quoted a computer security consultant who described the method only as an example of what he could do without detection.[16] The story has never been verified from first-hand knowledge and is thought to be a old accounting story predating computers.

It is stated by a number of computer security consultants, law enforcement officers and CPA firm members that the publicized cases of computer-related crimes constitute only the visible top of an iceberg. Many incidents involving dishonest and unethical acts are not prosecuted. They are settled by informal means because of fear of harmful publicity, lack of confidence that laws would apply and inability to find lawyers, law enforcement agencies and courts sufficiently aware of computer and communications technology to understand what happened. Another possible deterrent is the lack of detection capability to prove that an act actually happened.

The Future of Computer-Related Crime

The pervasive penetration of computers and data communications into the functioning of society is justification enough to predict a growing level of crime related to information processing systems. Computer-related crime will occur just as crime has always occurred in social activities which provide great leverage for human benefits. This much is evident, but how will crime manifest itself in an automated society? Will the incidence go up or down? Will the cost-to-protect/value-of-assets-protected ratio increase or decrease? Will losses by crime increase or decrease? Will the advent of automation cause a sufficient change in criminal methods to produce an actual change in types of crimes?

Consider the hypothesis that the incidence of all financial and informational crime will decrease, but the losses per incident will increase as direct results of increasing use of computers and data communications. Unfortunately the arguments for and against this hypothesis are themselves hypotheses but possibly closer to being self-evident or easier to support as more experience is gained and made subject to deeper study.

In opposition to this hypothesis the FBI Uniform Crime Reports show a steady increase in the incidence of crime and associated losses. One sample: 4,354 property crimes per 100,000 population were reported in 1970 in cities over 250,000 in population. The President's Commission on Law Enforcement and Administration of Justice Report of 1967 estimates the following annual losses:

Embezzlement	$ 200 million
Fraud	$1,350 million
Tax Fraud	$ 100 million
Forgery	$ 80 million

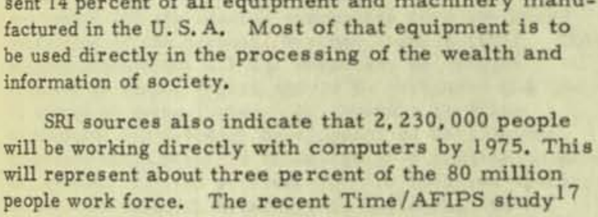
Computer technology has been increasing its impact on society for about twenty years. However, less than 100 computer and data communications crimes have been reported since 1964. This is not enough to matter relative to society and its many problems. The incidence of crime is determined by social values, concentrations of populations, increasing complexity of life, restrictions of actions, increasing freedoms of choice and advancing technology. Information processing and communications technology is just one small part of technology.

Even if information processing and communications technology were to have a major impact in areas of societal activities where financial and informational crime has its major impact, such technology does not appreciably change the number of potentially dishonest people and the opportunities for them to act. In fact, the concentration of data as a result of computer processing and mechanistic, inflexible rules of processing create new opportunities for dishonest people. Increasing security of computer and communications systems will prove to be no different in effect than security associated with manual systems of the past. Finally, losses per incident will decrease as the use of minicomputers in small businesses grows.

In support of the hypothesis that incidence of all financial crime will decrease and loss per incident will increase, first consider the impact of computers and data communications on business and government. Stanford Research Institute sources estimate that there are 80,000 computers currently in use in the U.S.A. By 1975 this should reach 140,000 and by 1980 the total will be 170,000. In 1980 annual sales are expected to reach 18 billion dollars and represent 14 percent of all equipment and machinery manufactured in the U.S.A. Most of that equipment is to be used directly in the processing of the wealth and information of society.

SRI sources also indicate that 2,230,000 people will be working directly with computers by 1975. This will represent about three percent of the 80 million people work force. The recent Time/AFIPS study[17] indicates that seven percent of the work force in 1971 claimed they worked directly with computers. Assume that a large proportion of the population with opportunity to perpetrate financial and informational crimes are those associated directly with handling wealth and sensitive information as shown on the diagram. The overlap with the three to seven percent directly working with computers provides a strong but not conclusive argument that financial and informational crimes will tend to be associated with computer processing and data communications.

Diagram. Work Force Sets



Financial information processing workers who are criminals.

The next step of the argument is to claim that such technology will result in reduced incidence and increased loss per incident. There are three relevant points to make in this regard. Computer processing and data communications result in greater concentration of data and produce more automation and uniformity of handling. This increases the opportunity for protecting sensitive data compared to previous manual systems. Secondly, the number of people required to process and communicate data is significantly less than in previous manual systems comparing equal volumes of data and amounts of processing. Finally, the expertise required of the smaller number of key people now engaged in the more sensitive aspects of computer processing and data communications is much greater technically and more narrow in subject matter. The increased complexity of technology and sophistication of security will likewise require greater expertise and knowledge to penetrate systems thus reducing the population of potential penetrators. However, when this expertise is used against automated systems which are processing great concentrations of wealth and vital information at high speed with little human intervention, the losses per act from successful but fewer dishonest acts should be expected to rise.

References

1. Alan Westin
Privacy and Freedom. Atheneum. New York. 1967.

2. Edwin H. Sutherland
White-Collar Crime, The Offender in Business and the Professions. Gilbert Geis, editor. Atherton Press. New York. 1968.

3. Quon Y. Kwan, etc.
The Role of Criminalistics in White-Collar Crimes. J. of Criminal Law, Criminology and Police Science. Vol. 62, No. 3, 1971 pp. 437-449.

4. Ray Zablocki
Credit Card Swindles Spur Development of Detection Systems. Stanford Research Institute, Investments in Tomorrow. Winter. 1971.

5. H. E. Petersen and R. Turn
System Implications of Information Privacy. AFIPS Conference Proc. SJCC 1967. pp. 291-300.

6. Clark Weissman
Trade-off Considerations in Security System Design. J. of Data Management. April, 1972.

7. W. D. Tiffany
Are Your Tapes Really Vulnerable to Magnets? Unpublished Stanford Research Report. March 29, 1972.

8. Donn B. Parker
Documentation of the Indictment of (________)* by the United States Attorney, Minnesota. A Report to the Council of the Association for Computing Machinery. January, 1967. U. S. District Court of Minnesota, Fourth Division. Information (Title 18 U. S. C., 1005). Charges filed by Stanley H. Green, Assistant U. S. Attorney.

*Name will be supplied by Donn B. Parker, upon request.

9. (________)*, Appellant V. The State of Texas, Appellee. No. 39526. Court of Criminal Appeals of Texas. April 20, 1966. Citation: 402 Southwestern Reporter 2d906. (________)* v. State. *Name will be supplied by Donn B. Parker upon request.

10. Complaint. The People of the State of California vs. (________)* Municipal Court, Oakland-Piedmont Judicial District, Alameda County, Calif. February 23, 1971.

11. Memorandum of Decision No. 51629. (________)*, Petitioner vs. Superior Court of California, Alameda County. John P. Sparrow, Judge. March 22, 1972.
*Name will be supplied by Donn B. Parker upon request.

12. Donn B. Parker
Documentation of the Indictment and Conviction of________* in Superior Court________*, Calif. Report presented to the Association for Computing Machinery Council, January, 1969.
*Names will be supplied by Donn B. Parker on request.

13. Private communication. See also Computerworld. December 15, 1971.

14. Byron E. Trapp, Assistant U. S. Attorney. U. S. Department of Justice, Southern District of Ohio, April 27, 1971.

15. Sabotage, Accidents and Fraud Cause Woes for Computer Centers. The Wall Street Journal. March 22, 1971.

16. Sheldon Dansiger.
Embezzling Primer. Computers and Automation, November, 1967.

17. A National Survey of the Public's Attitudes Toward Computers. American Federation of Information Processing Societies and Time Magazine. July, 1971.

AFOS: A PROGRAM FOR NATIONAL WEATHER SERVICE FIELD AUTOMATION

Gerald A. Petersen, National Oceanic and Atmospheric Administration
Silver Spring, Maryland, U.S.A.

Abstract

The Systems Development Office of the National Weather Service has been conducting a number of interrelated studies to determine the feasibility of introducing automation into a diverse and wide-spread field structure. The general conclusions of these studies indicate that many aspects of the National Weather Service's field operations can be cost/effectively automated through a suitable combination of minicomputer and communication technologies. Current efforts have concentrated on integrating the results into a unified Automation of Field Operations and Services (AFOS) concept.

This paper will present a descriptive discussion of the AFOS concept mainly from an internal operation viewpoint but with some suggestion as to how it affects the external users of the weather service.

Introduction

The "Automation of Field Operations and Services (AFOS)" Program has been established to provide a focus and method for introducing automation into the field operations of the National Weather Service (NWS). This program encompasses all the necessary activities required to identify what should be automated and how it should be accomplished. It includes hardware acquisition and software development as well as the experimentation required to design an automated operating system. It also provides for the detailed specifications and planning necessary to implement the automated system into the National Weather Service's field operations.

Present methods used in the field operations and services rely mainly on manual techniques for observation and collection of meteorological data and for forecast preparation and dissemination. In fact, only through the support and dedication of the man in the system has it been possible for the NWS to provide and maintain its present services. With the increasing demands on the NWS for additional service concurrent with limits on the number of personnel in the system, it has become increasingly difficult to meet existing weather service requirements, much less any new requirements. It is apparent that other ways must be found to meet these service needs rather than continuing to try to do so through the addition of people to the operating system. Automation offers a viable option for overcoming this problem.

The purpose of this paper is to provide a description of the AFOS program. Following a brief summary of the background leading to AFOS, the concept will be described functionally as well as operationally (i.e., how these functions are to be integrated into an automated field operation). Then, a preliminary evaluation of AFOS impacts will be given.

Background

The National Weather Service's Systems Development Office has been conducting several studies to determine the feasibility of introducing automation into the field structure. These studies mainly considered individual pieces of the operating system but they provided guidance and insight into how the total field system should be automated.

The major study results leading to the AFOS concept to be discussed in the next section are as follows:

Weather Service Forecast Office Data Handling Study--Recommends a distributed store-and-forward data handling and alpha-numeric/graphic display system consisting of on-site minicomputers and random access mass storage disks at each of the fifty Weather Service Forecast Offices.

Data Acquisition Automation Study--Recommends that preprocessing be accomplished using minicomputers at all sites with weather radar and/or upper air observation functions.

Communications--Trade-off analysis indicates that automated assistance for alpha-numeric message composition alone could be based upon a distributed on-site minicomputer system at approximately the same cost as a centralized system but with improved reliability.

Next Generation Upper Air (NEXAIR) Development--Techniques to be used for deriving upper air data in the NEXAIR system suggest that on-site minicomputer preprocessing is the only economically feasible approach.

Digitization of Radar Data Experiment--Again, the magnitude of the data preprocessing required makes it economically attractive for on-site minicomputer digitizing of radar weather information.

Each of the above efforts indicates that a combination of current minicomputer and communications technology will provide a cost-effective means for improving the NWS operational field environment from data acquisition to eventual dissemination to the end user. The trick, of course, is putting the pieces together into a coherent, total field automation program.

The AFOS Concept

Before describing the AFOS concept in functional and operational terms, the significant underlying assumptions will be given.

Assumptions--First, it has been assumed that maximum automation of field operations and services, within technical and economic realities, represents the policy and goal of both the National Weather Service and the National Oceanic and Atmospheric Administration.

Secondly, the field system will continue to provide existing as well as new services through the use of automated assistance in a "man-machine" mix type of operation. However, modification of these services may well occur if the automated system leads to better overall effectiveness from the viewpoint of the end user.

Also, it has been assumed that the present forecasting structure (Figure 1) will remain relatively unchanged over the next five to ten years. That is, the National Meteorological Center will continue to perform the central scientific processing functions in its preparation of national meteorological guidance for field operations. The National Hurricane and National Severe Storm Forecasting Centers similarly will continue to operate much as they do today. The AFOS concept, however, will materially enhance their ability to respond to severe weather situations.

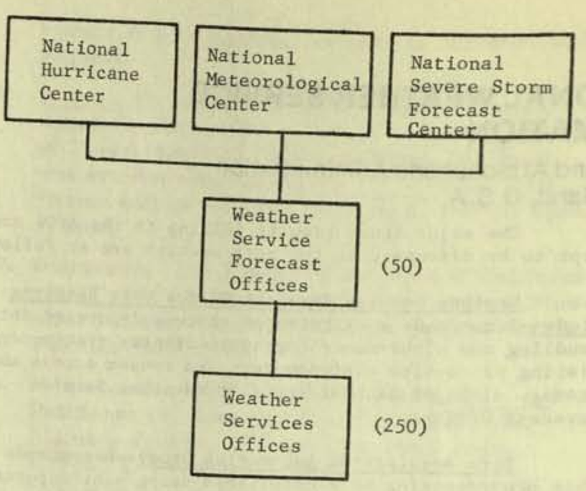


Figure 1. National Weather Service Field Forecast Structure

The Weather Service Forecast Offices (WSFO) will operate essentially as they do now with the prime responsibility for user products over an area approximately the size of an average state. Examples of products include state and local public forecasts, aviation terminal forecasts, and severe weather warnings.

The 250 Weather Service Offices (WSO) will continue to serve primarily as observation/dissemination sites but with short term local forecast and warning responsibility under the jurisdiction of the appropriate forecast office.

The aim of automation is to improve the ability of each level in the field forecasting system in providing user services without materially affecting the organizational responsibilities assigned. It should be recognized, however, that the introduction of automation on a large scale will bring about changes in many of the mechanical aspects of field operations and could ultimately result in significant organizational changes. Thus, automation will provide an environment that allows existing manpower levels to handle expansions in day-to-day operations in a more responsive and effective manner than could be achieved through continued increases in personnel.

Functional Description

As a result of the studies completed to date, the following major field functions have been identified as part of the AFOS concept.

Data Acquisition--Acquisition of raw sensor data will be automated to the fullest possible extent. In this area, automation of surface, upper air, and radar observations have been included as components of the AFOS system.

Preprocessing--Closely related to data acquisition is the preprocessing function which refers to the conversion of raw sensor data into meteorological parameters in a form suitable for effective transmission.

Processing--The role of translating meteorological and hydrological data into analyses and forecasts via scientific computer models will continue to be carried out by the National Centers and by the River Forecast Centers. It is possible that a limited amount of on-station processing involving localized forecast techniques (e.g., max/min temperature forecasts, statistical forecasts) could be done at the WSFO's by setting aside fixed portions of CPU core and disk storage. In addition, it is expected that many of the on-station record summarizations could be processed by the minicomputer during idle time periods.

Storage, Retrieval and Message Compositions--Minicomputers and random access mass storage disks at WSFO's will provide on-site storage of the entire data base (graphics and alpha-numerics) required for retrieval, display, and manipulation. The major effect of automation of this function will be to put the data where it can be easily and quickly retrieved for use in product preparation as opposed to the current system of distributing data via facsimile and teletypewriter circuits in paper form. Such things as precanned formats and previous products will be readily available from the computer subsystem to ease the message composition task and make it more responsive to the user needs.

Displays--Because of the need to display gray scale radar and satellite images, the technological state of the art options include only digital TV and storage tube scan converters. For alpha-numerics, the random scan CRT could be used. The analysis to date, however, supports the storage tube scan converter display subsystem as being most attractive based on its cost and performance characteristics for both alpha-numeric and graphical data.

Dissemination--The final distribution of products to end users characterizes the dissemination function. The modular nature of the AFOS concept provides for a number of possibilities that will enhance timeliness and quality as well as offering the potential for achieving a well-balanced mix of dissemination techniques. For example, automated transmission and switching over low and high speed circuits of an alpha-numeric warning that has been composed on a CRT using a computer stored precanned format, translates into quicker and more effective response to a severe weather situation. Also, delivery of graphical products from a minicomputer/disk subsystem to nearby subscribers poses no significant technical problems. Therefore, the major effort in assessing the AFOS impact upon dissemination is not a function of technology; rather, it involves determining the "best" mix of man/machine techniques.

Communications--A national digital communications circuit operating in conjunction with the Weather Service Forecast Office minicomputer will provide the means for a store-and-forward network configuration that includes automated switching on off-line low and higher speed localized circuits. The basic digital circuit will probably consist of a full-duplex 2400 bit per second voice grade line connecting all 50 forecast offices in a looped chain-like fashion with off-site preprocessed data entering the circuit through the nearest WSFO over the local area networks. Product distribution to offices and users beyond the WSFO level will be accomplished via the local networks which may range from higher speed (e.g., 1200 or 2400 bits per second) interactive to low speed (110 bits per second) teletypewriter receive only lines.

Operational Description

A "typical" example will be used in describing the operational characteristics of the AFOS system. As the inherent modularity of the concept represents one of its attractive features, exceptions in types and functions of offices will present no technological limitation for accommodation.

The heart of the AFOS system will be the Weather Service Forecast Offices. Each WSFO will be a node, and will have a minicomputer and random access mass storage disk (Figure 2) subsystem with two major functions. First, automated assistance will be provided to the forecasters responsible for the production of user products. This assistance will be in the form of TV-type displays, keyboards, and a hard copy device which constitutes the display subsystem. The analysis to date indicates that a storage tube scan converter utilizing high resolution TV displays adequately satisfies response time requirements while allowing for alpha-numeric, graphic, and gray scale image data presentation. The on-site storage capacity required amounts to somewhat less than 40 million bits for a full day's worth of current data. However, it is expected that selective storage for various types of data will range from hourly updates to as much as three or four days for specified weather maps of historical interest.

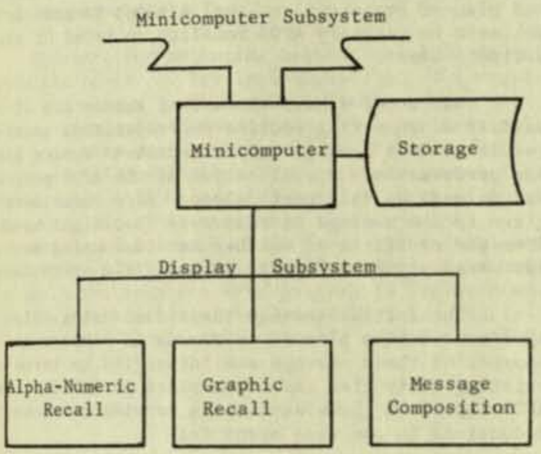


Figure 2. Automated Weather Service Forecast Office Configuration

The second major function of the minicomputer subsystem relates to the store-and-forward operation. The lifeline of the AFOS system will most likely consist of a looped 2400 bit per second voice grade telephone line through which graphical and alphanumeric weather data passes to and from all nodes in digital form (Figure 3). As each message comes into the computer, it will be parity checked and, if acceptable, an acknowledgment will be returned to the originating WSFO on a reverse 150 bit per second supervisory channel.

The minicomputer then checks its library to determine if the piece of data is required at the receiving station. If so, the data is stored on the disk. If not, it is forwarded to the next WSFO for receipt and acknowledgment. The operation continues until the data meets itself on the loop. It is important to note that, with such a looped arrangement using an ordinary voice grade line, a fail softly communication back-up becomes possible. Through software and, say, an auto-dial feature at each mini, if an acknowledgment is not received after two or three transmissions, another line can be dialed up by-passing the next node in the chain. In this way, the integrity of the line can be maintained at relatively inexpensive cost.

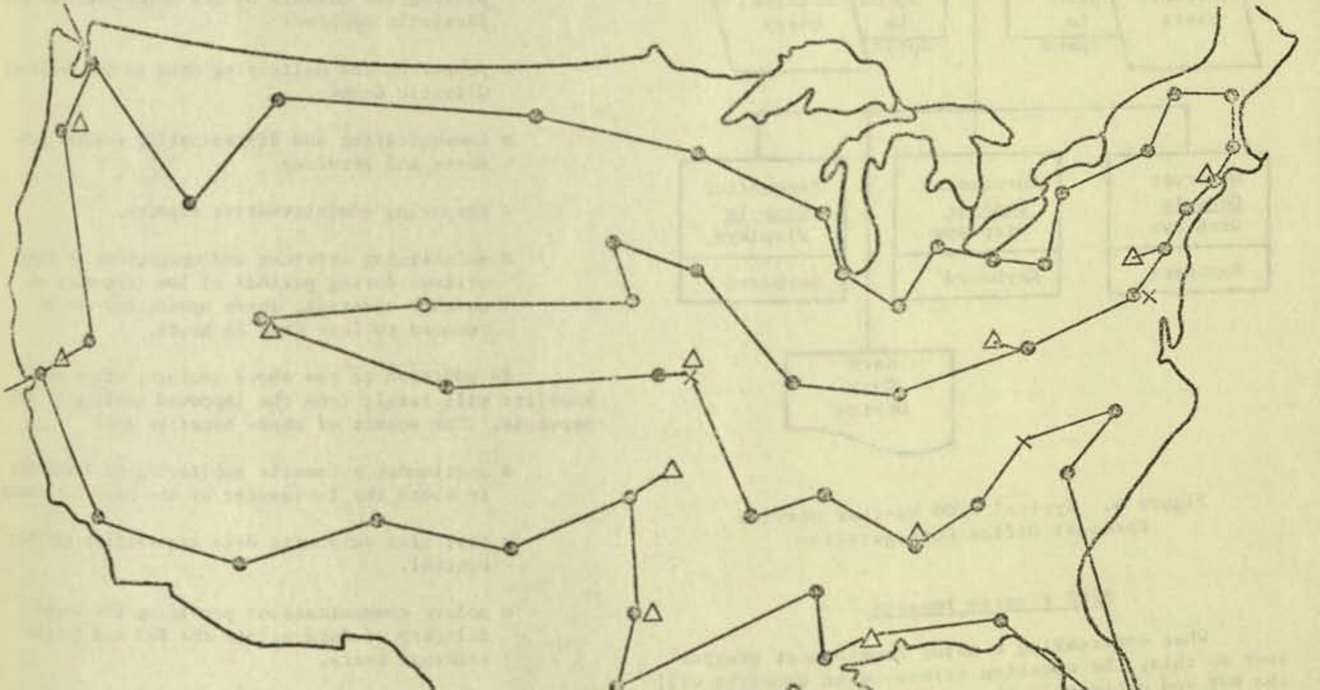


Figure 3. National Digital Circuit

Another related activity to be carried out by the minicomputer is that of data collection and dissemination for the WSFO region of forecast responsibility that usually includes an area of an average sized state. We visualize this involving both slow speed teletypewriter and higher speed graphic or printer displays for output to both internal NWS and external users. Conceptually, a forecast prepared for a portion of a state by the WSFO would be automatically distributed on either a leased line basis, if the communications load happens to be large enough, or on a dial-out line to a particular user (e.g., radio or TV station, newspaper, etc.).

The same principle applies with respect to data collection. If an upper air observation from a meteorological observatory that happens to have no other responsibility is to be collected twice a day, a dial-in line would be used. On the other hand, if the upper air site happened to be colocated with a weather radar and a Weather Service Office, a leased interactive higher speed line probably would be required. Putting these pieces together results in the typical WSFO configuration shown in Figure 4.

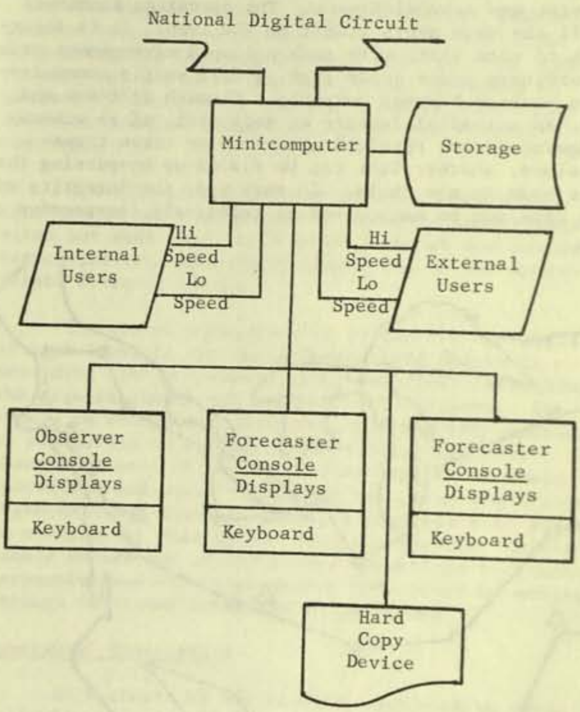


Figure 4. Typical AFOS Weather Service Forecast Office Configuration

AFOS Program Impacts

When undertaking a major development program such as this, the question arises--What benefits will the NWS and ultimately the public derive from the program? This question will be the subject of both a preliminary and continuing effort in the system analysis phase of the AFOS program. At this early stage, it is not possible to quantify benefit with a high degree of confidence. However, we do know enough about potential benefits to identify them; the difficulty, as in all attempts at determining benefits, is quantifying them.

The primary benefactor of the AFOS program will be the nation's population. The outputs of the program should be most apparent in terms of improvements in the quality, timeliness, and reliability of warnings, and in the accuracy of short range forecasts covering the zero to eighteen hour period. In addition, AFOS will make possible the provision of new services without significant increases in NWS staffing. Benefits to the nation from these improvements can be qualitatively estimated in terms of increased safety, convenience, and in the economic well being of the population. The quantification of these benefits in most cases must be approached subjectively. Since AFOS can be justified on the basis of more effective utilization of existing and planned resources, no real attempt is made in this document to quantify AFOS benefits in terms of the ultimate users.

Our studies have shown that automation of activities involving routine and repetitive tasks will result in cost savings and/or increased system outputs and performance. Justification of the AFOS program can be made on this basis alone. Here, consideration is given to the savings in resources that might be obtained from the provision of weather services using an automated versus a largely manual field operation.

The initial savings identified using this approach involve planned increases in personnel. The sources of these savings are identified in terms of existing activities that take place in field operations. AFOS impact on these activities provide for potential reductions in the time spent in:

- o preparing and verifying warnings and forecasts.
- o routine and repetitive tasks of monitoring and posting the outputs of the teletypewriter and facsimile machine.
- o preparing and delivering data to the National Climatic Center.
- o communicating and disseminating weather products and services.
- o preparing administrative reports.
- o maintaining services and operations at some offices during periods of low frequency of weather warnings, where operations can be reduced to less than 24 hours.

In addition to the above savings, other AFOS benefits will result from the improved quality of NWS services. The source of these benefits are:

- o continuous automatic monitoring of forecasts to alert the forecaster of the need for change.
- o real time automatic data acquisition quality control.
- o modern communications providing for rapid delivery of data within the NWS and to the external users.
- o provision of computer produced user oriented services (e.g., special graphics for cable TV).

In evaluating these sources of AFOS benefits, two types of impacts are noted. First, there is the impact on existing operations that allows resources to be more effectively utilized. The second involves a cost avoidance that allows for present five year plans to be implemented at a significantly lower cost, both in terms of manpower and dollars. The first type of impact provides the basis for the second impact--cost avoidance.

The five year plans used in this cost avoidance justification have generally been based on an expansion of existing manual operations. These plans were analyzed to determine the impact of AFOS on the increased number of personnel required to provide new and improved services covered in the plans. The first source of reductions in manpower expansion could be the elimination of planned manual activities which will be automated. The second source is related to more effective utilization of existing personnel in the field operation which allows service expansion without increasing manpower.

However, the potential benefits of AFOS do not provide the means for its implementation. The magnitude of the change in the current field operations resulting from automation dictates that implementation should be spread over a number of years. Also, the initial cost of development, equipment procurement and implementation will be significant. Because of these two factors, it is not realistic to assume that reprogramming of resources will cover the initial cost of automation. Many of the benefits will not materialize until the complete AFOS program is implemented. During the early years of implementation the initial development and investment cost will exceed the AFOS savings. Also, many of the AFOS savings are such that they increase the service capability of the NWS, but do not decrease the current costs of operations; e.g., freeing the time of the forecasters from routine duties does not eliminate the need for the forecasters to prepare products.

AFOS, in the short term, requires increased development and capital investment by the NWS, but its long term effect in providing for increased safety, conveniences and economic benefits to the nation's population makes it an attractive investment. These benefits are derived from more accurate, timely, and reliable warnings and short range forecasts, along with an expanded NWS capability for meeting new service requirements via the AFOS program.

Prior to any widespread field implementation, a "Model Facility" will be implemented for testing, evaluation, and demonstration purposes. Of particular interest will be the reaction of the man to a new system that effectively does away with the slow, cumbersome, and manually oriented paper world of facsimile charts and teletypewriter data. The current working environment has a long history and converting to an automated, soft copy screen operation presents a formidable human factors challenge that may exceed any of the technological problems.

INTEGRATED MUNICIPAL INFORMATION SYSTEMS: SOME POTENTIAL IMPACTS

James W. Evans, Robert A. Knisely
U.S. Department of Housing and Urban Development
Washington, D.C., U.S.A.

Summary

Many approaches to solving urban problems have been tried and suggested over the past two decades. One does not have to breathe the urban air deeply, or scrutinize the urban environment closely, to realize the lack of success of these various efforts. This paper describes a technological approach which may help focus and perhaps identify possible solutions to our urban crises. This approach utilizes computer and telecommunications technology. In this paper, and elsewhere, this approach is called an integrated municipal information system (IMIS). By itself it is not a solution. But, if its theory is valid, useful information will be aggregated from the voluminous bits of data that all cities generate daily. Presently, most of these data are lost, often in file cabinets, but also data are frequently not even collected. Computer technology now gives cities the ability to aggregate, synthesize, and store these data, so that seemingly disparate pieces can be related in useful ways.

Briefly discussed are a theoretical IMIS, and a complex federal program, the Urban Information Systems Inter-Agency Committee (USAC), which is researching the fundamental assumptions, and, where feasible, implementing them. Also discussed are two of the many potential areas of impact resulting from an operating IMIS. One impact is the ability the criminal justice system will have to predict and delineate potential aberrant individuals. It is pointed out that our criminal justice system (corrections and rehabilitation especially) is predicated upon the notion of the individual's free will. This raises several questions. If potential aberrant behavior can be predicted, what happens to the concept of free will? Moreover, what does this portend for the criminal justice system? If society can predict who its aberrants will be, what should it then do? An IMIS, it is believed, will give society this ability. A second impact of an IMIS is discussed which is called program coordination, or more recently in the federal government, services integration. This impact is viewed as a benefit, and occurs as a result of the horizontal and vertical integration of information, otherwise known as information sharing.

Integrated Municipal Information Systems (IMIS)

The decade of the 1960's was one of broad-based research and experimentation in the urban information systems field. As a result, several different approaches to information systems development emerged, generally called housekeeping, databank, and model building. Each approach focused on different problems of urban governance and administration, concerned different groups, and represented different conceptions of an urban information system.[1]

During the middle 1960's, it became clear that research was needed on how to build a broader information system capability for cities.[2] During these years a number of such efforts developed around the country most of which utilized a systems approach to the municipal use of computers. A systems approach to urban government is an attempt to view broadly, and within a unified framework, the processes of urban government, and the use of information for decision making within those processes. It also views the utility of information technology in its potential contribution to improving the operational information and decision processes in urban government. To achieve that potential, information is viewed first as a part of the governmental operating processes and secondarily as data to be handled. Concurrently with the use of the systems approach, the notion of an Integrated Municipal Information System developed.[3]

An IMIS is a comprehensive concept of applying computer technology to urban government which, when implemented, becomes an integral part of the operational, managerial, and administrative functions of the city government. Its more significant objectives are the following:

- An IMIS is to operate in direct support of operational activities and, in some instances, actually perform these operations.
- An IMIS seeks to increase staff productivity.
- The system includes information necessary for analyzing changing community conditions for the purposes of problem analysis and program development.
- An IMIS should facilitate the management of all operations.
- An IMIS seeks to protect the privacy of each citizen while providing an information base which will improve service delivery to the people generally and, where appropriate, to particular individuals.

Because the term automation is germane to IMIS theory, it is helpful to discuss it. Automation, rather than simply computerizing limited manual operations, is the process of identifying all aspects of activities which are functionally related and which have common data needs.[4] These operations or groups of activities are then analyzed in terms of the manner in which they utilize data to find their related objectives. In each instance the operations are performed by the computer. Automation of such processes as recording and computational activities, filing tasks, and document preparation, enables the municipal employee to address himself to those problems which require judgment and discretion. The computer, while relieving the employee from much of the daily routine characteristic of many municipal activities, thus becomes imbedded in, and an organic part of, the administrative and operational processes of city government.

The result should be an improved service delivery system for the citizenry.

Automation is often confused with "computerization." The latter refers to the simple mechanization of a given operation in which the computer usually acts as a printer and performs arithmetical exercises for one operation. Thus, the process of keypunching the contents of employee time sheets, and feeding the cards to the computer so that extensions can be made, withholdings computed and deducted, and checks printed is a process of "computerization." In cities today, this is the most common use of computers. Automation seeks to utilize computational techniques to link together a series of events which usually have required human intervention for a relatively low order of reasoning and decision-making.[5] Referring to the previous example, the preparation of payroll checks usually involves human review of manually prepared time sheets. The summarized results of the run are then recorded in books of account for budgetary and cost accounting purposes. With today's technology, it is possible to permit the computational equipment to take the rawest data concerning the number of hours a given employee worked and proceed to verify eligibility, look up all applicable rates, make all extensions, prepare the checks, record in the accounting records the funds affected, and prepare a work distribution analysis record for analytical purposes. In this process there is no need for human intervention because none of the steps which connect the sequences in the operation require human discretion.[6] The applicable rules and conditions are exactly prescribed, and the machine can apply them even better, in terms of time and accuracy, than a person can. This is the process of automation.

Because each word in the name "Integrated Municipal Information System" contributes to the meaning assigned to an IMIS, it is useful to explore the ideas included in those terms.

Integrated

This term indicates that an information system is treated as a whole. When integrated refers to the data elements, it implies the maximum utilization of each item of data with respect to every possible application. The objective is to eliminate redundancy and increase the overall efficiency of the system. When the term is applied to the components of the system and each of its subsystems, it refers to the relevance each part has to all others. For example, one record of a building permit might be utilized for code enforcement, land use analysis, fire records, and educational planning.

Municipal

This term refers to that unit of government which is closest to the source of data and which has the need for, and authority to collect, this data. Such a unit is usually a general-purpose government which is designated as a municipal corporation. Cities, villages, and boroughs clearly fall within this category, as do such jurisdictions as counties and towns. For purposes of the USAC Program, the term is restricted to cities.

Information

Often information is used as a synonym for data. In an IMIS, the difference between these two terms is critical. Data refers to the numbers, letters, or other symbols which are stored and processed. Each discrete item of data is, by itself, nearly meaningless. Only when data elements are combined and meaning attributed to them do they become information. In an integrated system, each item of data is carefully stored for use with respect to all the applications to which it relates. Thus, the same data, combined in different ways, provides different types of information. For instance, data concerning a given parcel of property has different uses and sometimes different meanings to planners, police officials, or transportation specialists. In short, information in this context is a generic term which includes the data, the meaning attributed to the data, the uses made of the data in performing municipal functions, and that combination of human and machine interaction which data and its meanings sustain.

System

With respect to an IMIS, system refers both to the municipal system for delivering services, and to the data processing system which permits the city to operate. If one regards the municipality as a system, organizational lines and traditional functional units are no longer significant. City activities can thus be seen as related subsystems within the total system.

In this context, system refers to a perspective for describing data and the manner in which it is processed. The systems perspective views data in a set of relationships. The sources from which the data is derived, the manner in which it flows in a time stream, and the manner in which it is used to provide municipal services are sets of relationships which have significance for the information system. Also, the techniques used for information storage and retrieval, the manner in which it is reported and used by people and organizations, and the way it is used for analyzing and assessing community problems and needs is significant for the system.

While an IMIS may be said to exist as an entity, from the USAC perspective it is possible and profitable to view it from two analytical views, horizontal and vertical development. Of these, horizontal development is the most critical since it acts as a decisive limit to payoffs to be achieved in vertical development. In this respect, both horizontal and vertical development decisively limit benefits which would otherwise be available through increasing the level of sophistication. It seems important therefore to have a better understanding of these conceptual perspectives.

Horizontal

The feature which best distinguishes an IMIS from other computer systems is its concern for the interfunctional uses of data and information. In the typical municipal organization, information flows are designed to follow organizational hierarchy. This results in redundant data collection and processing, as well as gaps in the collection of needed data, which inhibit the use of the systems data for operational or analytic use. From a horizontal perspective, an IMIS views data as an inherent part of routine operations flowing independently of organizational lines. As the system processes ordinary municipal operations, a data base is created which is constantly updated. In short, in its horizontal mode, an IMIS performs all operations, creates an integrated data base and makes data available for all purposes on a city-wide basis.[7]

Vertical

From this perspective, information is provided within the organization where and when it is needed. The information pases to predetermined levels to serve the full range of managerial needs. From the standpoint of management, the ultimate purpose of an IMIS is to serve these vertical needs. Managers frequently regard these needs as the "payoff" of the system. However, it is a mistake to regard these vertical applications as the system itself. Because of the immediate advantages for planning, direction, and control in vertical systems development, there is a great temptation to by-pass the development of the operationally-based applications and the dynamic, unified data base called for by the horizontal approach. Indeed, this temptation is the basic reason for the shortcomings in current attempts at building municipal information systems. It is a cardinal principle of an IMIS that vertical applications can function properly only if they are based on a soundly developed horizontal operation.[8]

An IMIS is not created in its entirety at one stroke; rather, it must be developed incrementally. In a sense, the system should evolve progressively through several levels of sophistication to the ultimate realization of the vertical benefits. In this manner, each state of development has its own operational value as the system grows through successive levels of capability.

During the process of evolution and growth, the system will pass from meeting the conventional operational needs of the municipality to fulfillment of the information requirements of all levels of management. In the process, it will provide for all types of planning and control activities. The computer system itself likewise will develop in complexity. In the early stages, it may operate in a batch processing mode whereby the input media are accumulated over a period of time as they come in and are then processed at one time by the computer. As the system develops, the computers will come to operate in a "real time" atmosphere whereby the computer can be queried at any time with respect to information in the data base. In an IMIS, the system is thus the people, the hardware, the programs, the data base, and the procedures working in a prescribed pattern.

The system which is comprised of these elements is designed to collect, store, update, and facilitate the automated use of data on a continuing basis. Flexibility in use is the key element sought as the system is designed to perform the following functions simultaneously: fulfill, on an automated basis, standard operational requirements such as issuing permits and licenses, maintaining accounting records, providing information for police and fire purposes, and recording and updating information on the conditions of streets. In addition, the system will generate continuous flows of information concerning community conditions and governmental operations so that the directives issued, the action taken, and the results of planning and operations are recorded and reported in a coordinated manner. In other words, the system provides management with feedback so it can identify the results of its decisions in terms of the action taken and the impact of the action on the community. In this sense, it is a cybernetic, i.e., directive and self-correcting, system. Finally, the system will permit the exchange of information between local governments, between local governments and state and federal agencies, and between local governments and the public served. If an IMIS were universally installed in all local governments, it would serve as a primary data base for state governments, and as a vitally important data base for the domestic programs of the Federal government.

In an environment of comprehensive automation, man and machine rely completely on one another. If responsibilities have been properly allocated, one cannot override the other. Machines will maintain records, make many routine, nondiscretional decisions, and indicate at appropriate times the need for human response. While initially this may seem farfetched, and even frightening, this simply means that the municipal employee works with the equipment much as he now does with a telephone or an earth mover. In all of these instances, reliance is based on the use of equipment where the equipment can perform better than the human.

With this background on a theoretical integrated municipal information system, the discussion now turns to the USAC Program, which is an attempt to explore IMIS concepts and develop them where possible.

USAC Information Systems Research Program

Overview

The USAC Program is a research and development effort to develop improved management and service delivery capabilities in municipalities through application of information systems technology. The five city projects are sponsored, and funded in part, by the Urban Information Systems Inter-Agency Committee (USAC), consisting of ten Federal agencies.[9]

For the purposes of this effort, city government has been broken into the following four subsystems: physical and economic development, public safety; public finance, and human resources development. Human resources development, for example, is further broken into health, education, welfare, libraries, manpower, vital statistics, voter registration, and recreation. These subsystems will be interrelated, so that within the bounds of well-defined privacy limitations, information from one subsystem will be available to another.

In this program, the Department of Housing and Urban Development is the lead agency; the other member agencies include the Office of Management and Budget; the Office of Economic Opportunity; the Departments of the Army (Office of Civil Defense); Commerce; Health, Education, and Welfare; Justice; Labor; Transportation; and the National Science Foundation.

All (369) American cities in the 50,000-500,000 population range were invited to submit competitive proposals to perform one or more of five types of municipal information systems efforts. Of these seventy-nine cities submitted proposals late in 1969, and six were selected. In March 1970, four cities were given two-year contracts for municipal subsystem development. These cities, and their tasks, are:

- Dayton, Ohio: Public Finance Municipal Information Subsystem
- Long Beach, California: Public Safety Municipal Information Subsystem
- Reading, Pennsylvania: Physical and Economic Development Municipal Information Subsystem

- St. Paul, Minnesota: Human Resources Development Municipal Information Subsystem

In addition, two other cities were given three-year contracts for total municipal information system development. These two cities are Charlotte, North Carolina, and Wichita Falls, Texas.

USAC's Formation

USAC was officially established September 10, 1968 for the express purpose of sponsoring research into municipal information systems. Its formation was the expression of a philosophy of intergovernmental relations that evolved out of Federal and local experiences in the 1960's. Basic to this philosophy was the idea that execution of Federal policies and programs ultimately rests upon the capacity of local governments to activate themselves to pursue improvement in urban conditions successfully. One important way to improve local government capacities appeared to be the development of information systems which would provide the data needed to conceive, plan, and evaluate government actions. Such systems could constitute a significant innovative mechanism for improvement, not only through the information they would make available, but also through the following:[10]

- The processes of analysis, design, and evaluation which inhere in the building of such systems.
- The rationalization and automation of routine information and decision processes.
- The provision of a current and multi-faceted data base to support non-routine decisions and actions.
- The facilitation of inter-governmental information flows.

Thus, USAC felt that information systems might constitute a means of bringing about fundamentally improved conditions and outputs in local governmental systems.[11]

USAC Research and Development Objectives

The objectives adopted for USAC's immediate research and development program were one, pragmatic recognition of urban information systems, two, research needs, and three, the potential benefits for municipalities and other governments to be derived from automation. The broad objective of the program was viewed as determining whether such information systems and subsystems could be built, and, if built, whether the concepts, designs, and experience could be transferred to other municipalities.[12]

Within this context, the USAC Request for Proposal (RFP) outlined the following specific objectives:[13]

- To improve the information and decision-making capabilities of municipalities.
- To provide a broader approach in the research and development of municipal information systems.
- To encourage the standardization of data and inventories of data, both vertically through successive levels of government, and horizontally at each level.
- To document solutions to the problems of data acquisition, data management, and data use.
- To develop solutions to the problems generated by sensitivity of information, viz., the protection of confidentiality.
- To develop solutions to the problems of subsystem and system linking, both vertically and horizontally, viz., the technique of facilitating systems compatibility and/or interfacing.
- To develop solutions to problems of scheduling, dispatching and control automation.
- To learn more of the secondary impacts of the implementation of municipal information, e.g., their implication for administrative organization and behavior.
- To test the payoffs of the broad approach.
- To ensure that the information systems developed are operationally based.
- To develop solutions to the problem of transferability.
- To generalize all of the above solutions so as to permit, with only minor modification, their application to other municipalities throughout the nation.
- To distribute reports of proven technical developments equally with reports of techniques which gave difficulty or which failed. This is intended to obviate the necessity of repeated research and development.

Broadly conceived, then, the USAC Program represents a major force for change in municipal government administration through the introduction of a new technology. Further, within the narrower framework of concepts for computer usage, USAC's Program represents a challenge to the concepts that traditionally have guided municipal Electronic Data Processing (EDP) development, i.e., USAC's emphasis on comprehensive and integrated systems that are operations-based directly challenges prior research and development emphasis on data banks, statistical control, model building, and management information systems.[14]

USAC Request for Proposal

Following its formation, USAC developed a research plan and designed an administrative vehicle for implementation of its objectives. The plan called for use of the contract mechanism and request for proposal (RFP) as the means of initiating the administrative vehicle. The RFP was the means of translating the plan into an action program. Therefore, a brief summary of the original RFP seems appropriate as a way of elucidating the nature of the USAC Program.

The RFP called for research and development of integrated information systems and subsystems within municipal governments. These systems were to be operations-based and were to be developed with a view towards their transferability to other municipalities with a minimum of modification being required. The research and development work was to be conducted by a consortium comprised of the municipality as prime contractor, a systems/software organization, and a university or other organized research center.

We have briefly discussed a theoretical IMIS, and the USAC Program. With this introduction, we will turn to two possible impact areas resulting from the implementation of an IMIS.

Computers and the Criminal Justice System

Surprisingly little research has been conducted on the possible secondary impacts of computers on the criminal justice system.[15] The problem is serious, however, because even the foreseeable impacts will greatly affect the current operations of the police, the courts, and corrections. Moreover, these impacts may call into question the very nature of the criminal justice system itself.

It should initially be noted the four ways computers handle data differently from manual systems. First, with the aid of terminal devices, they can rapidly move data from one location to another, limited only by the speed of light and the display or printing speed of the terminal devices. This contrasts sharply with the drafting, typing, and mailing of memoranda and reports. Second, computers can locate data almost instantaneously, once stored on disk, drum, or tape. Third, they have to be told to forget data, once accumulated. Fourth, computers allow "browsing" through a vast array of data for correlations which would be unimaginable to the human researcher, who is always limited by the learned logic of inductive and deductive reasoning, and cause/effect relationships. Computers can detect relationships, not because they should exist, but because they do.

The claim that computers really do nothing new can therefore be refuted. The four characteristics outlined above, especially the ability to sort rapidly through vast amounts of information, produce a qualitative change in information-processing. If all information about almost everything in a given area is available nearly instantaneously at any location, it is begging the issue to deny a qualitative change by noting that the same information is, after all, "in the files." It is thus possible that the qualitative leap in information handling resulting from the invention of computers equals the information leap resulting from Gutenberg's invention of movable type.

Our legal system, having its roots in Anglo-Saxon jurisprudence, is based upon the notion of individual responsibility for behavior.[16] The concept of responsibility necessitates the presumption of free will, otherwise guilt or innocence cannot be determined. Indeed, if free will cannot be presumed, the system has difficulty responding. For example, if a judge knew, beyond a reasonable doubt, that an individual would commit further criminal acts whether incarcerated or not, what should be the judge's response? Furthermore, if statistical analyses of anti-social behavior could predict with compelling certainty that a given adult would continue with anti-social behavior, or that an infant in a given environment was fated to a life of crime, can society and the criminal justice system respond in a manner which presumes free will? The use of the computer in an integrated municipal information system may demonstrate exactly this predictability--the involuntary character of the anti-social acts of certain individuals. This will add to the confusion in an already murky area of the law, that of "diminished responsibility."

An example of such predictability is found in Ohio. In 1963, a study relating Cincinnati street crime and social services was conducted for the Community Health and Welfare Council. The study identified, by name, about one thousand families in Cincinnati who were responsible for 60% of the street crime in that city. These families also absorbed (coincidentally) 60% of the social service benefits dispensed in the city. Let us assume high validity for this study. Multiple-problem clients and criminal recidivism are familiar topics in the literature of social service and criminal justice. For the most part, the Cincinnati study was a laboriously manual operation utilizing manual files. An integrated municipal information system would allow even greater specificity about the social and anti-social behavior of individual citizens at, of course, fantastic speeds. All of the information would be taken directly from public records; the computer would merely correlate it, and if asked, perform probabilistic calculations about future activity.

The purposes of incarceration are deterrence, punishment, rehabilitation, and separation. If an individual already has a criminal history, or comes from a social environment conducive to anti-social acts, the computer will predict quite accurately the individual's chances of entering, or returning to, the social system. At such a point, society's continued presumption of the individual's free will would simply be an excuse for ignoring the failure of another agency, the individual's family, or society itself. Society must ask if the individual has a viable chance of escaping his heredity and his environment, for it seems manifestly unfair to punish an individual whose life pattern could have been (and perhaps was) predicted with almost chilling certainty the day he was born. Punishment is an inappropriate response to conduct which is not voluntarily chosen.

Can society expect to rehabilitate the individual whose crime is not of his own choosing? Here the computer may prove to be more of an advantage, as the resources of society can be focused on the individuals and families who are the problem. If we assume that rehabilitative services after release have an effect upon the recidivism rate, then better focusing of these resources should have the desired effect--less repeater crime. Society would still be faced with the fact that incarceration itself probably does not improve the chances for rehabilitation. On the contrary, the best recourse is immediate probation, regardless of the offense. This concept has been proposed,[17] but runs somewhat counter to the accepted Anglo-American concepts of the criminal justice system.

What of the individual who is not at all rehabilitated? In time, an IMIS will likely predict with startling certainty which criminals will be unresponsive to any rehabilitation efforts.[18] If deterrence and punishment are ineffective, and both prison and social services prove unequal to the task of rehabilitation, what must be society's response? It may be the final listed purpose for incarceration--separation from society itself. At present, prison sentences are set so that an individual will have paid his "debt to society," or will be "safe for society." By examining the records of many prisoners and habitual offenders, the computer may make it inescapably clear that although one cannot predict with total certainty the fact that an individual will remain a criminal, the probabilities are so high as to raise a presumption that he will. Of course, separation works moderately well for the society at large; it is, however, rather expensive. It also runs counter to certain presumptions inherent in Anglo-American jurisprudence.

It will be interesting to note the judicial response to the probabilistic certainties presented via the computer. In the face of increasing evidence that neither statement is true, judges may continue with the fiction that punishment and rehabilitation work. If the only effective methods of alleviating crime are preventative, not rehabilitative, they must be addressed to the children of each city's "one thousand families." Have the parents of the children any say in the matter? If society continues to allocate insufficient resources for either preventative or rehabilitative services, will the logical choice eventually be made to spend all or almost all of the money on prevention? The criminal justice system will then become the end of the line for many individuals, there being neither hope nor expection of rehabilitation.

The judicial system may well refuse to accept the general theory outlined above, or its application in individual cases. Reliance on the present system, and on such presumptions as "innocent until proven guilty," may continue indefinitely. It is unlikely, however, that the general public, once made aware of the statistical probabilities, will allow the current system to continue. This is especially true if the concentration of street crime is elsewhere as it seems to be in Cincinnati. One can expect demands for long-term incarceration, banishment, and perhaps worse.

In summary, the data manipulating characteristics of the large scale digital computer represent a quantum jump over existing methods of information handling. Apart from short term changes in the function of existing social systems, fundamental changes in the systems themselves can be expected, and should be anticipated. Research can profitably be directed to the problems of the future. It is not enough to computerize existing systems, nor to anticipate changes in daily operations or personnel requirements. The criminal justice system, and other social systems, must look to the future to discern the likely impacts of the information revolution.

We turn now to a second possible IMIS impact area, Program Coordination.

Program Coordination

There is a constant problem in complex governance called "Program Coordination" by many people in the U.S. Government. A year ago, Elliot L. Richardson, Secretary, Department of Health, Education and Welfare (HEW), began referring to it as "Services Integration" which is a more accurate definition of the problem but, alas, is not a solution. The problem is exhaustively detailed and well summarized by James L. Sundquist.[19] Although the problem has been recognized for years, no adequate technology or methodology was developed for its solution. Information technology may help solve the services integration problem.

One may begin with the assumption that social services should be integrated at the delivery level. The delivery mechanisms must therefore be examined for generic similarities. For example, a manpower development program may have no direct service in common with an adult basic education program, but each has facilities, staff, funding, clients, and information. Attempts have been made to combine services in a single facility; thereafter, this becomes a multiple service center. Similarly, intake and referral staff for all services may have common training at initial diagnosis of multiple problem clients. Unfortunately, a building housing all services is so large as to repell the consumer. Moreover, the staff that performs the actual service function (fitting eyeglasses, or teaching reading) is usually too specialized to perform more than one or two services well. Such limitation has proved to be a bar to effective integration of services.[20] At present, however, there exists no theoretical barrier to the effective use of information technology to solve social problems. Moreover, information may match the problem to known available resources, schedule service delivery rationally, and monitor progress.

The USAC Program has come to this conclusion via information technology, while HEW's Services Integration staff has reached the same conclusion through its analysis of the problem of social service delivery. Because of information sharing, the staff members of USAC and Services Integration have even coordinated with each other, and are currently proceeding with a common agenda.

One good example of program coordination in information sharing is in St. Paul, Minnesota, one of the original USAC cities. In St Paul, there are three family planning agencies each supported respectively by the St. Paul Health Department (City), the Ramsey County Hospital (County), and Planned Parenthood (voluntary agency). Normally, nothing could be more difficult than to coordinate the activities of three independent agencies. They have separate budgets, separate administrators, perhaps even different fiscal years. As a result, the three agencies can never be sure that some women are not being allowed to drop out of all programs through oversight, or that some indeed are not receiving services from more than one agency. In addition, HEW's National Center for Health Statistics receives summary data on patient visits, not people served. Finally, all three agency directors resist amalgamation knowing that there could then be only one overall director.

The USAC St. Paul Project was able to bring the staffs of the three health agencies together with a representative of HEW's Communicable Disease Center in Atlanta, Georgia. After lengthy meetings, the three agencies agreed to a common form and basic medical records which were in computer readable form.[21] There were, of course, problems of patient privacy and legal responsibilities of the three agencies to be worked out. For a time, it seemed as if they would have to form a greater St. Paul Family Planning Association in order to legally share data. Unfortunately, the private agency developed internal problems and as of this date has not received permission from its state organization to rejoin the group.

At present, if a woman comes to any one of the two agencies' several clinics, she will need to complete only one form which will be processed after her initial visit and then sent to all clinics at all locations. The system is batch processed, and does nothing sophisticated, such as scheduling appointments in advance and notifying patients of the appointments by mail.[22] Currently, there are 3,500 patients (not visit records) on the file shared by the two agencies.

Note the elegance of the system. If viewed from the client's perspective, the two agencies are one--each will know her and her problems as well as the other. If viewed from the perspective of the two program managers, their planning and management must be very similar, since both are based on information which is largely common. Again, the organizational responses of the agencies must be in tune with each other. From the Federal perspective, the National Center for Health Statistics is receiving from St. Paul summary data that is free from distortions caused by overlapping reporting. From all three perspectives, the integration of the operationally based data provides a working relationship which allows two totally independent service providers to function as one. This is an example of services integration in parallel.

The USAC Reading, Pennsylvania, Project provides an example of services integration in series, if that term is appropriate.[23] There, the offices generate permits in the areas of fire prevention, electricity, heating, street maintenance, and sewage treatment.

The information generated is collected into the Bureau of Code Services. Where, formerly, one inspector from each functional area collected all data on a new construction site, now the basic information requirements of all are filled once only. Moreover, progress is being made toward consolidating the functions themselves which will allow one inspector to perform more than one function. More rapid response to requests for service, as well as a more efficient use of manpower, is thus provided.

With a theoretical IMIS in mind, as well as some understanding of the USAC Program, in addition to an acquaintance with two possible IMIS impact areas, we turn now to our concluding discussion, the Current Status of the USAC Program.

Current Status of USAC

The USAC Program is commencing the third year of its contract effort. The budget for fiscal year 1972 is about three and a half million dollars, and for fiscal year 1973 about four and a half million dollars.

In accordance with the USAC objectives, project cities have completed an analysis of all municipal functions, all city organizations, and those of the surrounding counties and special districts providing services and information to the city. The functions covered by the analyses vary from city to city, but some examples include public safety (police and fire); municipal finance; transportation; city engineering; public health; administration; sanitation; public works, traffic control and transportation; budget preparation; city assessment; cash and treasury management; accounting and disbursement; inventory costs; income tax; and civil defense.

In addition, a design of how information can best be provided to serve city operation and management personnel in the performance of their functions has been completed by five of the six USAC cities. Furthermore, the project cities are completing design, and entering development of applications serving the various functions outlined above.

Each city completed an early demonstration in the fall of 1971, the long-range effect of which varies with each city. In Dayton, Ohio, the capability shown in the early demonstration provides for increased return from city cash investment.[24] Moreover, it is based upon automated forecasting of returns from the investment mix and time options which can be varied by the city during continual use of the cash management model. In Charlotte, the early demonstration capability, as now augmented, provides for enhanced receipt and handling of a variety of service requests and complaints apart from emergency public safety operations.[25] The early demonstration in Wichita Falls, Texas, proved the data base design, and the application of the Geographic Base Index System.[26] In Long Beach, California, the early demonstration was recently implemented as a permanent part of city government. This capability provides for use of automation to increase the speed of access to, and comprehensive coverage of, two significant police information requirements, namely Field Interview (FI) reports, and Moving (Traffic) Violations reports.[27] In St. Paul, an early demonstration of a family planning component, as discussed previously, was demonstrated in October 1971. The system is now in operation and serves two agencies: the St. Paul Bureau of Health, a municipal agency, and the Ramsey Hospital, a county agency.

The topics of privacy, confidentiality, and security of information are of utmost importance to the USAC Program. A report on these topics has been prepared for the Program and all project cities by Dr. Edward E. Goldberg of California State University at Los Angeles. The Wichita Falls Project Team pioneered in this field. Their work culminated in an "Ordinance Creating a Data Access Control Plan," which was passed by the City Council. The ordinance aims to protect individual privacy while simultaneously maintaining public access to public records. It states that any person on whom data are filed has the right to review that data, and to correct inaccuracies. Moreover, safeguards are provided to protect the rights of various authorities, such as the police, personnel departments, public health agencies, and other organizations. In addition, the ordinance approved the creation of a data registrar, a data access control board, and defines the functions and responsibilities of the registrar and the board. The establishment of the ordinance, data registrar position, and data control board were initiated by the USAC Program. Another USAC city, Charlotte, North Carolina, has also passed a resolution (which was also initiated by USAC) which aims to protect individual privacy while maintaining access to public records. Finally, no USAC municipal information system contains personal credit data, or other non-municipal function-related information.

USAC is investigating what others are doing to develop service delivery information systems. In addition, USAC is searching reports of other systematic analyses of social service delivery systems. When it is known how applicable the integrating methodology is to other jurisdictions, and when more is known about the delivery systems themselves as well as the reporting requirements each faces, USAC will be in a better position to go into the full-scale development of one or more integrated information systems which serve the needs of many agencies. Since usable data will be on hand, the objective is to assure clients that they will receive services according to a rational plan, and in a manner that can be evaluated.

Because it is a developmental program, USAC is charged with the responsibility of developing operating integrated municipal information systems in each project city. But USAC is also a research effort designed to explore every aspect of an IMIS. Whether an operating IMIS will, in fact, be developed at the termination of the USAC Program thus remains, as it should, an open question.

As with many R&D projects, USAC is often schizophrenic in day to day activities. On the one hand, the staff realizes the importance of "getting something operating." On the other hand, the staff is constantly reminded that perhaps nothing will become operational because, after all, USAC is exploring the fundamental concepts of an IMIS. No one currently knows for certain if IMIS notions are valid. So no one can categorically say that USAC will develop and implement an operating IMIS. Still, at this juncture, USAC has generated evidence which suggests optimism for development. Many IMIS assumptions have been proven valid; and the problems are becoming less technical and more political and behavioral. If the latter can be solved, there is reason to believe an IMIS, as USAC is defining it, will help cities in the decade ahead.

Conclusion

This paper has discussed a theoretical IMIS, the USAC Program, and USAC's current status. In addition, two potential impacts of an IMIS, were discussed. One is the ability that the criminal justice system will have to predict and delineate potential aberrant individuals. It was pointed out that our criminal justice system (corrections and rehabilitation especially) is predicated upon the notion of the individual's free will. If potential aberrants can be predicted, what happens to the concept of free will? If society can predict who its aberrants will be, what should it then do? An IMIS, it is felt, may give society this ability. Finally, a second impact of an IMIS, (called program coordination, or services integration) was discussed. This was viewed as a benefit, and occurs as a result of the horizontal and vertical integration of information, otherwise known as information sharing.

Footnotes

[1] For a more complete discussion of the different approaches to urban information systems development see Kenneth L. Kraemer, "The Evolution of Information Systems for Urban Administration," Public Administration Review, XXIX (July/August, 1969) pp. 389-402; Dennis G. Price and Dennis E. Mulvikill, "EDP in State and Local Governments," Public Administration Review, XXV (June. 1965), pp. 151-155; Public Administration Service and the International City Managers' Association, Automated Data Processing in Municipal Government (Chicago: Public Administration Service, 1966); and Kenneth L. Kraemer and George Howe, "Automated Data Processing in Municipal Government: A Survey," 1968 Municipal Yearbook (Washington, D.C.: International City Manager's Association, 1968).

[2] Kenneth L. Kraemer, "The Evolution of Information Systems for Public Administration," Ibid.

[3] For a theoretical discussion of IMIS concepts see Kenneth L. Kraemer, William H. Mitchel, Myron E. Weiner, and O. E. Dial, Integrated Municipal Information Systems: The USAC Approach (Greenvale, New York: Long Island University Press, 1972).

[4] Kenneth L. Kraemer, et al., Integrated Municipal Information Systems, Ibid., pp. 3.10-3.12.

[5] Ibid., pp. 3.10-3.14.

[6] Ibid., pp. 3.10-3.14.

[7] Ibid., pp. 3.22-3.24.

[8] Ibid.

[9] For USAC development, see Kenneth L. Kraemer, "USAC: An Evolving Governmental Mechanism for Urban Information Systems Development," Public Administration Review, XXXI (September/October, 1971) pp. 543-551; the Department of Housing and Urban Development, RFP H-2-70, "Municipal Information Systems" Washington, D.C., 1970).

[10] See USAC RFP, Ibid.

[11] Kenneth L. Kraemer, et al., Manager's Guide to Integrated Municipal Information Systems (Washington, D.C.: Department of Housing and Urban Development, 1971).

[12] Department of Housing and Urban Development, RFP H-2-70, Ibid., p. C-27.

[13] Ibid., pp. C-11-C-12.

[14] Kenneth L. Kraemer, et al., Integrated Municipal Information Systems: The USAC Approach, Ibid.

[15] From a conversation with the Director of Research, Institute for Criminal Justice, Law Enforcement Assistance Administration, Washington, D.C., April, 1972.

[16] Lon Fuller, The Morality of Law (New Haven: Yale University Press, 1969), 2nd edition, pp. 162-167.

[17] The President's Commission on Law Enforcement and the Administration of Justice, The Challenge of Crime in A Free Society (Washington, D.C., 1967).

[18] In addition to the above example from Cincinnati, Ohio, evidence from the USAC Program also suggests this.

[19] James L. Sundquist, Making Federalism Work (Washington, D.C.: The Brookings Institute, 1969).

[20] Ibid.

[21] St. Paul, Minnesota USAC Subsystem Project, Systems Analysis Task Completion Report (Washington, D.C.: National Technical Information Service, 1972).

[22] Ibid.

[23] Reading, Pennsylvania USAC Subsystem Project, Systems Analysis Task Completion Report (Washington, D.C.: National Technical Information Service, 1972).

[24] Dayton, Ohio USAC Subsystem Project, Systems Analysis Task Completion Report (Washington, D.C.: National Technical Information Service, 1972).

[25] Charlotte, North Carolina USAC Project, Orientation and Briefing Guide (Washington, D.C.: National Technical Information Service, 1972).

[26] Wichita Falls, Texas USAC Project, Systems Analysis Task Completion Report (Washington, D.C.: National Technical Information Service, 1972).

[27] Long Beach, California USAC Subsystem Project, Systems Analysis Task Completion Report (Washington, D.C.: National Technical Information Service, 1972).

NETWORKS
AND
RELATED PROBLEMS

TELECONFERENCING: THE COMPUTER, COMMUNICATION AND ORGANIZATION

CHAIRMAN:
Professor David W. Conrath
Department of Management Sciences
University of Waterloo, Waterloo, Ontario, Canada

PROGRAM COORDINATOR:
Gordon B. Thompson, Manager,
Communications Studies
Bell Northern Research
Ottawa, Ontario, Canada

PAPERS:

Professor David W. Conrath, *"Teleconferencing: The Computer, Communication and Organization"*

A. J. Lipinski, H. M. Lipinski and R. H. Randolph, Institute for the Future, Menlo Park, California U.S.A., *"Computer-Assisted Expert Interrogation A Report on Current Methods Development"*

James A. Schuyler, Northwestern University, Evanston, Illinois, U.S.A., and Robert Johansen, Sociology Department, Upsala College, East Orange, New Jersey, U.S.A., *"ORACLE: Computerized Conferencing in a Computer-Assisted Instruction System"*

Dr. Murray Turoff, Office of Emergency Preparedness, Washington, D.C., U.S.A., *"PARTY-LINE and DISCUSSION: Computerized Conference Systems"*

Professor M. L. Constant and Professor P. L. Seeley, Department of Systems Design, University of Waterloo, Waterloo, Ontario, Canada, *"Computer Mediated Human Communications in an Air Traffic Control Environment: A Preliminary Design"*

DISCUSSANTS:

Douglas Engelbart, Stanford Research Institute, Menlo Park, California, U.S.A.

David Abbey, Ontario Institute for Studies in Education, Toronto, Ontario, Canada

THEME: The general theme of the session will be the role of the computer as an intermediate processor in communication between two or more individuals with a common purpose. What can the computer do in such a role? How might it enhance communication effectiveness over non-processed messages? Can it speed up the pace of sending and receiving between two or more parties to a communication interaction? Can computer aided communication bring about new dimensions to the communication experience? What are the constraints of computer aided communication? How should individuals interface with each other when using the computer as an intermediary?

TELECONFERENCING: THE COMPUTER, COMMUNICATION, AND ORGANIZATION

David W. Conrath, University of Waterloo, Waterloo, Ontario, Canada

When asked to chair a session on "teleconferencing", my first desire was to broaden the scope of the session to include all uses of the computer that would aid in multi-person interactions. That was for two reasons. I felt rather ignorant on the topic of teleconferencing, and thus I wanted to search for a terrain that would be a bit more familiar. Secondly, my feeling is that the computer is going to play an increasingly important role in the communication systems of virtually all organizations, and I wanted an excuse to actively seek what was being done toward studying this problem.

The search revealed several things of note. Most of the work relating the computer to multi-person activity has concentrated on the design and use (or misuse[1]) of information systems for large organizations. While this is perhaps as it should be, the disturbing discovery was that virtually all writers and prophets view an information system as a man-machine interface. Little attention is paid to the man-machine-man aspect of the system, that is, the role of the computer as an intercessor in inter-human communication. For example, the fact that man can receive data at a much higher rate than he can output them, and that a computer might use this asymmetry to an advantage for inter-personal communication is ignored[2]. Admittedly I may have not contacted the right people nor read all the relevant references. My search was sufficiently broad and unbiased, however, that I should have seen more recognition by information systems experts of the role of the computer in inter-personal communication.

Much to my surprise, most of the work that used the computer to further man-to-man interaction fit under the title of "teleconferencing". The existance of the ARPA network is an example, as is the exciting work of Douglas Engelbart at Stanford Research Institute. Thus, with but one exception, and that is on the borderline, the papers that follow are about teleconferencing. The one exception concerns the use of the computer during the communication between the air traffic controller and the pilot, and that too has some conferencing aspects.

PREVIEWS

To give the reader a sneak preview of what follows, let me give a brief introduction to each paper. The Schuyler and Johansen contribution discusses the evolution of a computer assisted instruction scheme into a system that provides for inter-personal interaction. The system at this time is primarily oriented toward computerized Delphi[3] type studies, but the authors describe their system as a research tool, and presumably that entails many configurations of human interaction. In fact, perhaps the most attractive aspect of the system is that it is basically a research tool. In particular, the authors are interested in studying the effects of different forms of feedback to the system's users, on the users. This should help provide guidelines for appropriate feedback characteristics of an interactive system. Most system designers have the tendency to take the form of feedback for granted, and yet it can so easily shape the behavior of the feedback recipient.[4]

Lipinski, Lipinski and Randolph (no, they are not linebackers for Notre Dame) present an outline of a very sophisticated approach to the use of the Delphi technique. As they put it, they are interested in the "joint modeling and problem solving by remotely situated experts," and their purpose is to accomplish this task in the most effective manner possible. Since the Delphi approach provides the *raison d'etre* of the system, rather than just one of its capabilities (as is the case of the Schuyler & Johansen, and Turoff systems) the authors concentrate on solving some of the more critical problems inherent in the technique. Hence, they provide an approach to the problem of varying degrees of uncertainty inherent in the estimates of different experts, as well as an approach designed to gauge just how expert a participant really is. The authors also show a considerable concern about the problem of the validation of Delphi results, a problem which most proponents of the Delphi technique choose to ignore.

The solutions proposed to the problems discussed clearly require an on-line computer system for their execution. Thus, we are provided with a good example of the use of the potential of the computer to solve well defined problems in multi-person communication.

Murray Turoff describes a computer based multi-person communications system with more built-in flexibility than the systems discussed in the other papers. His base of origin, rather than resting in computer assisted instruction or the Delphi technique, was founded upon the principle of the conference telephone call. This has led to the development of a more open format which is conducive to a wide variety of multi-person communication events, especially those involving individuals who are geographically disbursed. He emphasizes the value to the participant of the hard copy output that a computer based conference can provide. Unless there is a recorder at a face-to-face or telephone conference, a participant must rely on his memory or personal notes to recall what took place. This is often rather hazardous.

Turoff also put his system into an economic context. He first evaluated the time value of a computer based system using a typed input and a printed output, with the time value of an audio or face-to-face conference. He then calculated the economic costs of alternative ways to get multi-party contributions. Thus, we have one paper which stresses the research capability of its teleconferencing system, another which stresses the processes needed to make effective use of group interaction for forecasting and problem solving, and a third which looks closely at the economics of a computer based system.

In the remaining paper, Constant and Seeley depart from the more traditional view of computerized teleconferencing to take a different look at what the computer might do to augment inter-personal communication. Recently we all have become aware of the problems of the Air Traffic Controller in guiding planes through the air to the point of landing, keeping them sufficiently spaced so that no two attempt to occupy the same space at the same point in time. The problem has become almost the classical one of information overload.[5] Yet little has been done to relieve the ATC of the computational aspects of his job so that he can concentrate on the communication and management portions of it. On the one hand, the Constant and Seeley proposal looks similar to the classical information system, at least as it is described in textbooks. On the other, they point out that their system truly assists in the communication process among several persons, a process in which time is of the utmost importance. The computer has the capability of rapid data processing and feedback which allows the individuals involved to make more effective use of the increasingly scarce resource, time.

OTHER POSSIBLE USES OF THE COMPUTER AS AN INTERMEDIARY

Each of the questions posed by me to provide a broader theme for this session has been handled by one or more of the papers presented. Nevertheless despite these most welcomed contributions, I still feel somewhat at a loss since I could not find a wider variety of research on the uses of the computer for assisting man-to-man interfaces. For example, what might the role of the computer be during bargaining and/or negotiating sessions, an obvious multi-person interaction event? Could the computer help make the process more "rational" by clarifying the issues and indicating the consequences of alternatives as they arose? Or is the "politics" of such an event so crucial to the process that computer mediation would hinder rather than help the process?

What about the use of the computer as a device to assist in the evaluation of ideas, without the critic having to be identified so that the fear of public (or private) censure might be avoided? Does this have possibilities, especially when the idea rather than the person uttering the idea is presumably what is to be evaluated? Or are we sufficiently incapable of evaluating ideas independent of the person offering them that value judgements would become random rather than ordered? Also, might not the anonymous use of the computer to interact with others lead to major problems of organizational structure? Can we cope with the problem of the new structures that evolve, including the possibility of anarchy?

Why isn't organizational budgeting done on a real-time basis on the computer? Think of the time and frustration such a process might save. One could rather easily establish priorities beforehand that were in line with the objectives of the organization, as well as establish the necessary constraints. Then the input for financial requirements could be made directly from those who were going to be involved in the expenditures without all of the political ramifications that typically arise. But again think of the implications of the changes that would take place in the sources of power. Could these be handled to the benefit of all, or would chaos result?

Turoff reflected on the use of high speed computer output in conjunction with the high input capability of the human, in constrast to our low output capability. Why haven't others spent more time finding ways for the computer to utilize this imbalance to our mutual advantage during interaction processes? Also, while current input into the computer from the human communicator is either restricted by typing or speaking rates, cannot this restriction be lowered by better uses of coding and the like? Would not there be more impetus to do research in this area if the computer was seen as a device through which man might better relate to man?

The computer is already involved in inter-personal processes whether we wish to recognize that fact or not. The question is do we choose to use the computer to foster these processes, or do we wish to still view the role of the computer in terms of man-machine interface. If the computer is to become an undisguised blessing to mankind, then we had better turn more of our attention to how the computer might affect the relations between men.

Footnotes

1. For a critque of management information systems see Ackoff, Russell L., "Management Misinformation Systems," Management Science, 14 (1967), pp. B-147-157.

2. This has not been ignored by those who have developed computer assisted instructional aids, however.

3. The Delphi technique for obtaining estimates about the likelihood of uncertain events was developed at Rand Corporation. While the papers presented at this session should give adequate insight into the technique, further information can be found in: Dalkey, Norman and Helmer, Olaf, "An Experimental Application of the Delphi Method to the Use of Experts," Management Science, 9 (1963); and in many Rand Memoranda which have been issued about the subject.

4. A substantial amount of the literature in the fields of psychology and sociology has been devoted to the effect of "feedback" on both individual and group behavior. Two relevant examples are: Pryer, Margaret W. and Bass, B. M., "Some Effects of Feedback on Behavior in Groups," Sociometry, 22 (1959), pp. 56-63; and Zajonc, R. B., "The Effects of Feedback and Probability of Group Success on Individual and Group Performance," Human Relations, 15 (1962), pp. 149-161.

5. Perhaps the best statement of the problem is the classic: Miller, George A., "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information," The Psychological Review, 63 (1956), pp. 81-97. Also, I have been told that Richard Meier has done a study on information overload specific to the problems of air traffic control, but no one has been able to give me the reference.

COMPUTER-ASSISTED EXPERT INTERROGATION: A REPORT ON CURRENT METHODS DEVELOPMENT

Andrew J. Lipinski, Hubert M. Lipinski, Robert H. Randolph
Institute for the Future
Menlo Park, California, U.S.A.

Abstract

This paper describes a major program currently under way at the Institute for the Future to develop and test a system for on-line interaction among experts via a network of computer terminals. The authors review the Institute's existing methods for eliciting and processing expert judgments, then discuss the operational procedures by which these methods have conventionally been implemented. In light of the shortcomings of these existing procedures, the authors argue an urgent need for on-line group modeling capabilities as the next leap forward in judgmental-research methodology. A description is given of the Institute's recently completed prototype computer conferencing system, whereby remotely situated respondents can participate in Institute inquiries via the ARPA computer-communication network and the Rand Corporation's PDP-10 computer. Problems encountered thus far are discussed, as are future plans for system refinement and extension. The paper concludes with some brief philosophical observations on the state of our art, acknowledging that much work remains to be done on the still-primitive tools of decision making.

1. Introduction

The future of any society depends on how the material, spiritual, and intellectual resources of that society are applied to the problems that face it. This paper will review the progress being made and that we anticipate making at the Institute for the Future to improve the effectiveness with which our society can draw on the intellectual resources of its "wise men." More specifically, we shall describe recent developments in the Institute's techniques for eliciting and processing expert judgment, including the computer assistance upon which these developments depend. The high-point of these developments, toward which the bulk of the Institute's methodological efforts are presently directed, is joint modeling and problem solving by remotely situated experts. Special attention will therefore be given to the key tool which promises to facilitate such activities--the "D-net" computer system for on-line interaction of geographically distributed experts currently being developed at the Institute.

Group Judgment

Given the complexity and size of our society, and the world-wide explosion of information, no one is likely to be fully informed, not even in what he considers to be his primary discipline. Moreover, because the problems that beset us become increasingly interdisciplinary, it now appears that no one man is likely to be even adequately informed. Therefore, to utilize expert knowledge effectively, it is often necessary to arrange for communication among experts in addition to elicitation of information from them individually. Helmer and Dalkey's[1] original contribution in developing the Delphi inquiry method was to recognize this important principle--that to employ the intellectual resources of experts in problem-solving, one must assure adequate communication among them.

In the past, and to a large extent even today, decision-makers have usually obtained the advice of their experts in qualitative, rather than quantitative, form. Although to a degree such will always continue to be the case, some modern decision-makers recognize that qualitative advice is open to ambiguity. This ambiguity can be of several kinds. It can be in the mind of the adviser who may not be precise enough in describing what he really believes. Or it can be in the mind of the decision-maker who may interpret expert advice differently from what was meant by the adviser. Thus, even if the adviser understands perfectly the message he intends to convey, the problem of expressing the message and its subsequent interpretation may lead to a misunderstanding. Moreover, the adviser is often asked to express an opinion about events or trends which lie in the future. Such developments, however, are usually determined by stochastic processes which themselves are often poorly understood. For example, the future cost of crude oil, an important raw material for many industries, will depend upon future developments in the Middle East. Similar examples closer to home might involve a future decision of the Supreme Court or even the outcome of a presidential election. In such cases, the expert adviser is bound to be less than certain and ought somehow to state the degree of his uncertainty. His uncertainty might be expressed by using terms such as "it is likely," or "it is plausible." However, the desire for precision to reduce ambiguity--and, parenthetically, the desire to store and process such information--requires that the state of information of the expert be quantified.

The key contribution to such quantification comes from the field of subjective probability, in which the language of probability allows precise encoding of the expert's state of information. Thus, to continue with the example already mentioned, the future price of crude oil can be expressed as a probability distribution assignment, based on the opinion of anyone we wish to consult, typically an expert. Because of the imperfect understanding of the situation (i.e., absence of a fully credible model) and the fact that even if such a model were available, no one expert could be expected to have complete information, the decision-maker would usually consult more than one adviser. Throughout history, decision-makers somehow have been able intuitively to evaluate their advisers and combine their advice, and then merge it with their own opinion to arrive at an improved state of information. This process, absolutely key to obtaining better decisions in society, is currently becoming better understood and is gradually being quantified. A major purpose of this paper, then, will be to describe the gradual development of the Institute's techniques for quantifying, evaluating, and combining expert judgments, as well as the computer assistance without which these developments could not have been achieved (see section 2).

Group Modeling

However important the gathering of judgmental data may be, it is not fully adequate by itself to allow maximally valid decision-making. Drawing conclusions with respect to outcomes of contemplated actions requires an understanding of how things work, of how a particular segment of reality can be represented. Here, as before, a desire to reduce ambiguity would lead us to a preference for a quantified representation consisting perhaps of a set of mathematical equations, constituting a model. Until recently, modeling of complex interdisciplinary issues was seldom even attempted. When it was attempted, it was done by individual researchers and, as a result, the credibility of the model tended to be low. Currently, however, entire teams of expert co-workers are being employed to build models. For example, Dennis Meadows' dynamic model of world population and resources[2] incorporates contributions from a team of experts, each expert being responsible for a particular set of computer statements relating to his area of expertise.

One can visualize situations, however, when convening such a team at one time in one location would be a luxury that a decision-maker could not afford, either because of expense or because of the urgency with which the advice is required. There is a need, therefore, for methods which enable geographically separated experts to model jointly. Sections 3 and 4 of this paper, then, will describe the Institute's current attempts to improve understanding of how this could be done, concentrating on the computer-conferencing system currently being developed.

2. Methods for the Processing of Expert Judgments

Elicitation of Point Estimates

Even though the type of data typically used in the Institute's research is essentially judgmental in nature, there was an early realization among the Institute staff that the sheer amount of such data being collected would require computer assistance. It was also recognized, however, that computer assistance could not be obtained, at least not easily, without having the data in quantified form.

The essential first step toward being able to quantify judgmental data was to specify, in advance, the intervals (technically the "lottery") into which a possible answer from a respondent might fall. This meant constraining the expert--by not asking him, "What do you think might happen?" but, instead, specifying alternative ways of how things may happen and by which his judgment might be encoded. These alternatives could then, in turn, be given numerical values.

Specifying a lottery is a substantial intellectual exercise because, unless one does understand the subject, one very often does not know what answers are possible. The easiest lottery is one which is constrained by definition--for example, an assignment of probability to a specified event, which can only be made within the range 0 to 1. With other variables, not so conveniently constrained, it is necessary to specify the scales within which possible answers may fall. When this was done in early Institute studies, we allowed for bad guesses on our part by inviting the respondents to modify our scales if necessary, although that of course required that we incorporate a scaling factor in the subsequent encoding process.

When considering future occurrences in early Delphi inquiries, our lotteries normally treated these occurrences as Bernoulli events, one branch of the lottery representing that future "world" in which the given development occurred and the other branch representing all other possible "worlds." Given such a lottery, each respondent was asked to supply a single-number estimate of the probability of the first branch. Unfortunately, this procedure suppressed any knowledge that might have been obtained about the respondent's personal uncertainty regarding that number. When the results of such questioning were summarized, the uncertainty of the group or panel of respondents was represented by making a histogram of the number of respondents assigning particular values to each variable in question. Although this was definitely not a true probability distribution representing the group information about the subject, we now recognize that it did offer some approximation to such a distribution. If the encoding intervals were made uniform in width, it was possible to calculate the fractional median of the group response by dividing the area of the histogram in half (as shown in Figure 1). Quartiles were determined analogously. The resulting statistics could then be converted to indicate the corresponding values of the random variable (an example relating to probability assignment is shown in Figure 2).

Elicitation of a single number from each respondent had an important consequence: it was not possible to acknowledge that some experts are more "expert" than others and to use this knowledge explicitly. However, it was possible to help the reader of our reports to form his judgment regarding the event by giving him information about the relative expertise of panel members. The only measurement used to evaluate the respondent's expertise was his own self ranking. A histogram in which the expertise of each respondent is indicated by a number is shown in Figure 3.

Early Experimentation

In order to evaluate the initial Delphi approach to the elicitation and processing of expert judgment, we made a number of statistical measurements in our early studies,[3] trying to determine some fundamental properties of Delphi such as reproducibility of results (panel to panel), shift of median response after two or more rounds of questioning, and the amount of convergence of opinion to be expected from feedback and reestimation. For example, our experiments corroborated Dalkey's observation[4] that there is virtually no shift in the median round to round and that the shift is comparable to the random "noise" that would result if one panel of experts rather than another were consulted.

We noted that, on the average, the histograms of expert responses became tighter on subsequent rounds, indicating convergence towards a group median. However, convergence is a tricky concept. For example, if the random variable were to be the numbers generated by a roulette wheel, an expert knowing that the wheel is not biased might assign equal probability that the ball would land in any slot. If we would constrain an expert respondent to supply a single-valued answer, e.g., red sixteen, for the result of the next spin, any convergence resulting from group judgment would obviously be inconsistent with the state of information available to the experts.

Finally, regarding a shift of the median toward the true value of the random variable, our one result was disappointing. We had, by chance, a respondent who very nearly qualified as a "clairvoyant." The hypothetical event, which was to take place by 1985, was stated in the questionnaire as:

> "A time division switch for telephone service, or telephone and wideband service, will be placed on trial in one of the large metropolitan offices where the exchange trunks have been all converted to the T-1 and T-2 carrier systems."

Our respondent's comment was: "There shouldn't be any trials of this type in 1985, since the first trial is now scheduled for 1975 or 1976." Reading this comment, one detects a note of authority. Surely, the event is highly likely to have occurred by 1985 if this expert is correct about its occurrence ten years earlier. Yet, given that information, the panel in question shifted less than the control panel that did not have this information. We feel that in face-to-face conference, our expert would have quickly convinced the others about the impending trial, and would have caused a larger shift in the assignment of probability.

Elicitation of Distributions

A major improvement over the Institute's early procedures was the decision to encode the state of information of each respondent as a probability distribution. The probability distribution easiest to extract is the mass function, i.e., cumulative probability. Our knowledge of how to determine the value of an expert was so inadequate when this form of opinion quantification was first attempted that we took refuge in consulting many more experts than we would now. If one chooses to consult many experts, it is not practical to extract the full mass function from every one. Even with the assistance of a computer program we developed, such an interview is laborious, taking about 15 minutes per expert per question. We decided, therefore, to collect only points corresponding to the 10%, 50%, and 90% probability assignments. A typical format for a questionnaire of this type is shown in Figure 4. These three projections of Figure 4 supply three points on the mass function for each of the years considered; these points in turn are used to arrive at a density function and group density function as shown in Figures 5 and 6.

Measurement of Respondent Expertise

As our experience grew, it became more and more obvious that experts differ greatly in the degree of their understanding of a given subject, their substantive knowledge, and the attention they give to our questionnaires. That realization led us to another major improvement--made possible by the above-mentioned decision to express the state of information of each respondent in the form of a probability distribution--namely, the weighting of respondents' opinions depending upon their expertise, as shown in Figure 7. The idea here was to calibrate each expert in terms of his substantive knowledge of the subject and, having calibrated him, to combine individual judgments of individual experts in some relation to the degree of their expertise.

To do this, it was necessary that the criteria for self ranking be more explicit than those previously used. For example, to qualify as an expert, we would now require that our respondent be able to answer affirmatively such questions as these: "If the National Science Foundation or some other scientific body convened a meeting of the 20 or so best qualified people in the country, would you expect to be invited?" "Are you familiar with the literature in your subject?" "Have you worked in this particular area recently?"

At the same time, we employed expertise tests other than self-ranking. For example, we employed consultants to phrase questions in the vocabulary of the profession. If the respondent did not understand the terms, our reliance on his judgment would be reduced, because even if he did know what he was talking about, he would have difficulty in conveying his opinion convincingly to our other experts. Similarly, we checked the quality of respondents' comments. We found that our experts' comments vary from utter trivialities and platitudes to true scholarly discussions of the subject matter. Comments that convey most to us are those in which the respondent elucidates his assumptions and indicates: (1) about which of these assumptions he feels reasonably expert, and (2) which assumptions he would like us to question.

In addition, we evaluated the degree of attention which the expert paid to the problem because even a highly qualified expert may be interrupted while filling out a questionnaire. A simple way to check the expert's attentiveness is to ask him to project a trend on a linear scale and then, some 10 questions later, repeat the same question using a logarithmic scale. Ideally, an expert ought to complain that he is being asked to do unnecessary work. On the other hand, if he gives a straight line projection on both scales, we have reason to be dubious about that particular expert's answers. Another check on the internal consistency of the expert's views is to ask him for an estimate of X and Y, then sometime later in the questionnaire ask for an estimate of X/Y. While not expecting precise correspondence, we would nevertheless hope that these three estimates would be reasonably consistent with each other.

Once the experts in an inquiry have been ranked according to these various measures, we can determine the weighting factors to be applied to each by asking our client a question such as: "If you were the decision-maker and had to consider an opinion supplied by this expert in comparison with an opinion of the 'reference' expert, how much less seriously or more seriously would you consider it?" It may be observed that this judgment is an approximation to the likelihood function in the Bayes Theorem. To combine the expert opinions themselves, we currently use the method described by Winkler[5] of adding weighted probability distributions. The result of such addition is shown in Figure 6. When smoothed, it approximates a group distribution, or the distribution that might have been obtained by one imaginary "aggregate" expert representative of the entire panel.

Measurement of Assessing Ability

One of our latest efforts has been to introduce calibration of respondents with respect to their assessing ability (independent of their substantive expertise). The fact that most people tend to make assessments which overstate their certainty has been well documented.[6] To measure each respondent's bias in this regard, we have followed the suggestion of Peter Morris[7] in using an almanac procedure which is appealing because it is amusing as well as effective. Using the *Guinness Book of Records* as a reference, we ask the assessor to bracket the value of an unknown record. For example, if the record is that for the longest cigar ever, the assessor might reply that the true value probably lies anywhere within 3 and 11 feet. At the end of 20 questions of that type, the respondent is asked to estimate how many actual records (out of 20) he believes will, in fact, lie within the intervals he specified, how many below his lower bound, and how many above his upper bound. As measured by Alpert and Raiffa[6] and confirmed by our own measurements, usually no more than a third of the estimated number of items fall within the intervals specified. Therefore, after measuring each expert's tendency to overstate (or understate) his certainty, we can correct each probability density supplied by that respondent (such as shown in Figure 5) by an appropriate amount. This additional calibration step represents only a slight modification of the computer program already in use for weighting responses according to substantive expertise.

Pre-Modeling

Perhaps the most important recent development in the Institute's methodological approach deals with the value of information contained in a given item of inquiry. Simply stated, the problem is to decide which potential future developments to study. A variable that is omitted early in an inquiry is unlikely to be discovered later. All conventional modeling tests, such as sensitivity checking, are designed to eliminate variables, not to discover missing factors. We have therefore begun using several methods to identify key variables and key events prior to more detailed study.

One of these methods involves introducing the respondent to an imaginary "clairvoyant" and allowing him an imaginary five-minute interview. What questions would the respondent ask of the clairvoyant, given such an opportunity? Another approach is that of asking the respondent to describe two future worlds, one of which is considerably worse and one considerably better than the world he actually expects. The definitions of "considerably better" and "considerably worse" are that, in the respondent's mind, there is one chance in 10 that the actual world may be even worse or better than the world described. Yet another approach is for the respondent to identify those assumptions about the subject being studied in which he has least confidence—for which additional information might thus be expected to have greatest value.

Once key issues have been identified by any of these methods, they are then submitted back to the respondents for ranking according to two final criteria: importance and susceptability to intervention. Importance typically is evaluated as follows: if a given development were to happen or not happen, what would be its differential effect? In making these estimates, a respondent can use any time preference and time span. The second criterion is the ability of the decision-maker to do anything about the problem. If nothing can be done, the problem may still be very important, but the value of information about it will be small.

3. Existing Procedures for the Processing of Expert Judgments

In the preceding section of this paper, we have described the development of the methods currently used at the Institute to quantify expert opinion. Now let us turn to the operational procedures we have used in the past to implement these methods in actual research inquiries.

In the past, the Institute has obtained expert opinions on various subjects principally through use of a carefully designed program of sequential interrogations conducted by means of questionnaires sent through the mails.* These sequential interrogations are interspersed with information and opinion feedback, thus providing some exchange of reasoned opinion among the experts preparatory to each successive round of questioning.

This procedure for information gathering appears to have some unique merits and weaknesses. Some of the major advantages of this procedure over other ways of eliciting and processing judgmental data are listed below:

- it maintains attention directly on the issue;
- it provides a framework within which individuals with diverse backgrounds or in remote locations can work together on the same problem;
- it tends to minimize the tendency to follow the leader and other psychological barriers to communication;
- it tends to minimize specious persuasion;
- it provides equal opportunity for all to be heard as individuals; and
- it produces precise documented records.

Some of the disadvantages of the conventional mailed-questionnaire procedure are the following:

- it is relatively time consuming, both in the mailing and receiving of questionnaires (even one round takes a month or longer) and in the coding and keypunching of the data required for batch computer processing and analysis of the results;
- the interaction between the expert and the investigator occurs infrequently and only through questionnaires--as a result, an expert cannot easily obtain clarification on what is meant by a particular question, seek additional background data, and so forth;
- the interaction between experts occurs only infrequently and only in the form of information feedback--as a result, the panel may not be able to frame a fully agreed-upon lottery or model for any given question under consideration;
- the procedure may lack the stimulation provided by face-to-face conferences;
- complete anonymity does not inspire credibility among the experts; and
- because of the infrequency of interaction and because the procedure is already overly time-consuming, the credibility of the results are not easily checked. The pressure for expeditious production of substantive results usually prohibits adequate attention to significance testing, replication, proximate validation, and the like, as well as the more complex methodological features discussed in the previous section (e.g., probability training, measurement of assessing ability, premodeling, etc.).

This last difficulty is by far the most serious, for it calls into question the validity of all study results obtained by conventional procedures. If we wish to make efficient use of the complex procedures available for analyzing and validating data, to have more interaction among the experts, and to do all this within practicable limits of time and cost, it now seems clear that we must use some form of computerized conference. Thus, researchers at the Institute for the Future and elsewhere[8] have come to take a growing interest in the concept of "on-line" expert interaction.

4. New Procedures for the Processing of Expert Judgments: The D-Net Conferencing System

On-Line Interaction of Experts

The basic idea of on-line group conferencing is roughly as follows. To begin an inquiry, the investigator(s) proceeds just as he would in a conventional mailed-questionnaire elicitation study; that is, he defines a topic of discussion, selects a panel of experts on that topic and invites them to participate in the inquiry, assembles background materials relevant to the topic, and perhaps formulates specific questions designed to illuminate the topic. Then, instead of drawing up a questionnaire, he feeds the preliminary information into a computer system and arranges for each

---

*This procedure has been called the Delphi method and is discussed in Dalkey and Helmer, *An Experimental Application of the Delphi Method to the Use of Experts*.

expert to establish communications with the computer system via a computer terminal convenient to him. The computer system then conveys questions, answers, group feedback, and other information to and from each expert individually, according to programmed procedures, while the investigator monitors the proceedings and intervenes as necessary. Information gathered at each step in the inquiry is tabulated by the computer to aid the investigator in directing the conference, for use as feedback, and of course as a record of the inquiry's results.

Although a computer conferencing system of this sort could serve as more than an automated and thus somewhat more efficient version of the conventional questionnaire method, it offers several potentially powerful advantages. First, it allows full consideration of one question at a time, if desired, including perhaps several rounds of question, answer, feedback, and even rephrasing of the question--thus permitting the experts to deal fully with that question while it is fresh in mind rather than having to wait a month or more between rounds of questioning. Second, the computer conferencing system allows any respondent to call for background data, explanation of questions, or other assistance at any time. Third and most important, such a system offers great flexibility in the types of information to be exchanged among conference participants, including not only the usual forecasting questions and answers but also specialized materials such as tutorials, expertise-calibration measures, problem-modeling questions, and so forth. Because the expert's contact with the system is interactive, it is entirely possible for the system to give each user highly individualized treatment, responding to his special needs in the manner made familiar by the concept of "programmed learning". It is precisely because of this flexibility that on-line group conferencing offers a real hope for implementing the various procedural innovations previously discussed as remedies for the problem of the incredibility of inquiry results.

It must be acknowledged that the basic idea of group-via-computer interaction is not new. Olaf Helmer, the Institute's Director of Research, first referred to the concept in print during 1967.[9] Other experimenters have conducted a number of interesting experiments in this field, notably Murray Turoff of the Office of Emergency Preparedness.[8] At least one commercially leased program (available at a cost of about $10,000 per year) is said to perform some of the functions we have been discussing.[10] It appears, however, that the performance of previously developed systems is generally inadequate for the full battery of procedures needed for obtaining maximally credible judgmental data For this reason, system development efforts were begun at the Institute in late 1971, directed toward provision of a computer conferencing system which can eventually incorporate any features deemed desirable in an operational judgmental-research tool.

Prototype D-Net System

In its present prototype version, the Institute's on-line conferencing system consists of some thirty interlocking computer programs, totaling roughly 5,000 lines of assembly language code. This basic system, now being tested with three terminals on-line, will serve as a test bed for implementing additional features and for conducting experimental conferences. The system resides on the PDP-10 computer at the RAND Corporation in Santa Monica, California, with access to the system being provided via the computer-communication network operated by the U.S. Advanced Research Projects Agency.

This early version of the system is simpler than that ultimately envisioned, in several ways. First, it assumes that conferences will initially be more interrogative than conversational; in other words, the investigator is expected to pose questions and the experts to respond, with little opportunity for direct communication among respondents other than by means of feedback which the investigator chooses to give them. And second, the investigator is assumed to be an individual, working at one high-speed terminal, rather than a team of individuals working at separate terminals as is eventually intended.

Even in this initial system, however, performance of two basic tasks must be provided for. The first task is leading each expert through an inquiry, feeding information and questions to him, and accepting input from him. The second task is allowing the investigator to control the substance, structure, and flow of the inquiry. To implement these tasks, we have written two main program packages, known appropriately enough as EXPERT and CHAIRMAN, which communicate with each other through shared files. In addition to these two main programs, we are in the process of developing an ever-increasing repertoire of numerical and textual information-processing routines as well as a number of other auxiliary programs for such specialized tasks as editing of keyboard inputs, control of unusual types of terminals, establishing direct communication links between the investigator and a respondent seeking assistance or information, and so forth.

The basic structure of the main programs can best be described from the standpoints of the individuals using them. Upon logging into the RAND computer, each expert receives a copy of the EXPERT program, which then leads him through the inquiry in accordance with directives obtained from the CHAIRMAN program. Thus in any inquiry there are several copies of the EXPERT program, each communicating with and controlled by the one CHAIRMAN program. When a respondent logs in, EXPERT first gives him an opportunity to familiarize himself with background briefing statements and reading materials appropriate to the inquiry. If the investigator desires, tutorials on such subjects as probability theory would be administered at this point. Next, the respondent enters the main body of the inquiry where he is called upon to answer a series of questions posed by the CHAIRMAN program. In answering each question, the respondent proceeds through three phases: the question-review phase, the verbal-modeling phase, and the numerical-response phase. Progress of the entire panel through these phases is governed by a series of software "switches" which are set by the CHAIRMAN program. Input provided by the respondent is stored in indexed files which can then be accessed by processing and display routines.

Unlike the EXPERT program, which operates according to its own built-in logic and allows little control by its user (the respondent), the CHAIRMAN program is under complete control of its user (the investigator). At all times, CHAIRMAN is either executing some command explicitly given by the investigator or else awaiting another such command. By means of these commands, the investigator can call for various routines which monitor and direct the flow of the inquiry. These include, for example:

- a display showing which respondents are on the system, along with their terminal numbers and other identifying information;
- a display summarizing the progress of each respondent (or of the panel as a whole) through a particular question;
- a display showing the current status of the

inquiry control switches; and

- a routine to set these switches and thus direct the flow of the inquiry.

In addition, the investigator can call a variety of text and numerical processing routines:

- a routine for replying to requests from individual respondents for specialized background data (it is assumed that the investigator will have access to reference materials appropriate to the subject of the inquiry);
- routines to search the indexed files of input information for response "packages" of any given type, display these responses on the investigator's terminals, and return selected responses to the panel;
- routines for rephrasing or deleting existing questions from, and introducing new questions into, the inquiry; and
- routines for gathering, processing, and displaying (in alternative formats) respondents' estimates of:
  - single numerical quantities, and
  - three-point probability distributions.

Initial Problems

The above-described basic system, which is now operational, has, in fact, all of the features of the Institute's previous batch processing capabilities and, further, permits a degree of on-line interaction. We are now using the system to experiment with and seek remedies for a number of problems encountered or anticipated. Some of these problems and their tentative solutions are listed below:

- Many experts either do not know how to, or prefer not to, type. This problem is compounded by the fact that there are some inevitable differences between using a computer terminal and using a typewriter (e.g., control characters, separate carriage return/linefeed, etc.). Currently, we are attempting to eliminate from our system as many confusing elements as possible and to build in provision of full instructions for users who require them. As a last resort, we recognize that it may be necessary for some experts to use a research assistant or secretary as an amanuensis.
- Similarly, the procedure by which each participant in an inquiry must gain access to the network, log into the computer, and initiate his participation is unavoidably complex. Eventually we hope to have at least part of this procedure performed automatically, perhaps with the user being contacted by the computer rather than vice versa.
- Assembling a sizable panel of high-level experts to participate in an inquiry at one time, even though at diverse locations, may prove impossible. One promising answer to this difficulty is to make the inquiry system asynchronous, allowing each expert to enter, leave, and reenter an inquiry at whatever times he finds convenient, over a period of, say, a day or two. This solution, unfortunately, carries some problems of its own, particularly regarding procedures for iteration of questioning on each given issue, but even these difficulties seem to be not insoluble.
- Once a respondent has entered an inquiry, he may resent the constraints imposed upon his participation by the strict preprogrammed logic of the inquiry system. To combat this problem, we are taking steps to increase the flexibility of the user's interaction with the system. Ultimately, inquiries will be conducted in a parliamentary mode, with numerous types of inputs acceptable at any time but with such inputs processed according to an orderly hierarchy of priorities. The basic structure for this mode of operation, based on Robert's Rules of Order, has already been prepared. For the time being, it is hoped that the dialogue nature of the interaction will permit the expert to feel as though he alone were being queried, thus allowing him to concentrate as much as possible on substantive considerations.
- In the present version of the system, the investigator uses his single program, CHAIRMAN, to control the entire inquiry. It is evident, however, that the conduct of a full-scale inquiry will create a serious overload on the investigator. (Even in initial tests using only three terminals, the investigator has been a bottleneck because respondents must often pause while he enters commands telling the system what to do next. This problem will be somewhat alleviated when the investigator's responsibilities are divided into three specific functions to be performed by three separate individuals using subsets of the original CHAIRMAN program [to be known as CHAIRMAN, UMPIRE, and EDITOR].) Further, we intend to alter the investigator's programs so that more of the conduct of an inquiry is automatic. Not only will greater automation of the investigator's role speed up the flow of an inquiry, but it will also relieve the investigator of his present obligation to have a detailed understanding of the inner workings of the system itself. One of our goals is to eliminate any need for users of the system, either respondents or investigators, to have any understanding of these inner workings.

Future Refinements and Extensions

Some of the important changes contemplated for incorporation into later versions of the D-net system are described in the previous section. Among the other specific features tentatively planned to be added and experiments to be run are the following:

- A programmed-learning tutorial on subjective probability theory will be developed, whereby a reasonable degree of uniformity may be obtained among respondents in their understanding of subjective probability concepts.
- In order to encourage a group of interacting participants to ask for relevant information in "modeling" underlying societal processes, programs will be prepared and tested which permit the participants to send information requests to and receive information from the umpire. If the umpire cannot meet these requests from available Institute sources, he will retrieve the necessary information by using a telephone line to a reference librarian.
- Because respondents' time and effort are valuable, it would be helpful if the optimal number

of respondents, questions, and iterations for each topic could be determined with an eye to (a) the marginal amount of information expected to be gained by adding a respondent, question, or iteration, and (b) the associated marginal costs (which may be far from negligible). Programs will be prepared which generate estimates of these various factors, thereby providing useful "stopping criteria" for each phase of the inquiry.

- When asked to answer a question in the course of an inquiry, a respondent may ask to delay his response in order to be able to look up, compute, or otherwise obtain information he regards as helpful in arriving at his response. The question then is whether it is better to accommodate such a request, at the risk of upsetting the synchronism in the processing and feedback of the respondents' inputs, or to insist on an "off-the-cuff" estimate, at the risk of frustrating that respondent and perhaps diminishing the quality of his response. Experiments will be conducted to throw light on this problem, using the above-mentioned concept of the value of information.
- Experiments by Institute staff members and by Norman Dalkey at Rand have shown that respondents tend to be swayed more by the weight of the opinion distribution of their copanelists than by arguments presented in defense of opinions different from their own. A balance has to be found between those forces which draw individual responses (assumed here to represent probability density functions) toward the previous response distribution and those which might draw them toward the "correct" distribution. We will carry out controlled experiments in an effort to ascertain the relative importance of various forms of information feedback in causing a respondent to move his response closer to the "true" distribution.
- In any real-life Delphi study it is necessary to calibrate the expertise of the respondents. We propose to conduct an experiment which uses experts to answer almanac-type questions, and which, in addition, requires each respondent to rate himself as well as each of the other participants (identified by name or otherwise) as to his relative expertise regarding each of the questions posed. This will permit an evaluation of the absolute as well as the relative merits of using self- and peer ratings to improve the accuracy of group estimates.
- One component of the validation of the Delphi process involves the stability of judgment. We will conduct a controlled experiment comparing the opinions derived from a panel of respondents with (a) those of a comparable panel polled at the same time and (b) their own opinions taken a few days or weeks later. The subject matter for this experiment will be short-term forecasts (that is, forecasts which can be verified within a few months).

5. Concluding Reflections

This paper was couched as a progress report because we realize only too well that our understanding of how to apply expert judgment to current decision problems is in its infancy--and this despite six millenia of highly organized society behind us. We cannot with complete confidence ascribe this lack of progress, as does Ardrey,[11] to the lack of communication between our "reptilian" brain and our "human" brain overlay, because we already start with the assumption that group modeling, or group judgment, should be arrived at only by logical inference from commonly assumed axioms (scarcely reptilian activities). Our modesty regarding our accomplishments is heightened by the realization that, intuitively, most people are able to process expert advice quite well, yet we cannot explain how we all do it. In our increasingly distrustful society, the question will surely be asked, over and over again: "Tell me, how did you arrive at that decision?" Therefore, we consider it worthwhile to continue chipping away, even if the chips are small and the wall of ignorance is large.

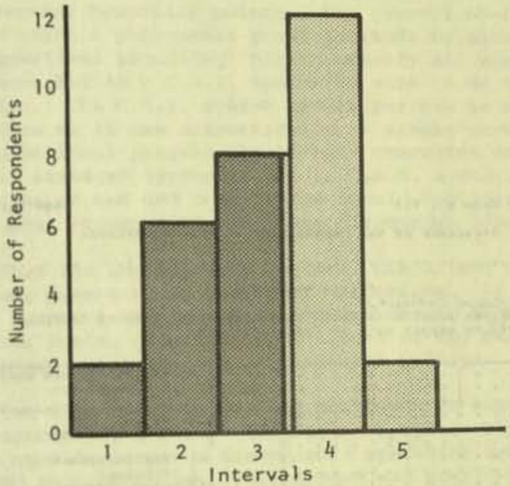


Figure 1 AREA MEDIAN OF HISTOGRAM

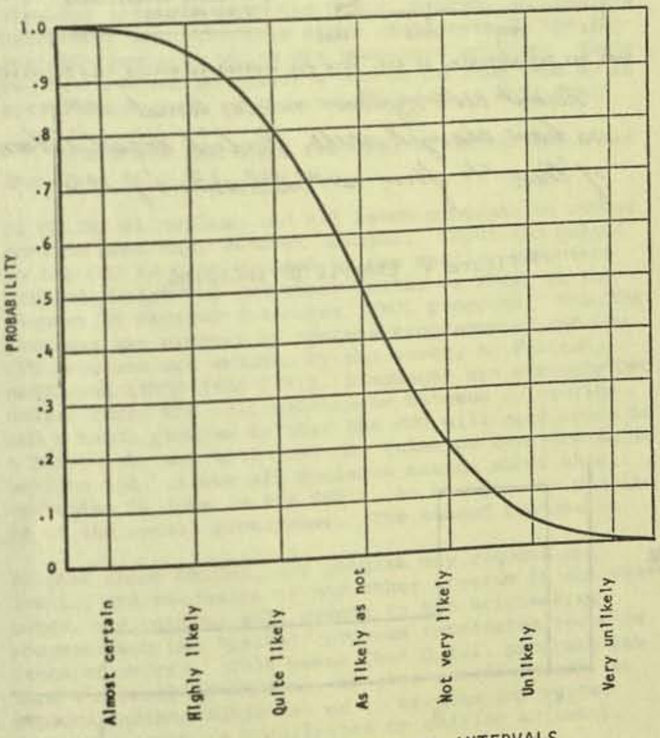


Figure 2 TRANSLATION OF CODING INTERVALS INTO PROBABILITY ASSIGNMENT

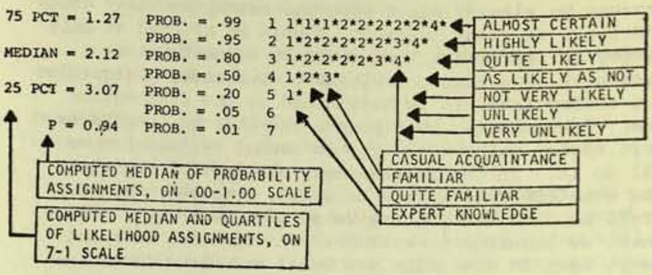
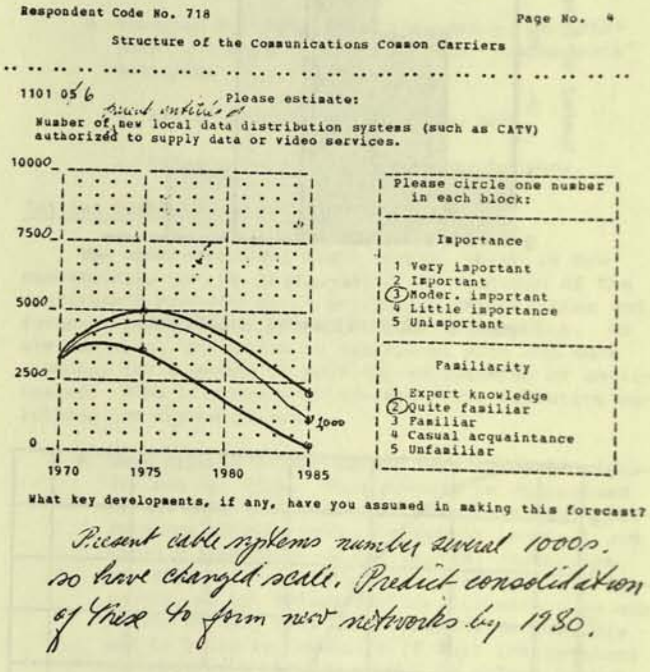


Figure 3 ENCODING OF EXPERTISE



Respondent Code No. 718 Page No. 4

Structure of the Communications Common Carriers

1101 056

Please estimate:

present entities of

Number of new local data distribution systems (such as CATV) authorized to supply data or video services.

Please circle one number in each block:

Importance

1 Very important
2 Important
[✓ 3] Moder. important
4 Little importance
5 Unimportant

Familiarity

1 Expert knowledge
[✓ 2] Quite familiar
3 Familiar
4 Casual acquaintance
5 Unfamiliar

What key developments, if any, have you assumed in making this forecast?

Present cable systems number several 1000s.
so have changed scale. Predict consolidation
of these to form new networks by 1980.

Figure 4 EXAMPLE OF ENCODING

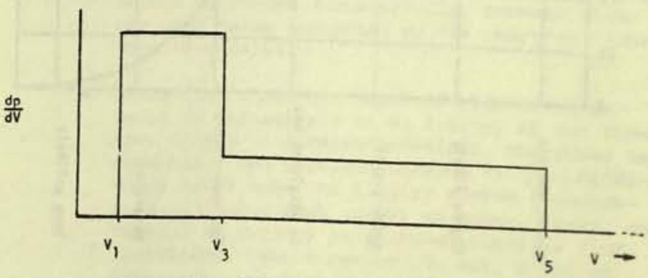


Figure 5 DERIVATION OF THE PROBABILITY DENSITY FUNCTION

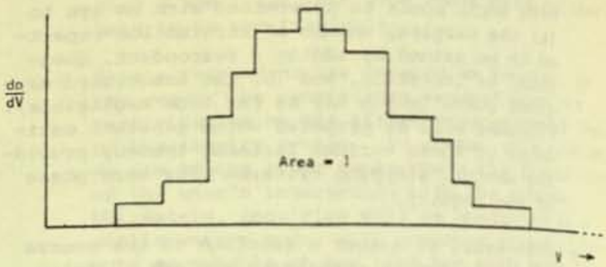


Figure 6 PROBABILITY DENSITY OF GROUP'S OPINION

Expertise	Ranking
Expert Knowledge	16
Quite Familiar	8
Familiar	4
Casually Acquainted	2
Unfamiliar	1

Figure 7 RESPONDENT FAMILIARITY SCALE

References

1. Norman Dalkey and Olaf Helmer, "An Experimental Application of the Delphi Method to the Use of Experts," *Management Science*, Vol. 9, No. 3 (April 1963).

2. Donella H. Meadows, Dennis Meadows, et al., *The Limits to Growth* (New York: Universe Books, 1972).

3. Paul Baran and Andrew J. Lipinski, *The Future of the Telephone Industry, 1970-1985*, Report R-20, Institute for the Future (September 1971).

4. Norman Dalkey, *The Delphi Method: An Experimental Study of Group Opinion*, Memorandum RM-5888-PR, The Rand Corporation (1969).

5. Robert L. Winkler, "The Concensus of Subjective Probability Distributions," *Management Science*, Vol. 15, No. 2 (October 1968).

6. Marc Alpert and Howard Raiffa, "A Progress Report on the Training of Probability Assessors," unpublished document, Harvard University (September 1970).

7. Peter A. Morris, "Baeysian Expert Resolution," doctoral dissertation, Stanford University (1971).

8. Murray Turoff, "'Party-Line' and 'Discussion': Computerized Conference Systems," International Conference on Computer Communication, Washington, D.C., October 24-26, 1972 (February 22, 1972).

9. Olaf Helmer, *Systematic Use of Expert Opinions*, P-3721, The Rand Corporation (November 1967).

10. "'X-Basic' Backs Delphi Conferencing", *Computerworld* (January 26, 1972), p. 17.

11. Robert Ardrey, *The Social Contract* (Dell Publishing, 1970).

"ORACLE": COMPUTERIZED CONFERENCING IN A COMPUTER-ASSISTED-INSTRUCTION SYSTEM

James A. Schuyler, Northwestern University
Evanston, Illinois, U.S.A.
Robert Johansen, Upsala College
East Orange, N.J., U.S.A.

Abstract

The purpose of this paper is to describe the evolution and operation of ORACLE, a computer program for computerized conferencing and interactional research. ORACLE exists as an essential part of the NUCLEUS Computer-Assisted-Instruction system at Northwestern University. The paper describes the operation of ORACLE as a communications facilitator for teachers and systems designers, a research tool (a kind of dialogical questionnaire), a "record-keeper", and as an instrument for curricular feedback. Thus ORACLE extends beyond the traditional realm of C.A.I. systems, and suggests alternative uses for computers in education. Both the systems requirements for the program and its sociological dimensions are discussed.

Introduction

To this date, most people in the fields of education and computer science still associate the term Computer-Assisted-Instruction (C.A.I.) with that old bugaboo - Programmed Instruction. It is true that Programmed Instruction (P.I.) is the most frequently used mode of interaction in C.A.I., but there is much more which should and could be done with C.A.I.

When the first C.A.I. computer systems were developed, in the late 1950's, they were patterned after Programmed Instruction models, and the computer had to be able to:

1) Present information to the student by typing on a typewriter or "writing" on a cathode-ray-tube (C.R.T.),

2) Receive and interpret information typed by the student, or entered by a light-pen. This included making a match with author-specified responses stored in the computer, which were generated in advance and in very great detail.

3) And all of this had to be done so that no more than a second elapsed, so that the student wouldn't begin taking the machinery apart.

To meet these demands, computer manufacturers built special data-terminals, special computers and special software (programming) systems. The IBM 1500 system was one of these. Even though a number of these systems are still in use, they did not have the expected market impact, and are not being promoted today.

Almost at the same time, computer scientists (though they had not taken that name yet) began to realize the power of having a computer at each researcher's fingertips. So-called "on line" programming (or real-time) was developed, where many computer users have simultaneous access to the same computer, with a complex set of programs (an Operating System) doing the job of coordination necessary to give each of them the appropriate amount of space in the computer and the appropriate speed of service. The general on-line system was designed to service a number of people, each requiring fair amounts of computation or text-manipulation. But C.A.I. systems were designed primarily to store and retrieve materials to be displayed for the student, with infrequent calculations to determine branching points. The general on-line system gives a programmer great latitude by allowing (and sometimes requiring) him to specify *all* steps to be taken; but in a C.A.I. system we wish to do the opposite. The C.A.I. system should perform as many functions as it can *automatically*. A single error in a computational program should halt execution (as it does in standard systems), but in C.A.I. a warning should be issued and compilation should continue, since most errors will not cause disastrous results.

Thus the two types of systems, C.A.I. and general on-line, have gone in different directions. Recently there has been a lot of talk about bringing them back together again, primarily as a result of the BASIC revolution which started at Dartmouth College.

The first step we took at Northwestern was that of constructing a C.A.I. language and system which could operate inside a large-scale operating system on our CDC 6400 computer. This system was previously used only for batch-jobs (research and students). The SCOPE operating system runs up to seven jobs at one time, and the ONLINE system (developed here at Northwestern) allows one or more of those seven "jobs" to be sliced up among many on-line users. In addition, the on-line users may create and submit batch jobs, or they may interact with the C.A.I. system. A pseudo-batch mode lets the user enter statements in the job-control-language, which are processed on-line. These may be combined to create a job to be submitted as an actual batch job, or they may be used to call "CAI".

There are two major features of ONLINE which are important to C.A.I. users:

A) ONLINE allows any and all batch programs to run as on-line programs, without changes. Input and output on the CDC 6400 is handled by ten smaller computers (called *Peripheral Processor Units, or PPUs*) at the request of Central Processor (CPU) programs. The PPU programs are written by Systems Programmers, and the CPU programs are written by the users, so Fortran, BASIC and LINGO (the C.A.I. language) are strictly CPU code. Thus, the only difference between an on-line and a batch program is that the PPU will read cards in a batch job, and will read the teletype keyboard in an on-line job. Since all decision making about this operation is done in the PPU's, no changes are required of the normal programmer. The second feature is

B) that under ONLINE, any program may request the loading and execution of any other program in the computer, and control will return to the originating program when the "called" program terminates (even in cases of error). This means that C.A.I. programs can form a *network of lessons*, to give remedial work, to explore further subjects, or to execute any system functions (such as compilation or editing of text).

This is even more important when you realize that a student now needs practically no knowledge of the operating system. One single control program, *CAI*, is able to ask the operating system for execution of any other program, or to diagnose errors, or to ask the computer operators to mount backup tapes from which lessons can be re-loaded. It can call the EDITOR to let the student (or teacher) create his own lessons, call the compilers to process these programs, and when the author approves it can make them a permanent part of the C.A.I. library. These are things which otherwise would only be available to those persons able to remember the proper job-control-language. And since ONLINE does not alter the basic structure of the SCOPE operating system, the Systems Programmers still have all the power they need, while the casual user has everything taken care of automatically.

This *mixed* C.A.I./on-line system has now evolved into a larger project, *the NUCLEUS* (Northwestern University Computer-based Learning and Educational Utility System) which combines teaching-programs with the very large program-library available on our 6400, for data-analysis.

Computerized Conferencing Within the NUCLEUS

It should be clear from the introduction above that the NUCLEUS is oriented toward expanded or exotic instructional uses of the computer. The ORACLE, which is the focus of this paper, is a prime example of this kind of extended C.A.I. It functions as a kind of *record keeper* for the system by allowing feedback of information which the student initiates, rather than just the information the lesson author requests (the standard C.A.I. model). ORACLE also functions completely outside of traditional C.A.I. by serving to lubricate communications in various kinds of groups and decision-making bodies. Perhaps the best place to begin an explanation of ORACLE is to quote directly from the user's write-up:

> *This program is intended to serve as a catalyst for group communications and information exchange. Through this program, a group leader or committee chairman can (with no programming knowledge) establish an "ORACLE group" for his own committee. Group members can then sign into the ongoing conference at their own convenience, making this type of interaction especially useful to groups which meet infrequently, or are widely separated geographically. This system of continuous conferencing also offers the opportunity to consider such things as long-range goals, which are often passed over in normal committee meetings due to more pressing immediate concerns.*
>
> *Thus computerized conferencing such as ORACLE offers can often provide a more efficient and less expensive means of information transfer than such things as committee meetings, conference telephone calls, panels, etc. Also the initiator has a variety of options in terms of the structure of group interaction. These are outlined in the diagram included with this write-up, but each potential group initiator need only call ORACLE on the Northwestern C.A.I. system, identify himself as a group initiator, and the options will be expressed in the form of questions from that point on. Several "public" conferences are usually listed at the start of the ORACLE, and these can serve as models for users who have little experience with computerized conferencing. It is helpful to do some planning in advance.*
>
> *Though this kind of conferencing is quite new, some experimentation has been done with what has been called Delphi Conferencing. (See list of references at the end of this write-up, particularly those by Murray Turoff[8,9]). There is a rather broad literature on the RAND-developed Delphi technique, but it is basically a method for gathering information about alternative futures. It has traditionally involved the polling of "expert" opinion about the probability of certain events happening by a particular date in the future. Also, results are usually sent back to participants in the form of "rounds", giving each person the option of changing his or her ratings based on the responses of others in the sample. Thus Delphi has been used as a consensus encouraging device in certain groups.*
>
> *ORACLE offers numerous Delphi-like options which can be used by groups to consider events relevant to them. These may be events which are established by the group leader and/or added continuously by group members.*
>
> *Oracle in Interactional Research:*
>
> *Using ORACLE, social science researchers can design interactional questionnaires which are much more flexible and dialogical than traditional paper-and-pencil instruments. From the perspective of the researcher, much of the busywork of test administration and organization can thus become automated. Responses are stored by the system as they are given by participants according to a format established by the researcher. Comments and reactions of each person can be recorded in a similar fashion. This data can then be retrieved periodically by the researcher in whatever form is most usable to the ongoing purposes of the project. In this way, orderly summaries of the responses can be available at any time to the researcher or (to the degree determined by the researcher) to each participant.*
>
> *This type of research raises some uncertain questions, however, which need to be explored further. Traditional survey research has usually sought to assess group attitudes and provide possible insight into a social system. The interactional computer model introduces another potential purpose: providing the participants with feedback about themselves in relation to particular social groups. Their individual responses can also be stored with others in the sample for later collective analysis. The uncertainty comes at the point of the relationship between the collective research purpose and the individual feedback purpose. Much work needs to be done to test the effects of various kinds of feedback in this kind of research format. ORACLE can provide the vehicle for such testing and experimentation.*

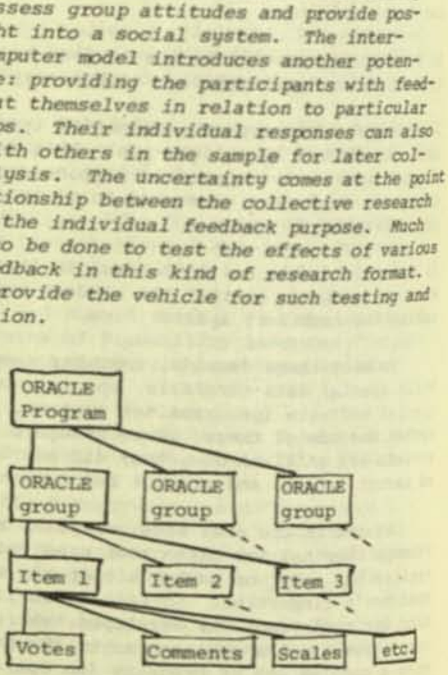


Figure 1. ORACLE Structural Diagram

Basic Options Offered by the ORACLE

I. Initiator options:

Anonymous conference, or names recorded ?
Mandatory events for each participant ? How many?
Can events (items) be added by participants ?
Additional events or events to be deleted by
the initiator ?

Type of conference: Nonvoting (comments only).
Voting - a. Comments recorded ?
 b. No comments ?
 c. Secret ballot ?
Delphi - a. Probability ? What scale ?
 b. Desirability ? What scale ?
 c. By what date will item take place ?
Other - a. Design your own scale ?
 b. Likert-type scale ?

Feedback options:
No feedback.
Feedback after each event is completed by
the participant.
Feedback only after all events have been examined
by the participant.

II. Participant options:

Variable, depending on the format established by the initiator of each conference.

III. Eavesdropper options:

Complete printout (on high-speed printer) of
conference proceedings ?
Comments only since a certain date ?
Comments only on a particular event ?
Comments only from a particular person ?
Data for analysis ?

Technical Dimensions of ORACLE

The ORACLE is written in the C.A.I. language LINGO, making the interactions quite easy to program and change at will. The SCOPE (CDC 6400) operating system allows two kinds of files of information - *local*, which exist only during a particular job while the user is on-line, and *permanent*, which exist after the user has left the system. ORACLE itself handles one permanent file for *each* ORACLE group. Each user of ORACLE gets his own copy of the ORACLE program, so many users may be active, each using a different ORACLE-group permanent file. At the current time, because participants may modify information on a group file, the file for a particular group may be accessed by *only one participant at a time*. However, an "eavesdropper" may gain access and take a snapshot of the conference while a participant is using it.

File Structure:

Our files are sequential (as opposed to direct access). This means that it takes processing time to access an item far down the file; however it also allows us to produce variable-length items quite easily, and the processing time on the CDC 6400 is negligible (amounting to no more than a second in a participant's session). Information for a single conference is stored in a file, and is divided into *records*, each record corresponding to one "item" or "event". An additional first record contains the name of the conference initiator and participants, or it may include a *flag word* indicating that the entire conference is anonymous. If the conference has some mandatory items, there is also an indication of how many there are. The rest of the records contain the events of the conference. Each item's first line is a title supplied by the person who entered the item. All titles are printed when a participant enters the conference. Following the title is a complete description of the item or event (exactly as entered by the item's creator), which is printed only when the participant selects the items he wishes to see. A description may contain as many lines as necessary; there is no limit. In a direct access file system, we would have had to limit the size of an event in order to allocate disk space properly. We also have avoided the need to allocate a fixed-length block of central memory for use as an item-buffer; we only need enough to read one line at a time. The one fixed piece of information is a special line indicating the date by which the item or event is to take place. A date is stored *exactly as the initiator enters it*, no conversion required, so "by fall 1973", or "Jan 1" or "3/14/84" or "never" are all acceptable.

The existence of the LINGO C.A.I. language has also made possible another programming trick; a number of *flags* then appear in the record, one per line and in any desired order. Since ORACLE is built around this approach, it is extremely easy to change by adding new types of flags, or deleting old ones, and because these do not change the record structure, all old ORACLE groups will still be acceptable:

1) ORACLE scans each "flag" line, one at a time, and ignores any flag-word it doesn't recognize. Therefore if we throw away VOTE scales at some future time, we don't need to go through all ORACLE groups deleting the flag, since it will be ignored in the future.

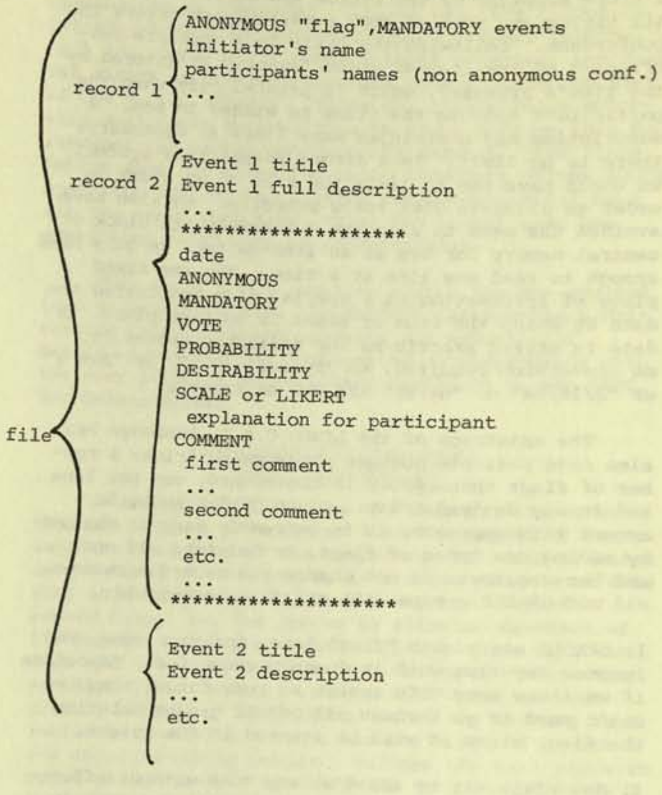
2) New flags may be added at any time without affecting old groups, since those groups will not contain any of the new flags. And if a participant adds a new item to an old ORACLE group, he may freely use the new flag, since there is no fixed record structure.

Flags currently used by ORACLE include:
ANONYMOUS (Indicating that the particular *item* is
 anonymous, not the whole group)
MANDATORY (Requiring that all participants consider
 the event)
VOTE (A yes/no vote is to be taken on the item)
PROBABILITY (Probability ratings are requested for
 this event or item)
DESIRABILITY (Desirability ratings)
SCALE (A self-designed scale or Likert-type scale is
 desired. Information to be displayed to the
 participant is also included with this flag.)
COMMENT (Indicating that open-ended comments follow)

The scale to be used for each possibility is also included with the flag. For example, DESIRABILITY usually ranges from -100 to +100, and following that is a tally of all participant ratings so far. When a self-designed scale is used, the initiator decides what the scale values mean.

Because of this approach, we can easily add new options to ORACLE as new applications arise, though the standard ORACLE contains all options we discuss in this paper, and the group initiator never does any programming himself.

Figure 2. ORACLE Permanent-File Structure



The Evolution of the ORACLE

The development of ORACLE was definitely related to the frustration which social science researchers have historically had with questionnaires as a medium for gathering information in social systems. In fact the methodologists in the field have emphasized that those social scientists who were *not frustrated* by questionnaires *should have been*. Yet, in spite of a rather general dissatisfaction with questionnaires, they are often seen as an inevitable and unchanging tool for social research.

The limitations of questionnaires became especially clear to us within the last year when we were faced with the problem of developing a systematic feedback process within a loosely constructed curriculum for graduate study in religion (at Garrett, which is affiliated with Northwestern and on the same campus). The curricular model itself is highly individualized and is based on a process orientation toward education. Research in an environment such as this dramatized the inadequacies of a feedback network which relied only on questionnaires as a source of information. There was a need for a method of research which was consistent with the dynamic nature of the social system with which we were dealing.

It was at this point that we became aware of the work with *citizen sampling simulations* which is being done by Stuart Umpleby[10] and others at the University of Illinois PLATO project. The idea of giving immediate feedback to each participant in a survey prompted a further exploration of this work and similar efforts which used the computer to facilitate group communications and feedback. Though our own interests were focused on specific types of social systems, the PLATO simulations prompted thinking about alternative possibilities. Why couldn't curricular feedback benefit from computerized questionnaires which would provide immediate feedback to each participant ? Why couldn't conferencing systems be developed which would enhance feedback and communication at a more subjective level? Could such things as course evaluation be done more effectively if they were computer-based? It was questions like these which prompted the development of ORACLE.

The first attempts to implement a computer interactional research format at Northwestern and at Garrett were oriented around simple multiple choice questionnaires. Our goal at that point was merely to test the feasibility of this type of computer application on our system. After the potential for this type of applicationa was demonstrated, we began programming a modified Delphi format which we had already developed as a paper-and-pencil instrument. The format was computerized in the same form as the paper-and-pencil version, except that each person was given the cumulative mean scores of other participants after he had given his own ratings. Thus each person was given immediate feedback about how the other persons in the sample had rated various events being considered.

Though the program mentioned above was primarily a test of the operating system at Northwestern, it also revealed some interesting factors in research methodology. For instance, what effects will continuous feedback of peer group ratings have on the overall results of the survey? In other words, how will the individual-feedback-purpose affect the original collective-research-purpose? Questions of this sort arose in the early stages of pre-testing, and made it clear that considerable experimentation would be necessary to determine the effects of this method of social inquiry. The problem with the original program, however, was that variations in it to test different formats were difficult to implement without changing the entire program.

For this reason, we decided on the idea of constructing a central program (later called ORACLE) which would not involve any fixed content, but only serve as a framework for various uses in different kinds of projects. Since the interactional research format is so similar to the programming needs of computerized conferencing, we decided that these options should be included in one program. In taking this route, we were attempting to provide a medium for the experimentation which seems necessary to feret out the potentials for interactional research and computerized conferencing. Though the latter may be somewhat different in its purpose, it relates directly to the purposes of computer-interactional research as we are now using them.

The ORACLE program has added a new dimension to the concept of curricular feedback and communication within our own educational system. It might appropriately be considered an expansion of sociological research, or a branch of the already-existing C.A.I. system. At this point, however, the primary purpose of ORACLE is to provide the framework for experimentation in developing alternative feedback channels and communication processes.

Summary: Specific Characteristics of ORACLE

Since ORACLE is similar in some ways to other conferencing systems, it seems appropriate to concisely state its characteristics:

1) ORACLE serves as a central *library program* which allows many uses within the same program. Each par-

ticular conference group, with its own data requirements, is then stored on a separate file.

2) The major types of conferences available through the ORACLE are:

a. Non-voting conferences, merely raising items or issues for discussion,
b. Voting conferences,
c. Delphi-like conferences, including pure Delphi, anonymous, probability only, etc., or variations,
d. Interactional research (e.g. dialogical questionnaires)

3) The use of the C.A.I. language LINGO allows for considerable flexibility in program design, with most responses given in normal English, rather than in coded symbols. LINGO is a user-oriented language and thus is especially suited for users with no programming knowledge. These persons are major users of the ORACLE.

4) ORACLE provides a framework for persons to develop and test various formats for computerized conferencing and interactional research. This seems important at this point, since there is still a shortage of information on the effect of computerized conferencing on group process and decision-making.

5) ORACLE is an integral part of a C.A.I. system (the NUCLEUS) with many possible applications within the educational system. It is intended to serve as an impetus for development of more creative uses of the computer in education. Thus, ORACLE is available to all members of the university community - including students, professors and administrators.

References

(1) *Alpert, D. and Bitzer, D. "Advances in Computer-based Education", Science, Vol 167*
(2) *Dick, W. "Sources of Information on Computer-Assisted-Instruction Today", Educational Technology, March 8, 1970*
(3) *Feingold, Samuel "PLANIT - A Language for Computer Assisted Instruction", Datamation, Vol. 14, Number 9*
(4) *Hall, Thomas W. "Implementation of an Interactional Conference System", Proceedings of 1971 Spring Joint Computer Conference*
(5) *Hickey, Albert Computer-Assisted-Instruction: A Survey of the Literature, Entelek Corp, Newburyport, MA, 1968*
(6) *Johansen, Robert "Sociological Applications of Interactional Computer Systems: Toward a Dialogical Research Model", to be published in 1972 in Arthur B. Shostak, Sociology in Action, Second Ed.*
(7) *Schuyler, James The LINGO System , Vogelback Computing Center, Northwestern University, Evanston, Il 60201*
(8) *Turoff, Murray "Delphi Conferencing (i.e. Computer-Based Conferencing with Anonymity", Journal of Technological Forecasting and Social Change, Volume 3, Number 2, 1971*
(9) *____________ "Delphi and Its Potential Impact on Information Systems", Proceedings of 1971 Fall Joint Computer Conference*
(10) *Umpleby, Stuart "Structuring Information for a Computer-Based Information System", Proceedings of 1971 Fall Joint Computer Conference*

Appendix: Sample ORACLE Run on Teletype

```
 06/14/72  15.14.45
***HI
TYPE YOUR STUDENT ID NUMBER? 98.6
TYPE YOUR NAME? jim schuyler
WHAT DATA-TERMINAL ARE YOU USING? t.i. 725
***NUCLEUS
$$$$$$$$$$ NORTHWESTERN UNIVERSITY $$$$$$$$$$
                 LINGO SYSTEM

TYPE THE NEXT LESSON NAME, OR TYPE "HELP" FOR HELP,
OR "LESSONS" TO SEE A LIST OF PUBLIC LESSONS.
? oracle
 06/14/72  15.15.48
***.1

YOU ARE PARTICIPATING IN A CONTINUING
EXAMINATION OF ALTERNATIVE FUTURES.

TYPE "HELP" FOR A LIST OF PUBLIC CONFERENCES, OR TYPE
THE NAME OF THE CONFERENCE YOU WISH TO ENTER
? computers in education
THE CONFERENCE YOU NAMED DOESN'T EXIST.
DID YOU MAKE A MISTAKE OR DO YOU WISH TO INITIATE
A NEW CONFERENCE
? a new one, please
IS THIS TO BE A PUBLIC CONFERENCE
? sure
IS THE CONFERENCE TO BE ANONYMOUS
? nope
WILL PARTICIPANTS BE ALLOWED TO ADD ITEMS
? yes
HOW MANY MANDATORY ITEMS WILL THERE BE
? 1
TYPE YOUR NAME EXACTLY AS YOU WISH TO HAVE IT
RECORDED FOR THE CONFERENCE.  THIS WILL IDENTIFY YOU
AS THE INITIATOR OF THIS CONFERENCE.
? jim schuyler
TYPE YOUR ITEMS FOR THIS ORACLE GROUP NOW.
MANDATORY ITEM...
GIVE A ONE-LINE "TITLE" FOR THE ITEM.  IT SHOULD BE
FAIRLY DESCRIPTIVE BY ITSELF.
? instructions to participants
NOW GIVE FACTUAL DETAIL OF THE ITEM. DO NOT INCLUDE
YOUR OWN OPINIONS (THIS IS IMPORTANT) SINCE YOU WILL
BE ALLOWED TO COMMENT LATER.  TYPE "///" WHEN YOU
HAVE FINISHED THE DETAILS.
? please consider all items in this conference. we
? are examining alternatives for uses of computers in
? education.  add your own ideas as additional items.
? ///
BY WHAT DATE WILL THIS TAKE PLACE
? no date
IS THIS ITEM TO BE ANONYMOUS
? no
DO YOU WANT PROBABILITY RATINGS
? no
DO YOU WANT DESIRABILITY RATINGS
? no
DO YOU WANT PARTICIPANTS TO VOTE ON THE ITEM
? no
DO YOU WANT A SPECIAL SCALE
? no
WILL PARTICIPANTS BE ALLOWED TO COMMENT
? no
WOULD YOU LIKE TO ADD OTHER ITEMS
? sure thing
GIVE A ONE-LINE TITLE FOR THE ITEM.
? computer conferencing replaces trustees meetings
NOW GIVE FACTUAL DETAIL OF THE ITEM.  TYPE "///"
ON THE LAST LINE OF DETAIL
? computerized "oracle" is used by trustees of the
? university to decide university problems from
? their own homes and offices. No longer are they
```

```
? required to appear on campus for meetings. ///
BY WHAT DATE WILL THIS TAKE PLACE
? 1980
IS THIS ITEM TO BE ANONYMOUS
? no
DO YOU WANT PROBABILITY RATINGS
? yes
DO YOU WANT DESIRABILITY RATINGS
? i do
DO YOU WANT PARTICIPANTS TO VOTE ON THE ITEM
? no
DO YOU WANT A SPECIAL SCALE
? yes
TYPE THE QUESTION TO BE ASKED FOR THIS SCALE
? how will the trustees vote on this item (1=yes,0=no)
WHAT IS THE LOWER BOUND ON THE SCALE
? 0
WHAT IS THE UPPER BOUND
? 1
DO YOU WANT ANOTHER HOMEMADE SCALE TOO
? no
WILL PARTICIPANTS BE ALLOWED TO COMMENT
? yes
WOULD YOU LIKE TO ADD OTHER ITEMS
? no
YOUR MODIFICATIONS AND COMMENTS HAVE NOW BEEN ADDED
TO THE CONFERENCE PERMANENTLY.
```

Sample of Participant's Run on Teletype

```
06/14/72  17.30.01
***.1

YOU ARE PARTICIPATING IN A CONTINUING
EXAMINATION OF ALTERNATIVE FUTURES.

TYPE "HELP" FOR A LIST OF PUBLIC CONFERENCES, OR TYPE
THE NAME OF THE CONFERENCE YOU WISH TO ENTER
? computers in education
YOU MAY EITHER EAVESDROP ON THE CONFERENCE OR YOU
MAY FULLY PARTICIPATE. WHICH DO YOU SELECT
? participate
PLEASE TYPE YOUR NAME BELOW, EXACTLY AS YOU WISH
TO HAVE IT RECORDED FOR THE CONFERENCE
? bob johansen
HERE ARE THE NAMES OF THE PARTICIPANTS IN YOUR GROUP

JIM SCHUYLER

HERE'S A LIST OF THE "TITLE" OF ALTERNATIVE FUTURES
WHICH WE ARE CONSIDERING IN THIS EXCHANGE.

1. INSTRUCTIONS TO PARTICIPANTS.
2. COMPUTER CONFERENCING REPLACES TRUSTEES MEETINGS

***.10
SINCE YOU ARE A PARTICIPANT, YOU MAY BE ASKED TO RATE
THE ITEMS YOU VIEW.  YOU MAY ALSO COMMENT ON THOSE
ITEMS YOUR CONFERENCE INITIATOR HAS MARKED FOR COMMENT
***.2

ITEM 1
INSTRUCTIONS TO PARTICIPANTS.
PLEASE CONSIDER ALL ITEMS IN THIS CONFERENCE. WE
ARE EXAMINING ALTERNATIVES FOR USES OF COMPUTERS IN
EDUCATION.  ADD YOUR OWN IDEAS AS ADDITIONAL ITEMS.

ITEM ENTERED BY JIM SCHUYLER

WHICH ITEM NEXT ? (TYPE "DONE" TO QUIT)
? 2

ITEM 2
COMPUTER CONFERENCING REPLACES TRUSTEES MEETINGS
COMPUTERIZED "ORACLE" IS USED BY TRUSTEES OF THE
UNIVERSITY TO DECIDE UNIVERSITY PROBLEMS FROM
THEIR OWN HOMES AND OFFICES. NO LONGER ARE THEY
REQUIRED TO APPEAR ON CAMPUS FOR MEETING.
ITEM ENTERED BY JIM SCHUYLER
HOW PROBABLE IS THIS ITEM BY 1980 (0 TO 100 SCALE)
? 0
HOW DESIRABLE IS THIS ITEM BY 1980 (-100 to 100 SCALE)
? 50
HOW WILL THE TRUSTEES VOTE ON THIS ITEM (1=YES,0=NO)
? 0
DO YOU WANT TO SEE THE COMMENTS ON THIS ITEM SO FAR
? no
NOW ENTER YOUR COMMENT. USE AS MANY LINES AS NECESSARY.
PUT THREE SLASHES ("///") AT THE VERY END OF IT.
? trustees are social animals, they won't give up
? their social occasions and meetings. ///
WHICH ITEM NEXT ? (TYPE "DONE" TO QUIT)
? done
***.5
WOULD YOU LIKE TO ADD ANY ITEMS
? not today
YOUR MODIFICATIONS AND COMMENTS HAVE NOW BEEN ADDED
TO THE CONFERENCE PERMANENTLY.



JOB COST  $ 1.70
```

"PARTY-LINE" AND "DISCUSSION" COMPUTERIZED CONFERENCE SYSTEMS

Murray Turoff
Office of Emergency Preparedness
Washington, D.C., U.S.A.

> "The machine yes the machine never wastes anybody's time."
> The People, Yes
> Carl Sandburg

Note: The views and opinions expressed in this paper are those of the author and do not necessarily reflect official policy of the U. S. government.

One of the most time consuming functions facing management and staff in most organizations is the need to participate in various forms of group communications. The methods available for group discussions (as opposed to presentation formats) are actually quite limited in number. They include committee or panel meetings, telephone conference calls, Delphi exercises and Delphi conferences. The purpose of this paper is to describe two computerized versions of the basic telephone conference call. These two systems, titled "PARTY-LINE" and "DISCUSSION", provide a completely new and unique group communication form and have particular utility for geographically dispersed groups desiring immediate written records of the proceedings. Considerations of both relative effectiveness and cost between computerized and verbal conference modes are quantified and discussed.

Introduction

Both systems described in this paper were created as adjuncts to the EMISARI system (Emergency Management Information System and Reference Index) operating at the Office of Emergency Preparedness. EMISARI is a Management Information System which emphasizes the concept of providing a lateral communication structure among staff members working on information of interest to management. EMISARI was first introduced at OEP to handle the reporting and information dissemination functions associated with Phase I of the Wage-Price Freeze (Fall 1971). The introduction of a supportive real-time computerized conference call capability to EMISARI fulfilled the requirement for conducting immediate discussions among about 15 individuals spread out over the 10 regional OEP offices and the National Headquarters. Most of these discussions dealt with clarifying sudden changes in the reporting formats or providing instruction and/or clarification on changes to the EMISARI system. In all these situations, the existence of a written record for the conferees to take away from their terminals was highly desirable.

The paper first explains the operation of this computerized mode of group communication by leading the reader through an illustrative example. Then an examination of relative effectiveness and costs of various communication modes is presented in order to illustrate under what circumstances the computer mode offers *economic* advantages over other communication modes. In the quantitative analyses of effectiveness and costs all psychological factors, pertinent to any particular form of group communication, are left out in order to limit incorporation of argumentative or subjective factors. In order to examine the psychological advantages of any communication form, the objective of the communication process would first have to be defined explicitly. The model presented will, however, provide the reader with a basis for relating the relative capacity and costs of group communication modes to his own observations on the discussion objectives and the relative merits of various psychological factors.

Example

"PARTY-LINE" operates very much like a conference telephone call. Group members wishing to hold a discussion go to their respective terminals at an agreed upon time and call up the system on the computer. If some are late in joining the conversation, they will receive first a copy of the discussion as it stands. A member of the conference call has only two basic modes of operation: (1) he can be typing in a message and when he signifies (by typing a plus sign) that he is finished, the computer adds his input to the end of the message list, assigns a unique message number to it and signs the author's name; (2) if the conference member is not writing, or waiting, then he is receiving any messages written by others that have been added to the list since he last received messages. This is similar to either talking or listening on a telephone, with the one difference that everyone is asynchronous with the system, i.e., they could all be typing or receiving at the same time or any mix of the two. There are a number of convenience features, for example, a way of retrieving earlier messages, a way to cause the individual terminal to pause for a fixed length of time or until a new message is generated, and a choice of special text editing commands.

The system is perhaps best understood by observing the example discussion that follows. In this example the reader is in effect "looking over the shoulder" of a participant named Nancy Goldstein, who is interacting on the computer terminal. What Nancy Goldstein types in is presented in lower case, and always follows a question mark. Everything in upper case is typed out on the terminal by the computer. The reader will note that after Nancy Goldstein finishes putting in a message, she first receives any new messages completed by other participants (and therefore already assigned sequence numbers) before her own newly entered message appears. The example conversation presented is a slightly edited version of an actual conversation.

```
DO YOU WISH AN EXPLANATION?  no

ARE YOU USING A CRT?[1]  no

WHAT IS YOUR NAME?  Nancy Goldstein

CURRENTLY THE FOLLOWING GROUPS ARE TALKING:

GROUP #1
MURRAY TUROFF     ROD RENNER     RICHARD WILCOX

DO YOU WISH TO:
    JOIN ONE OF THESE GROUPS (1)
    START A NEW GROUP        (2)
    TERMINATE                   (BYE)
CHOICE (1 OR 2 OR BYE)?  1
```

WHICH GROUP NUMBER? 1

#1 FROM MURRAY TUROFF:
THE PURPOSE OF THIS DISCUSSION IS TO SUMMARIZE SOME OF THE COMMENTS WE AND OTHERS HAVE MADE ABOUT THE PARTY-LINE SYSTEM AND ITS UTILITY AND APPLICATION. THIS IS NOT RESTRICTED TO JUST THE APPLICATIONS IN OUR OWN ORGANIZATION.

#2 FROM MURRAY TUROFF:
ROD, YOU HAVE USED THE SYSTEM QUITE A BIT, EVEN FOR TWO PEOPLE COMMUNICATION. WHEN DO YOU PREFER IT TO A TELEPHONE CALL.

THAT'S ALL THAT IS NEW AS OF TIME: 153839
ENTER A COMMENT (UP TO 8 LINES) OR A SPECIAL REQUEST
COMMAND:
L#1? I used the party line to try to train new terminal operators in the
L#2? IRS when we switched over EMISARI. It would be a very good way
L#3? to get together a group of people to educate them,
L#4? make announcements, etc. One possible problem is this:
L#5? someone has to get onto party line 15-30 minutes ahead of the others
L#6? to put in the main points to be covered, the explanations necessary,
L#7? etc. otherwise, the discussion gets away from the leader, and it
L#8? can never be retrieved.
THAT'S ALL THE LINES ALLOWED AT ONE TIME.
+ ASSUMED NOW.

#3 FROM MURRAY TUROFF:
TWO OBVIOUS POINTS MADE ABOUT THE PARTY-LINE HAVE BEEN THAT NO ONE GETS MISQUOTED AND EVERYONE WALKS AWAY WITH A WRITTEN RECORD OF WHAT TOOK PLACE.

MORE PEOPLE HAVE NOW JOINED IN:
TOM HALL

#4 FROM ROD RENNER:
MURRAY, I'VE CARRIED ON APPROXIMATELY 10 PARTY-LINE DISCUSSIONS–MOST RECENTLY WITH C. B. ATKINSON IN ATLANTA, GA. BEN WANTED TO SUGGEST CHANGES TO PART OF OUR EMISARI SYSTEM AND WE BOTH FELT COMMUNICATION WOULD BE BETTER IN WRITTEN FORM, SO WE AGREED ON A MUTUAL TIME FOR A PARTY-LINE. WE RESOLVED OUR DIFFERENCES AND EXCHANGED USEFUL IDEAS IN ABOUT AN HOURS WORTH OF TIME. WE HAD A "COPY" OF THE ENTIRE DISCUSSION. (TOTAL CPU^2 TIME=3 SECS.)

#5 FROM MURRAY TUROFF:
NANCY, YOU HAD AT LEAST ONE UNIQUE IDEA ON APPLICATIONS OF THE SYSTEM. I'D LIKE YOU TO BRING THAT UP AGAIN.

#6 FROM RICHARD WILCOX:
FROM AN OPERATIONAL STANDPOINT, PROBABLY THE MOST OBVIOUS UTILITY OF "PARTY-LINE" IS THE ABILITY TO CONDUCT A RELATIVELY INFORMAL DISCUSSION AMONG GEOGRAPHICALLY SEPARATED PEOPLE IN ORDER TO ATTACK (AND, HOPEFULLY, SOLVE) PROBLEMS OF MUTUAL CONCERN; THE ADVANTAGE OVER A TELEPHONE CONFERENCE CALL BEING, OBVIOUSLY, THAT A RECORD OF ALL COMMENTS EXISTS BOTH FOR IMMEDIATE REFERENCE (IN CONFERRING ON COMPLEX MATTERS EFFICIENTLY) AND FOR PERMANENT REFERENCE LATER.

#7 FROM NANCY GOLDSTEIN:
I USED THE PARTY LINE TO TRY TO TRAIN NEW TERMINAL OPERATORS IN THE IRS WHEN WE SWITCHED OVER EMISARI. IT WOULD BE A VERY GOOD WAY TO GET TOGETHER A GROUP OF PEOPLE TO EDUCATE THEM, MAKE ANNOUNCEMENTS, ETC. ONE POSSIBLE PROBLEM IS THIS: SOMEONE HAS TO GET ONTO PARTY LINE 15-30 MINUTES AHEAD OF THE OTHERS TO PUT IN THE MAIN POINTS TO BE COVERED, THE EXPLANATIONS NECESSARY, ETC. OTHERWISE, THE DISCUSSION GETS AWAY FROM THE LEADER, AND IT CAN NEVER BE RETRIEVED.

THAT'S ALL THAT IS NEW AS OF TIME: 154339
ENTER A COMMENT (UP TO 8 LINES) OR A SPECIAL REQUEST
COMMAND:
L#1? If you mean by a unique idea my "blind dating" use of the
L#2? party line, there it is. In some instances it seems easier
L#3? to meet people via the party line than face-to-face. It
L#4? is also easy to meet the computer through the party line.
L#5? +

#8 FROM TOM HALL:
I PERSONALLY ALMOST NEVER PREFER A PARTY-LINE LIKE THIS TO A 2 PERSON DISCUSSION, BUT I CAN'T DRAW A DIVIDING LINE TO SAY HOW MANY PEOPLE MUST BE INVOLVED BEFORE THIS SYSTEM BECOMES PREFERABLE. CERTAINLY FOR MORE THAN 4 PEOPLE, THE NON-INTERRUPTABLE, SYNCHRONIZING FEATURES OF THE SYSTEM ARE PRETTY MUCH OF A REQUIREMENT.

#9 FROM NANCY GOLDSTEIN:
IF YOU MEAN BY A UNIQUE IDEA MY "BLIND DATING" USE OF THE PARTY LINE, THERE IT IS. IN SOME INSTANCES IT SEEMS EASIER TO MEET PEOPLE VIA THE PARTY LINE THAN FACE-TO-FACE. IT IS ALSO EASY TO MEET THE COMPUTER THROUGH THE PARTY LINE.

#10 FROM MURRAY TUROFF:
ALTHOUGH WE HAVEN'T USED IT MUCH AS YET, THE ABILITY FOR EVERYONE TO SIGN IN WITH A FAKE NAME ALLOWS FOR A DELPHI DISCUSSION TO TAKE PLACE. THIS FEATURE COULD BE QUITE USEFUL WHEN SOMEONE DESIRES AN UNINHIBITED EXPLORATION OF A TOUCHY ISSUE.

#11 FROM TOM HALL:
NOT ONLY DOES IT PROVIDE A RECORD OF THE COMMENTS TO LOOK BACK AT IN CASE OF UNCERTAINTY, BUT IT EVEN FACILITATES THIS "RECALL" OPERATION WITH THE +N TO M FEATURE FOR RETRIEVING EARLIER COMMENTS. ALSO HAVING THE COMMENTS NUMBERED ALLOWS ONE TO REFER BACK TO THEM EXPLICITLY RATHER THAN USING CIRCUMLOCUTIONS LIKE, "REGARDING WHAT YOU SAID ABOUT SO-AND-SO ABOUT 5 MINUTES AGO. . ."

THAT'S ALL THAT IS NEW AS OF TIME: 154952
ENTER A COMMENT (UP TO 8 LINES) OR A SPECIAL REQUEST
COMMAND:
L#1? +P

```
#12 FROM ROD RENNER:
I PROMISED MY WIFE I'D BE HOME ON TIME FOR A
CHANGE-BYE ALL.

#13 FROM MURRAY TUROFF:
TOM, COULD YOU SUMMARIZE IN ONE COMMENT THE
TECHNOLOGY ASSOCIATED WITH THIS SYSTEM FOR
NON-COMPUTER TYPES.

SIGN OFF BY:  ROD RENNER

#14 FROM RICHARD WILCOX:
WHEN THE GROUP IS LARGER, THERE IS UTILITY IN
HAVING A PROTOCOL WHERE CERTAIN DISCUSSANTS
MAY BE RESPONSIBLE FOR RESPONDING TO CERTAIN
ISSUES. THIS TAKES FULL ADVANTAGE OF THE FACT
THAT EVERYONE CAN, IN PRINCIPLE, BY TYPING AT
THE SAME TIME.

#15 FROM TOM HALL:
BASICALLY ALL THAT IS INVOLVED IS THAT THE
VARIOUS PEOPLE ON THE PARTY-LINE ARE ALL
WRITING THEIR COMMENTS INTO A COMMON AREA OF
STORAGE, WHILE THE PROGRAM TAKES THE
RESPONSIBILITY FOR KEEPING THEIR DATA IN ORDER
BY THE USE OF "LOCKING" AND "UNLOCKING" OF A
CERTAIN LIST OF DATA. THEN EACH USER'S PROGRAM
READS THIS COMMON AREA TO SEE WHAT EVERYONE
ELSE HAS WRITTEN AND PASSES THAT INFORMATION
ON TO THE USERS.

THAT'S ALL THAT IS NEW AS OF TIME:  155346
ENTER A COMMENT (UP TO 8 LINES) OR A SPECIAL
REQUEST
COMMAND:
L#1?  +P

#16 FROM MURRAY TUROFF:
ANOTHER POSSIBLE APPLICATION WOULD BE THE USE
OF THIS COMMUNICATION FORM TO CONDUCT A
SENSITIVITY SESSION.

THAT'S ALL THAT IS NEW AS OF TIME:  155302
ENTER A COMMENT (UP TO 8 LINES) OR A SPECIAL
REQUEST
COMMAND:
L#1?  I've seen people who were nervous about using the
      computer
L#2?  completely forget their fears after a short time on the
      party-
L#3?  line. If you want to capitalize on anonymity to get
      a sensitivity
L#4?  discussion started, maybe, but it would be difficult if
      the
L#5?  participants cannot type.

#17 FROM TOM HALL:
MURRAY, RE: #16. WE ARE OF THE OPINION THAT THIS
SYSTEM IS THE ANTITHESIS OF A SENSITIVITY SESSION,
BEING DRAINED OF ALL POSSIBLE NONVERBAL AND
EMOTIONAL FACTORS.

#18 FROM NANCY GOLDSTEIN:
I'VE SEEN PEOPLE WHO WERE NERVOUS ABOUT USING
THE COMPUTER COMPLETELY FORGET THEIR FEARS
AFTER A SHORT TIME ON THE PARTY-LINE. IF YOU
WANT TO CAPITALIZE ON ANONYMITY TO GET A
SENSITIVITY DISCUSSION STARTED, MAYBE, BUT IT
WOULD BE DIFFICULT IF THE PARTICIPANTS CANNOT
TYPE.

#19 FROM MURRAY TUROFF:
THIS FORM OF COMMUNICATION, AS A STARTING
MODE, FOR A SENSITIVITY GROUP MIGHT SHORTEN
THE LONG PERIOD USUALLY NEEDED TO START UP
EFFECTIVE COMMUNICATION. I THINK THE ABOVE IS
SUFFICIENT AS A DEMONSTRATION, THANK YOU AND
GOODNIGHT.
```

When everyone has signed off, the file of messages that has been stored in the computer disappears and only the individual terminal printouts remain as records. It is in this property that the "DISCUSSION" system differs from PARTY-LINE. Where PARTY-LINE requires the group to coincide in time, the DISCUSSION version does not. The DISCUSSION file stays in the computer until the moderator (the one who started the particular discussion) decides to delete it. The members of the discussion may get on the terminal whenever they wish to observe what comments have been added and to enter additional comments. The DISCUSSION system may then be used to discuss a topic over days or weeks. The moderator of the discussion also has the ability to shape the list of comments into a more compact set by either editing or deleting items. With code names or pseudonyms it may also be used for a Delphi type discussion. The DISCUSSION system is, in essence, a nonvoting version of the DELPHI CONFERENCE system that has also been implemented at the Office of Emergency Preparedness.

This latter system allows voting scales to be associated with comments so that the group may vote on various issues. The computer tallies and displays to the conference group the voting results on a specific issue. An individual may go back and change his vote at any time if he has been so influenced by the discussion. It has been observed that vote change rates tend to be much higher when using the anonymous format provided by the computer.

The seemingly straightforward concept of automating the conference call on a modern time-shared computer system offers a unique ability to allow effective communication within larger groups than would normally be possible in a telephone conference call. While the current version is arbitrarily set at a limit of 15 individuals per conference, it is feasible to include 30 to 50 people in such a conversation. The problem that may exist for those who do not type may be overcome by utilizing secretaries in a dictation mode. The system has mainly been used on terminals operating at 30 characters per second. This is far more desirable than the normal teletype speed of 10 characters per second, since it is closer to reading speed for most people, as is shown in the next section. A group of 10 to 15 people on 30 character per second terminals can exchange a great deal more information in a given time span than would be possible in a verbal discussion. The additional psychological advantage of allowing each conferee to interact with the group at his own speed is also a significant factor in fostering an effective exchange of information.

Effectiveness

It is a fairly straightforward analysis, and quite illustrative, to compare the effectiveness of a computerized conference call versus a verbal committee meeting or telephone conference call.

In a verbal exchange, we assume only one person is talking at a time and that there are no gaps in the dialogue. We define:

W_v = number of words conveyed over the time span of the verbal exchange.

T_v = the time span of the verbal conference.

R_v = the average rate of talking; e.g., words per second.

The variables are related by the expression

$$R_v = \frac{W_v}{T_v} \tag{1}$$

Assuming all participants listen when not talking, then the upper limit on the amount of information[3] that the conference members can exchange, represented by (W_v), is obviously independent of the size of the group. However, as the number of people increases, the proportion of total information which the average member can contribute will decrease accordingly.

For the computerized conference call we define:

W_c = number of words conveyed over the time span of the computerized exchange.

T_c = the time span of the computerized conference.

R_r = the average reading rate or terminal print rate, whichever is lower; e.g., words per second.

R_t = the average typing rate; e.g., words per second.

t_t = the average time a conference member spends in typing.

t_r = the average time a conference member spends in reading.

N = number of conference members.

The total number of words contributed to the discussion is:

$$N\ R_t\ t_t$$

It should be noted that a "conference" for the explicit purpose of having one person give directions, with only minor feedback from the rest of the group, is not the type of discussion being considered here. Such "orders" or "lectures" can be communicated ahead of time by mail, TWX, or computer, and pre-reviewed by the concerned individuals. Any subsequent discussion, however, used to reach an understanding or common interpretation of the "orders" would be appropriately represented by this analysis.

The number of words received by everyone in the discussion is:

$$R_r\ t_r$$

In order for the information to be conveyed, we must have

$$W_c = N\ R_t\ t_t = R_r\ t_r \tag{2}$$

i.e., whatever is typed in by each individual must be read by every individual. However, we also have the condition that[4]

$$T_c = t_t + t_r \tag{3}$$

Equations (2) and (3) are used to eliminate t_t and t_r, giving

$$\frac{W_c}{T_c} = \frac{R_r}{1 + \frac{R_r}{N\ R_t}} \tag{4}$$

The ratio of equations (4) and (1) may then be used as a relative effectiveness measure E(N) of the computerized versus the verbal exchange given N participants:

$$E(N) = \frac{W_c\ T_v}{W_v\ T_c} = \frac{R_r}{R_v\left(1 + \frac{R_r}{N\ R_t}\right)} \tag{5}$$

If a conference group has a given amount of information to exchange (i.e., $W_c = W_v$ = constant), then E(N) is the ratio of the times they will have to spend in making their exchange (i.e., $T_v/\ T_c$). On the other hand if the group has a fixed time in which it can engage in the exchange (i.e., $T_v = T_c$ = constant), then E(N) is the relative amount of information that can be conveyed (i.e., W_c/W_v). The quantity E(N) for some representative cases is plotted in Figure 1 as a function of the number of conferees.

A fast talking rate in a conference can be considered to be two words per second. The reader may verify this by seeing how fast he would read an item to a group without thinking about the content of what he is saying. A slow talking rate, and probably more typical of group discussion, is about one word per second. A good stenographer can take dictation at about one-and-a-half words per second, which we will consider to be an average talking rate. Fast typing can be considered to be about 0.2 words per second (i.e., one character per second, or "two finger" typing). More representative of most young professionals or casual typing is half a word per second, which we will consider as an average rate. The PARTY-LINE system has been used, for the most part, on 30 characters per second terminals (i.e., six words per second), and most users seem to have been able to read at this speed, so this value for R_r is assumed in the curves plotted. Usual computer terminal speeds of two words per second considerably reduce the range over which the computerized system is more effective.

Note that as N goes to infinity E approaches the limit

$$\lim_{N \to \infty} E(N) = \frac{R_r}{R_v} \tag{6}$$

that is, the maximum relative effectiveness obtainable is the ratio of reading to talking speed.

The measure of effectiveness used here shows only the relative "capacity" of these two forms of communication. The more significant question of how much information actually is transferred from individual to individual is considerably more difficult to answer quantitatively, because it is dependent upon various psychological factors determined by the composition of a particular conference group and the objective of its discussion.

However, many of the obvious factors that have been left out of the analyses here tend to work in favor of the computer system. For example, nothing in the calculations reflects any added benefits from the immediate availability of hard copy. Clearly, individual conferees may review during a discussion explicitly what was said earlier, and they may keep transcripts for later review. Also, conferees already familiar with some of the material being communicated do not have to read it, or give the appearance of listening to material containing no information for them. Furthermore, if individuals do not read

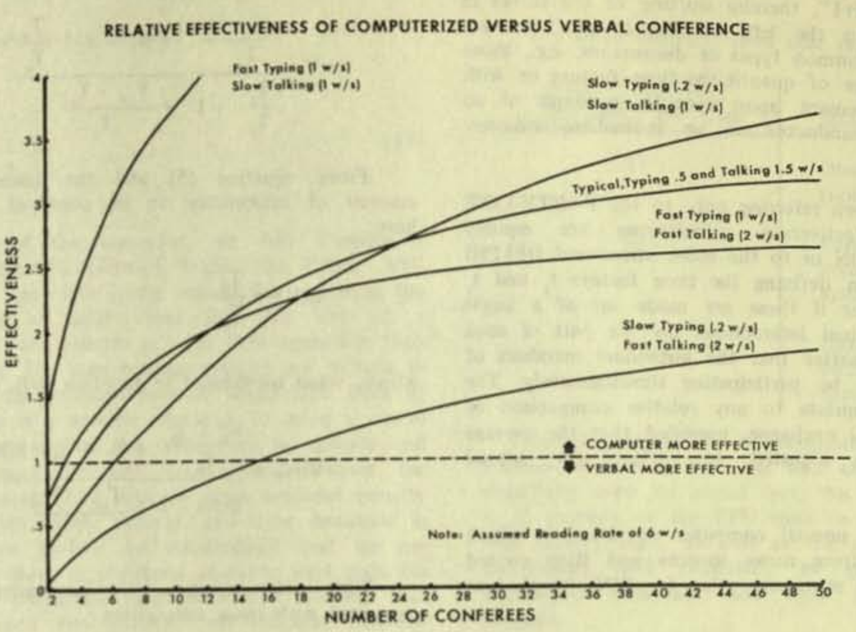


Figure 1

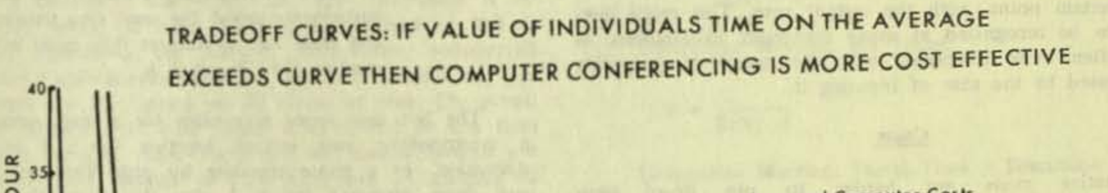


Figure 2

their own comments in the computerized system, the "N" in equation (5) becomes "N+1", thereby shifting all the curves in the figure one person to the left. Finally, it has been our experience that certain common types of discussions, e.g., those dealing with the exchange of quantitative-type factors or with the editing of and agreement upon specific wordings of an item, are more easily conducted on an immediate-hard-copy basis.

Although we have been referring only to the PARTY-LINE system, the same effectiveness calculations are equally appropriate to DISCUSSION or to the more structured DELPHI CONFERENCE system. In defining the time factors t_t and t_r it really does not matter if these are made up of a larger number of separate terminal interactions on the part of each individual, nor does it matter that the individual members of the conference may not be participating simultaneously. The analysis is in fact appropriate to any relative comparison of computerized versus verbal exchange, provided that the average R_v, R_r, and R_t are adjusted to reflect any special circumstances.[5]

As an example, for normal computer operation in which information is collected from many sources and then poured out on a fast printer, the maximum value for E(N) (reached at infinite print speed) is

$$\lim_{R_r \to \infty} E(N) = \frac{N\ R_t}{R_v} \tag{7}$$

which is independent of printer speed. This suggests that the relative efficiency of the computer as a data collection device is a function only of the data sources and has little to do, beyond a certain point, with the output rate. This result does not seem to be recognized at many computer installations, as they are often more concerned with the speed of outputing data as opposed to the ease of inputing it.

Costs

Comparative costs are harder to pin down than comparative effectiveness. Our approach will be to take some representative values of the parameters which are conservative with respect to making the computerized system look good.

Costs for a verbal conference (face-to-face or by telephone) may be represented by

$$C_v = N\ (a_0 + a_1\ T_v) + N\ V\ (T_v + T_{sv}) \tag{8}$$

As before, N is the number of people in the conference and T_v is its duration. The parameters a_0 and a_1 represent the conference telephone cost, over and above any long distance charges per se (long distance line charges for a conference group would normally be at least as high for a verbal as for a computerized conference). The last term accounts for the hourly value (V) of each participant's time, including the extra time (T_{sv}) the participants have to spend in getting to the conference.

For the computerized conference there is a similar cost expression

$$C_c = N\ (b_0 + b_1\ T_c) + N\ V\ (T_c + T_{sc}) \tag{9}$$

To compare the costs, we ask the following question: How valuable must the time of the members of the conference group be for the cost of the computerized conference to equal that of the face-to-face or telephone conference, assuming that the group has a fixed amount of information to convey?

To answer this, we equate equation (8) and (9) to obtain

$$V = \frac{\dfrac{b_0 - a_0}{T_c} + b_1 - a_1 \quad \dfrac{T_v}{T_c}}{\dfrac{T_v}{T_c} - 1 + \dfrac{T_{sv} - T_{sc}}{T_c}} \tag{10}$$

From equation (5) and the assumption of an equal amount of information to be conveyed (i.e., $W_v = W_t$) we have:

$$E(N) = \frac{T_v}{T_c} \tag{11}$$

which, when substituted in equation (10), gives the result.

$$V(N) = \frac{\dfrac{b_0 - a_0}{T_c} + b_1 - a_1 \quad E(N)}{E(N) - 1 + \dfrac{T_{sv} - T_{sc}}{T_c}} \tag{12}$$

With this equation the computerized conference is compared with three alternatives.

(1) A conference telephone call.

(2) A face-to-face committee meeting with no time wasted in travel (i.e., everyone in the same building).

(3) A face-to-face meeting where, on the average, participants spend the same time traveling to and from the meeting as they spend in the meeting (i.e., $T_{sv} = T_v$).

The last case seems reasonable for a group spread around a metropolitan area getting together for a 1 to 2 hour discussion, or a group traveling by plane (including times to and from airports) for a 1 day discussion. Note that no additional factors are included for actual travel costs such as taxies or plane fares.

The detailed assumptions used to obtain equations for each one are as follows:

The cost of a conference telephone call within the Washington, D. C. area was found to be adequately approximated by assuming a_0 to be zero and taking a_1 somewhere between $2.50 and $3.00 per person per hour.[6] We will assume the $2.50 rate, and also assume that the start-up times are the same (i.e, $T_{sv} = T_{sc}$). This local use case is the worse comparison for the computerized conference, since the introduction of long distance charges on both systems should favor the computer over the telephone because of the reduced time to accomplish the same flow of information.

For the face-to-face committee meeting situation, both a_0 and a_1 are of course zero. We assume in the "no travel" meeting case that the start-up time is the same as the computer case (i.e., $T_{sc} = T_{sv}$). In the travel case, we assume that $T_{sv} = T_v$ and that the computer conference start-up time is 10 percent of the computer conference time span (i.e., $T_{sc}/T_c = 0.1$).

For E(N) we assume the typical values of

R_r = 6 words per second

1

R_t = 0.5 words per second

R_v = 1.5 words per second

which results in

$$E(N) = \frac{4}{1 + \frac{12}{N}} \tag{13}$$

For the cost of the computer, we will consider a UNIVAC 1108 computer operating under the EXEC VIII time-sharing system, since this is the vehicle for operating the conferencing systems at OEP. Note that the use of a mini-computer might well provide a lower cost operation than the one used below. The conferencing systems are written in XBASIC[7] language; the program contains about 350 lines of XBASIC code. There is a notable advantage to using a macro language of this sort, so that the system can be quickly and easily modified to meet new user requirements. However, the use of a macro language does produce some overhead penalty in the actual operation of the systems, and those described in this paper have now become so standardized that we are considering rewriting them in re-entrant assembly level code for operation under a transaction processor. The costing that follows is based upon the current, less efficient XBASIC operation.

If a participant in the conference is receiving messages at six words per second without pause, then the maximum amount of CPU (Central Processing Units) time he can use in 1 hour is between 8 and 9 seconds. If he is doing nothing but typing in messages, than the maximum CPU time consumed in 1 hour is between 7 and 8 seconds. The core usage is 50 blocks (512 words) of core. While about a third of this is re-entrant, representing the XBASIC subroutines, commercial time-sharing charges usually make no allowance for this and per user charges must be figured on 50 blocks of core. The actual time the program is in core (using core) should be less than three times the CPU time. Therefore, we assume overall per user demand as 8 seconds of CPU time and 24 seconds of usage of 50 blocks of core. Typical commercial rates for the 1108 are:

$0.25 per CPU second.

$0.0025 per core block per second.

Thus the basic cost is about $2.00 of CPU time and $3.00 of core time per user per hour. Other charges, such as for I/O transfers, should amount to less than $2.00 per user, per hour based upon the same "no pause" assumptions about user interaction.

Therefore, the *maximum* processing charge per user is $7.00 per hour. In addition to this processing charge, we must consider the terminal hookup charge. There are three cases to be considered:

The user who does not own a computer and is an infrequent time-sharing user would pay about $10.00 per hour per terminal for prime time hookups.

The user who has a large volume of use and can usually obtain the non-prime time rate, based upon a minimum use contract, would pay about $5.00 per hour per terminal.

The user who has his own computer with lines and terminals into his system not saturated by his user community would get a realistic cost (on a marginal basis) for the operation of this system by taking zero as the terminal hookup charge.

The three resulting total cost rates for b_1 are thus:

User Charge (b_1) Per Hour Per User	Situation
$7	Owned computer[8] (not saturated).
$12	Large volume user on commercial time-sharing system.
$17	Infrequent user on commercial time-sharing system.

Commercial sellers of computer time usually factor the start-up rates into their per unit of time costs, so that b_0 is essentially zero. In actual fact, this is represented by about 1 to 2 seconds of the CPU time to initiate each user program. This facility cost, as well as the cost of core use on a per user basis, will essentially be eliminated by going to a transaction processor and fully re-entrant version of the program.

Under the above assumptions we have the following three equations resulting from equation (12):

Conference Telephone Call

$$V = \frac{b_1 - 2.50\ E(N)}{E(N) - 1} \tag{14}$$

Committee Meeting, No Travel

$$V = \frac{b_1}{E(N) - 1} \tag{15}$$

Committee Meeting, Travel Time = Discussion Time

$$V = \frac{b_1}{2E(N) - 1.1} \tag{16}$$

These three equations for the "break-even" value of participant's time are plotted in Figures 2-4 as functions of N and for b_1 equal $7, $12, and $17 per hour per person. The curves are interpreted as follows: For a group of size N, the computerized conference can be considered to be cheaper than the verbal conference when the average dollar value of the individuals involved lies above the curve in question. In essence, the curves represent indifference levels at which the two communication modes are equal in cost for a given value and number of conferees.

As before, there is no consideration in this analysis of the added benefits of hard copy or any other psychological factors.

It is quite interesting to note the behavior of these particular tradeoff curves. For example, at values of $10 and $20 per person per hour, the curves indicate the following conference sizes beyond which computer conferencing becomes cheaper than the compared verbal mode.

For the most expensive situation ($17 per hour per user), we have contrasted also the results of verbal discussion rates of 1.5 and 1.0 words per second, in order to illustrate the strong effect of this parameter. The more expensive computer cost with the slower talking rate is equivalent to the less expensive computer cost with the somewhat faster talking rate. One

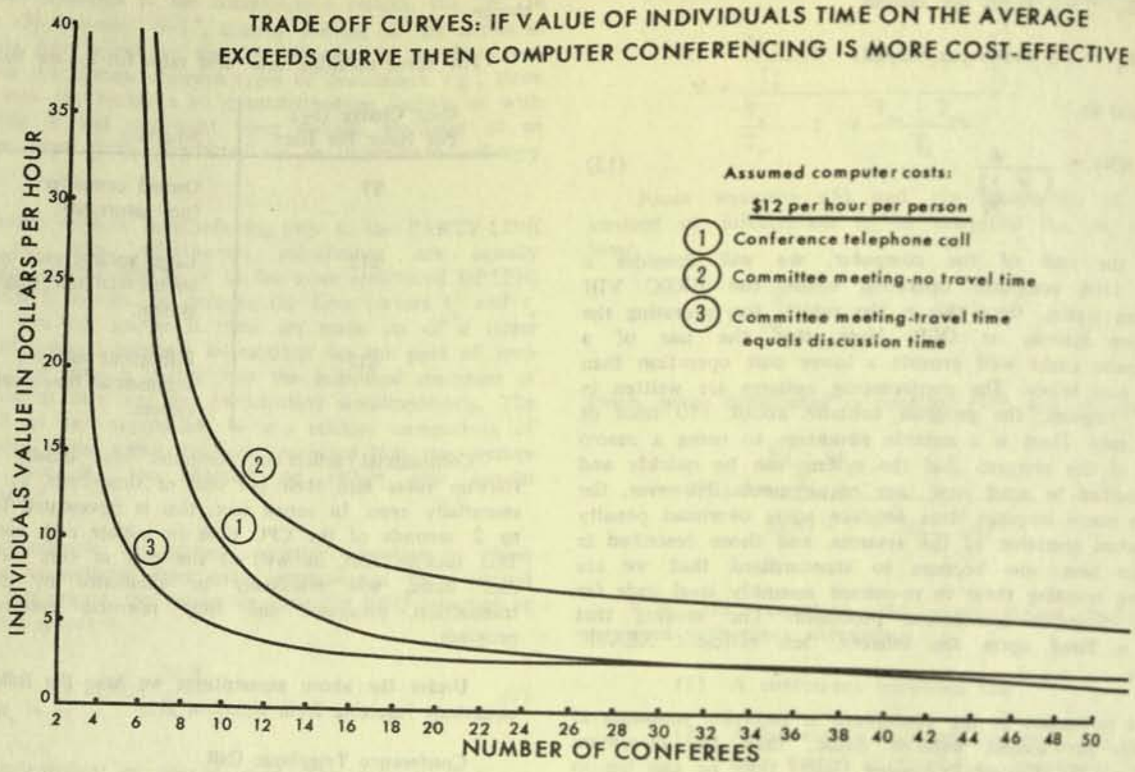


Figure 3

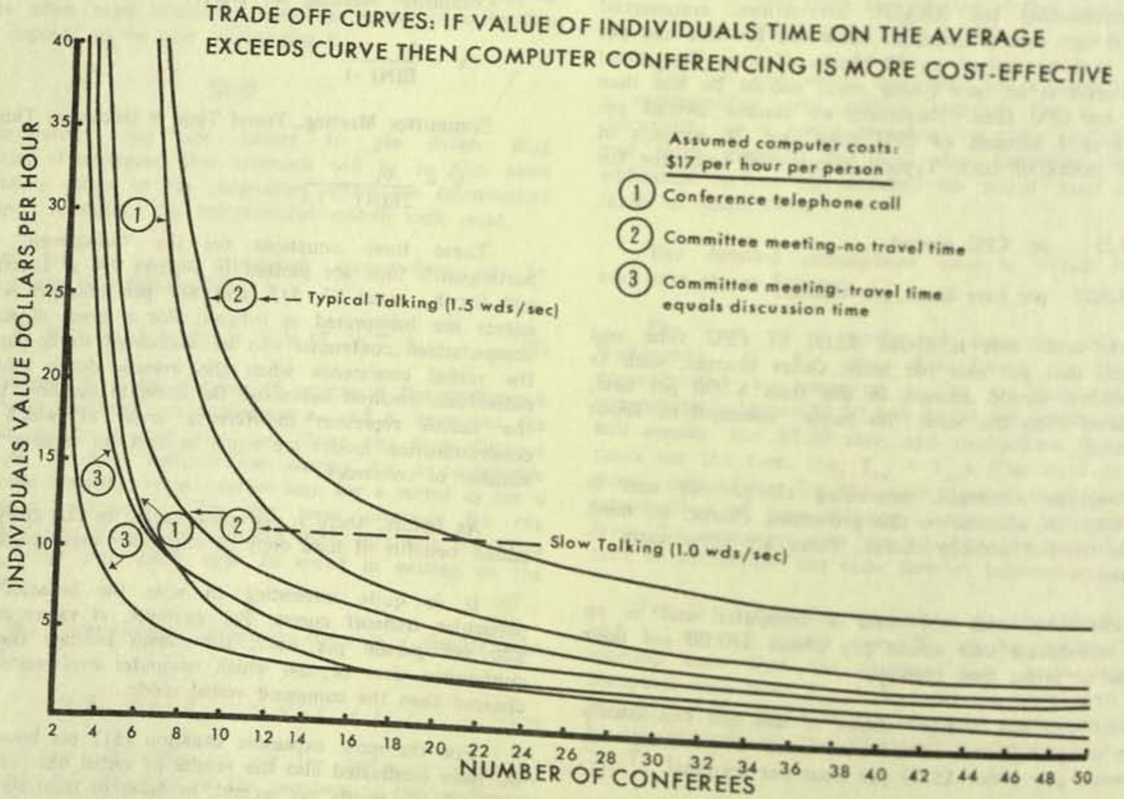


Figure 4

Tradeoff Points Assuming the Value of a Participant's Time to be $10 per Hour

Computer Costs $/hour/person	$7	$12	$17	$17
R_v (talking rate words per second)	1.5	1.5	1.5	1.0
Conference Telephone Call	7	10	15	7
Committee Meeting (No Travel Time)	9	15	25	10
Committee Meeting (Travel Time = Discussion Time)	4	5	7	4

Tradeoff Points Assuming the Value of a Participant's Time to be $20 per Hour

Computer Costs $/hour/person	$7	$12	$17	$17
R_v (talking rate words per second)	1.5	1.5	1.5	1.0
Conference Telephone Call	5	7	9	5
Committee Meeting (No Travel Time)	6	8	11	6
Committee Meeting (Travel Time = Discussion Time)	3	4	4	3

would expect that more complex discussions, involving editing and/or analyses of quantitative type data, would certainly proceed no faster than the one word per second rate if accomplished on a verbal bases. Variation of participant travel time also provides quite dramatic cost-effective tradeoffs at rather small group sizes, even without incorporating actual travel expenses.

It appears that an organization that wished to create a conferencing capability optimized to serve a relatively large volume of use could, by dedicating a computer such as one of the new mini's, provide such services at rates of $2 to $5 per hour per user, as opposed to the $7 to $17 figures used here. Actual experience in typical computerized conferences indicates that the CPU time and resulting processing cost per user are about half of the saturation figure of 8 seconds per hour used in the calculations. This is because users spend time thinking and reviewing while on the terminal. The commercial penalty for thinking which is applied to users by commercial rates is contrary to what some experiments have shown to be the most effective use of time sharing[9]; it is a bone of contention that this author has previously discussed.[10] Considerations of this sort should more properly influence the design of input/output sections of time-sharing computer executive systems, so as limit computer facility use during user pause situations.

For large group discussions (i.e., 25 or more people) which are typical of most Delphi exercises being carried out today, a value of $10 per hour per person always makes it cheaper to computerize than to attempt verbal discussions--in most circumstances a value of $5 per hour per person is sufficient to make a cost-effective tradeoff. Analogous economics illustrate one of the underlying reasons why Delphi exercises utilizing paper and pencil feedback, rather than face-to-face conferences, have become so popular. Although pencil and paper exercises, like computerized discussions, are comparatively cheaper than verbal exchanges, the unnaturally long time delays that result for feedback greatly hamper the communication process.

Finally, it has been the view of many of us who have had the opportunity to participate in discussions via computers that a very different and unique atmosphere of group communication occurs which allows each member to control his communication with the group more effectively. A much more heightened feeling of individual participation appears to be generated, which in turn seems to provide better cooperation in discussing complex issues. The actual quantification of this subjective view by the author is a problem which psychologists will hopefully address when these systems are, at some point, more generally available.

Bibliography

The work that has been described here is a direct follow on to earlier Delphi work by this author, and to the general view of Delphi as the art of tailoring group communication processes to fit particular problem areas. Articles which discuss the relationship of Delphi Design to Management Information Systems are:

Delphi and Its Potential Impact on Information Systems by Murray Turoff, Fall Joint Computer Conference Proceedings, 1971, Vol. 39, AFIPS Press.

An Alternative Approach to Cross Impact Techniques by Murray Turoff, Journal of Technological Forecasting and Social Change, Vol. 3, No. 3, 1972, American Elsevier Publishings.

Delphi Conferencing by Murray Turoff, Journal of Technological Forecasting and Social Change, Vol. 3, No. 2, 1971.

The Design of a Policy Delphi by Murray Turoff, Journal of Technological Forecasting and Social Change, Vol. 2, No. 2, 1970.

The EMISARI system as a whole represents a good example of a Management Information System viewed as a communication process among humans and is documented in the following:

EMISARI: A Management Information System Designed to Aid and Involve People by B. Bechtold, C. Clark, N. Goldstein, D. Marbray, R. Renner, and R. Wynn; submitted to the fourth Intl. Symposium on Computer and Information Science, Dec. 14-16, 1972.

EMISARI–An On-Line Management System in a Dynamic Environment by R. H. Wilcox and R. H. Kupperman; submitted to ICCC 1972.

Considerations in implementing Conferencing and EMISARI type Systems may be found in:

Implementation of an Interactive Conference System by Thomas W. Hall, Proceedings of the 1971 Spring Joint Computer Conference, Vol. 38, AFIPS Press.

Survival of Data for On-Line Data Acquisition Systems by R. Wynn, UNIVAC's Scientific Exchange, Technical Papers of Use Conference, San Francisco, California, March 1972.

References

1. Cathode Ray Tube (Visual Display screen terminal). This option causes automatic pauses in the printout of messages to avoid information moving off the screen before the user has caught up.

2. CPU is central processing unit and represents the amount of computer processing that took place, i.e., 3 secs. in one hours time.

3. Within the context of this paper the term "information" or "amount of information" will denote the capacity to transmit words. By leaving out all psychological factors we are assuming a noise free and non-redundant transmission process where pure capacity is an appropriate effectiveness measure of information flow.

4. The conferencing program uses so little computer resources that there is normally no delay time in which the user has to wait.

5. It is also possible to expand the model to incorporate at least some psychological factors such as think time and to introduce the more formal parameters for information and noise. However, "real" data is not readily available and only parametric tradeoffs could be illustrated.

6. Since the local telephone companies do not make the rate algorithms public, these estimates were obtained by obtaining quotes for a significant number of potential conference calls varying in size and conferee location.

7. A version of the BASIC language developed by Language and Systems Development Corporation, Silver Spring, Maryland which has the capability for designing group conference

structures.

8. The author is aware of two UNIVAC 1108 operations where internal costs for computer use are figured at 13 cents and 30 cents per CPU second, with no other charges. This would place the cost of conferencing at $1.00 or $2.40 per user per hour.

9. For example: Interactive Problem Solving–An Experimental Study of "Lockout" Effects, by B. W. Bochm, Proceedings of 1971 Spring Joint Computer Conference, Volume 39, AFIPS Press.

10. Immediate Access and the User, Datamation, August 1966 and Immediate Access and the User Revisited, Datamation, May 1969.

Appendix

The following is a copy of the one page users guide for OEP's conferencing systems. It illustrates a number of the features made possible by the introduction of the computer: selective retrival of messages by context, author, date of occurrence; the ability to vote on items; the ability to 'whisper' to any member (confidential messages), etc. This newer version of the conference package was developed after the body of the paper was written and represents an effective cost of approximately $5 per hour per user as it was written in re-entrant 1108 machine code. However, it does represent about two man months of programming effort compared to the one day effort on the part of Mr. Thomas Hall for writing the original version in XBASIC for use during the Freeze.

USER'S GUIDE
OEP COMPUTERIZED CONFERENCING (5/1/72)

Party-Line, Discussion, Conference Commands:

+	new messages
+M #	retrieve messages by numbers (+M 2,10,5 TO 8,25)
+M NAME	retrieve messages by author
+? KEY PHRASE	retrieve messages by content
+X	get explanation
+C NAME	send confidential message to NAME
+EC NAME	send previous confidential message to NAME
+C or +EC	kill sending confidential message
+EC AUTHOR	to see what you wrote
+S	status of all members
+S NAME	status of NAME only
+W	wait until someone else does something
+W #	wait # seconds
+W NAME	wait until NAME does something
[illegible]	to sign off (then press CTRL & D keys)

Discussion and Conference Commands:

+EM #	edit messages by number	
+EM -#	delete messages by number	
+D DATE	retrieve messages written or edited on DATE	
	Form: MM/DD/YY	
+D DATE-TIME	retrieve on DATE after TIME	
+D DATE TO DATE	retrieve over range	
+D DATE-TIME TO DATE-TIME	retrieve over range	
+D TIME	retrieve anything edited today after TIME	
	DATE-TIME Form:	MM/DD/YY-HH:MM:SS
	Examples:	+D 1/1/72 TO 1/5/72
		+D 14:30
		+D 1/5/72-13:35

Conference Commands:

+EP	enter proposals
+EP #	edit proposals
+EP -#	delete proposals
+P #	list proposals by #
+V #	vote on proposals
+PV #	combines +P & +V
+VP #	provides just vote results on proposals
+PS #	short summary of voting status
+VS #	your vote summary
+VC CODE	allows user to specify new vote CODE

Note: All proposals are listed if # is omitted.

Calling Statements:

```
@CONFERENCE
@CONFERENCE          conference-name.
@CONFERENCE,options  conference-name.
```

```
Examples:   @CONFERENCE     AGENDA.
            @CONFERENCE,CN  AGENDA.
            @CONFERENCE,CN  EOBA*MEETING/PASS/CODES.
```

Option Letters:

C	automatic pause for CRT
L	list explanation
M	don't echo author's messages
N	short interaction mode

Note: Use '+0' command to turn options on or off

```
Examples:   +0 N
            +0 -M+C
```

Message Editing:

←	character delete (CTRL & 0 keys)
CTRL & X keys	delete the current line
↑	back up one line
=	leave line as is
=N	jump to line N (1 to 9)
RETURN key	carriage return (CR) key blanks the line

Voting Scales:

Desirability (DES)	Feasibility (FEA)	Numeric Code
Very Desirable	Definitely Feasible	1
Desirable	Possibly Feasible	2
Undesirable	Not Determinable	3
Very Undesirable	Possibly Infeasible	4
Not Pertinent	Definitely Infeasible	5
No Judgment	No Judgment	6

Answers to Questions:

Designer - Murray Turoff
Implementor - Rod Renner

COMPUTER-MEDIATED HUMAN COMMUNICATIONS IN AN AIR TRAFFIC CONTROL ENVIRONMENT: A PRELIMINARY DESIGN

M.L. Constant and P.L. Seeley
University of Waterloo
Waterloo, Canada

Summary

This paper describes some communication requirements characteristic of current manually-operated Air Traffic Control systems, and the design of an improved computer-mediated alternative. It focuses on the implications of this alternative system for the human communications network.

At present the controller is not only a manager and decision maker but also a data processor, manipulator and recorder. As well, he is a data transmission device and the organizer of data flow. During the processing of information the controller is in continuous communication with a network of humans (other controllers, pilots, supervisors, coordinators, etc.) over a variety of channels (radio, telephone, radar, etc.). As a consequence he is subjected to a high stress level (due to many conflicting demands on his time), and the system is subjected to constraints often solely dictated by the time associated with human data processing and transmission. To the extent that the controller becomes more a monitor and manager and less a data processor, manipulator and recorder, the stress level will decrease and the system constraints will be relieved.

Increasing aircraft densities, variety and speeds are placing an unacceptable load on the controller, the central figure in the ATC communication system. At present the current flight data required by controllers are recorded, displayed and updated manually on Flight Progress Strips, mounted in holders.

The design of an alternative system requires that pressure on the controller be relieved without sacrificing reliability, safety or the efficient movement of air traffic.

The Present Air Traffic Control Situation

Whenever an aircraft flies between two airfields under Instrument Flight Rules (IFR) conditions, a wealth of data concerning the aircraft is passed from origin to destination and to all the responsible Air Traffic Control (ATC) units along the route of flight. These data are contained in the flight plan which must be filed for any IFR flight, and which may also be required for a flight under Visual Flight Rules (VFR).

Flight plan data are usually moved by verbal, person-to-person communications. To achieve this, a network of direct lines between adjacent ATC jurisdiction is maintained, with specific positions within a unit being obtained by dialling. Within each Centre's area of jurisdiction, further direct lines to airfields and radio beacon stations are maintained.

The Flight Progress Strip

As the flight plan data are passed from jurisdiction to jurisdiction, they are recorded on a horizontal strip of paper, which fits into a holder. This is known as a Flight Progress Strip and records the progress of a single flight. It is marked by lines which divide the surface into squares and rectangles of various sizes.

When initial flight data information is received by one ATC sector of jurisdiction from another, it includes a time estimate for the flight's arrival over an agreed point, known as the "transfer of control" fix. The word "fix" is used as a generic term to include both points defined by radio navigation aids and points defined by longitude and latitude. The receiving controller calculates a time estimate for each further reporting point along the route of flight, based upon the aircraft's true airspeed (TAS), as contained in the flight plan data, and the prevailing wind at the aircraft's altitude. Additional strips are then written for those points which are normally fixes where control decisions may have to be made. The final time estimate within the sector is then passed, together with the remainder of the flight plan data, to the next sector.

The ATC Sector

The ATC sector of jurisdiction has two main communcations interfaces; telephone communications and radio communications. The direct radio communications between pilot and controller represent one control position; with telephone communications being the responsibility of a second controller, who also maintains the progress strip data. Executive control may be exercised by either controller, based upon the available displayed data. On some sectors, radio communications are indirect, through radio operators at HF and CW facilities, or through beacon operators located at Aeradio Stations which contain the navaid facilities. The latter define the reporting points within the sector. In these circumstances, the radio operator reports the substance of all communications to the controller via the telephone circuits, and only one controller will have responsibility within the sector. On other sectors, radar coverage may be available and will be normally used by the controller manning the radio position.

The Handling of Information and Information Categories

On each sector a controller looks at a sloping board divided vertically into bays. Each bay contains a tray into which stripholders fit one above the other. Some stripholders contain a permanent strip bearing the designator of a particular fix. These fix posting strips are arranged on the board to approximate the geographical layout of the area of jurisdiction. All the calculated aircraft strips for each fix are located under the appropriate fix posting. All such groups of strips are then time-ordered on the calculated ETA, with the earliest strip at the bottom. The total display provides the fundamental data base upon which executive control action is predicated, and from which supervisory control can be exercised. At each sector the use made of these data can be categorized into three types, though in practice, clear lines of demarcation are hard to find, and any executive action will be a subtle blend of all three. A thorough study was made.

GENERAL DESIGN CONSIDERATIONS

Whatever display system concept is considered as a replacement for the current flight progress strip, it must not increase the controller's overall workload.

The interface with the computer will involve a measurable workload, which will also be a new workload. The sub-system design should demonstrate a measurable saving in total workload.

The controller's primary function, managing his area of jurisdiction, necessitates that he have time to think. As was made clear in the Montreal TMA Study, it is not possible to measure thinking time, but it is possible to analyse, quantify and then limit all other workload areas, and by such limitation leave the remainder of the controller's time available for his primary function.

At present, in the manual type of system, the procedural controller has a significant percentage of his time occupied in data maintenance, data transfer and verbal coordination functions. The Digital Methods study[1] of Toronto Centre shows as much as 75% of the controller's time occupied under these headings on the one comparatively routine period of traffic which was examined in detail. GPS Sciences, in the Montreal study, found between 50% and 60%, depending on the particular job function, as a maximum that should be planned for or tolerated.

The particular Operational Sequence Diagram (OSD) from which the DML figures were obtained has been examined in detail. From it, a communications model was constructed which reflects the interface workload of the controller during the traffic period analysed. This was the London Low Sector procedural position in Toronto Centre at the time of the DML study, under conditions categorized in the report as moderate to light traffic. See Figure 2, Total Interface Occupancy Times.

It should be borne in mind that there is not necessarily a linear relationship between communications (or interface) workload and the traffic density, due to the complex factor of the human in the loop. The figure of 75% interface occupancy time seems very high, especially when related to the qualitative description of 'moderate to light traffic'; a description which was operationally accurate. Work in this area seems to indicate that various interpretations are possible, including the implication that under pressure, the trained and experienced human being performs a motor task, or a series of interrelated tasks, not as a multiplexer, but rather as a time-sharing device. More motor functions to perform, seems to lead to more efficient sharing of the available time, with non-immediate tasks being postponed.

Clearly all such activities are suitable for, and amenable to automatic processing. If these interface workloads and their physical limitations are virtually eliminated, the only significant motor workload becomes that of the new computer interface, insofar as data maintenance is concerned. The SATCO experience at Amsterdam indicates that if this is the only significant interface, it is an acceptable one; and the Montreal study indicates that if the total measurable workload is kept, by design, at an acceptable level, then the controller can discharge his traffic management function adequately.

The items of information currently displayed on a paper strip have been examined critically to establish realistically what use is actually being made of each item of data, and who really needs it. In terms of new display system concepts, and particularly the EDD, it seems reasonable to conclude that the need for controller acceptance initially will inhibit any radical surgery on the accustomed data items displayed. With use and familiarity, supported by a computer-organized analysis of data manipulation and usage, progressive re-design should be possible to optimize and simplify the display.

The system should be designed to leave data processing, manipulation, storage, retrieval, transmission, etc. as much as possible to the computer and the exercise of judgement to the controller.

Preliminary Thinking

Initially, we accepted that technological evolution would lead Air Traffic Control in the direction of greater automation and greater reliance on the computer. As a result, the controller would become more a monitor and decision-maker, and less a calculator, estimator and data-manipulator.

From the point of view of communications the hierarchy below indicates that the maximum information exchange would come from a dynamic, visual, inter-active, computer-driven display:

1. Manually displayed data.
2. Printed displayed data.
3. Computer-generated printing.
4. Computer-generated symbology, formatting and printing.
5. Interactive, dynamic, computer-generated symbology, formatting and printing.

In view of our belief in ATC's growing dependence on the computer, it followed that our principal design concern should be focused on the nature of the man/machine interface, both devices and language. In addition, we felt that any commitment to a computer-based ATC system would be irrevocable: in the event of failure, reversion to manual methods would be impracticable and every design decision would have to be made subject to this premise. We accepted as a fact that every part of the system would fail at some time. Consequently, the concern initially must be with operational reliability rather than with the reliability of hardware and software. This implies a design for operation under conditions of failure, wherein the traditional system reliability studies are only intended to reduce the frequency of operation in this mode.

The design solution must be capable of ready acceptance and assimilation by the controller. It was also clear that the design solution would have to be implemented in a period of accelerating technological change with its implicit high rates of obsolescence. This in turn led us to the conclusion that the design should be based on the current technology embodied in off-the-shelf hardware of demonstrated reliability.

Constraints

All automatic data processing and data transfer equipment should be self-checking.

All display equipment which shows callsigns, position reports and clearance instructions should be fail-safe.

To satisfy legal requirements and to meet requirements for the investigation of incidents, the automatic ATC equipment should be capable of keeping a record of data displayed by that equipment to the controller.

DESIGN CRITERIA

Minimally, information input to the system and output from the system is to be equivalent in content to that currently in use. In effect, this means retaining the same information capability as exists today with present methods of controlling aircraft. The information sources are to include those planned for the 1970-80 period, as specified at the beginning of the study by the Department of Transport.

Consideration will be given to the determination of the reliability of individual hardware components and the possibility of residual hard copy in the case of catastrophic failure.

Considerations will also be given to the various problems attendant on the implementation of a successful design solution: minimal familarization, start-up, parallel implementation with the old system, etc.

The solution must apply directly to radar, non-radar and mixed environments, including departures and arrivals operations. It should be capable of interfacing with a NAS-type system such as that currently under development in the United States, and, as well, provide functional separation between radar data and FPS data.

A. Hardware Criteria

1. Displays - devices that can display and maintain the probable amount of required information at computer-compatible speeds. They must be reliable as components (proven record) and contribute to the operational reliability of the system through modularity and serviceability.

Display information must be available to other controlling personnel such as coordinators and supervisors.

2. Input Devices - these should be ergonomically acceptable. They should require minimum specialized training and combine a low probability of data input

errors with ease of recovery and correction.

B. Man/Machine Interface Criteria

1. Avoid Information Overload -

(i) Data Overload - a function of the quantity to be scanned to determine the operational status of a sector. This overload can be avoided by identifying high probability data needs requiring continuous display, and separating them from those of low probability, displayed on demand.

(ii) Attention Demand Overload - conflict warning, handoff acceptance and verification, emergency situations, data change notices, etc. The demands made by the computer on the controller can be minimized by arranging a hierarchy of priorities, a variety of cue modes (underlining, intensity levels, flashing, formatting, scale changes, font, type size, symbols, etc.) by restricing the number of simultaneous cues of one kind (e.g. no more than 3 flashing symbols); and by the use of a variety of sensor modes (audio alarm, buzzer, etc.).

(iii) Manual Overload - trading the present manual activity demanded by the FPS for an even more onerous manual activity demanded by the computer input devices; keyboards, function buttons, joysticks, etc. The input devices must be ergonomically acceptable. They should require minimum specialized training and combine a low probability of data input errors with ease of recovery and correction.

2. The controller should not feel that the machine is driving or pacing him; nor should he feel that he is waiting "too long" for the machine to respond.

3. The controller should enter the new man/machine relationship without a sense of cultural/technological shock.

4. The controller's image of himself should not be violated by requiring him to perform operations he might consider as menial or trivial.

5. In general, certain confidence requirements must be met:

(a) The controller must be provided with all the information he feels he must have, with provision for the progressive reduction of information with accumulating experience.

(b) The controller must feel that he has at least the same control facilities as he now enjoys. Again, there should be provision for a progressive change in control facilities according to experience.

(c) The controller must feel that he can rely on the system under conditions of high pressure, catastrophic sub-system failure and intermittent disturbances.

6. The information display should enable the controller to maintain a three-dimensional understanding of his airspace and there should be provision for frequent refreshing under conditions of changing information. This suggests memory-jogging and attention-getting devices such as "cocking", formatting and the use of special symbols.

C. Software Criteria

The hardware/software should be designed to minimize the man-machine response time without jeapordizing the processing capacity of the CPU. Clearly, to the extent that the CPU is overloaded, the response time of the machine to the controller will be increased.

Input commands of purely local significance (domestic chores, etc.), should not be permitted to make demands on a heavily loaded CPU, with all the internal bookkeeping that this would require to cope with processing of a comparatively trivial kind. This requires an interface between the controller's input facilities and the data base in the refresh memory, previously specified. For the sake of the flexibility demanded by the hardware and man-machine criteria (noted above), some form of programmable local processor is required for the display terminal.

For the controller, all this unloading of the CPU would reduce the machine's response time. For the CPU, there would be increased capacity to carry out those sophisticated tasks for which it was originally intended.

As a consequence of the above considerations, the following software design criteria can be applied:

1. The configuration of the display terminal software should be such as to provide for the kind of programming flexibility that would permit specific, custom-tailored programmes for each control centre, each control sector, and each control environment, including even the preferences and idiosyncracies of the individual controller.

2. For ease of manual imputting and for minimization of input errors, the local processor should present the controller with a menu of pre-formatted choices; a hierarchy of amendment functions based on the frequency and/or probability of use.

3. Man-man and man-machine communications can be reduced by increasing machine-machine communications. In order to reduce the number of communications transactions required of the controller over existing channels (intercom, R/T, telephone, audio, non-verbal), these should be allocated to the degree possible to computer data channels. Such a software design will provide for the technological advances expected in the next few years; satellites, data links, positive intermittent control, discretely-address time-squenced radar beacons, etc.

DESIGN AND SIMULATION

It was decided that a model of the optimum design, in the form of a hardware/software simulation, would be compared with the communications model from the analytic phase. The purpose was to elicit the kinds and degree of improvement, if any; the probable effects on the system hardware, software, personnel and procedures; and the effects and requirements unforeseen prior to actual hands-on experience.

Conclusions re Simulation

The simulation has proven to be a powerful tool in the evaluation of the design proposals for the data display. Some effects and requirements which became apparent only after seeing them on the screen were: (Fig. 6)

1. Data should be ordered in justified columns. It then becomes very easy to scan ETAs or altitudes when making control decisions.

2. Strips should not be indented on the planning display as an attention device. It is confusing to see a break in the columns particularly if two or more consecutive strips are indented. A special symbol to the right or left of the strip is much better on this type of display.

3. A difference in character size at each end of the planning strip helps to denote the direction of the aircraft. The convention is to print the aircraft identity at the end of the strip that signifies the direction of flight. In columnar form this loses some of its impact. If the character size is made smaller the planning strip takes on a sense of direction, but there is a significant loss of readability of "IDENTIFICATION". By reducing the character size of "NEXT BEACON" the sense of direction is preserved without affecting the quick visual registration of "IDENTIFICATION".

4. Displaying the ETA at each reporting point on the full strip helps in the planning function. Further refinements in the use of a moving indicator are needed to judge its usefulness.

5.(a) Data packing on the planning display can be accomplished by single row spacing of those aircraft whose ETAs are more than twenty minutes from sector penetration. The second row contains only the SSR and CLEARED and REPORTING altitudes; they are not as relevant for aircraft that far way - if needed, they can be obtained on the full-strip display. For aircraft which are within twenty minutes of sector penetration the new data

appear automatically when there is display room under the appropriate beacon.
(b) A cueing symbol should be displayed with each item of new data until it is acknowledged.
6. Underlining can be an effective visual cue on the planning display. In the simulation, an underlined ETA denotes that it is actually on ATA.
7. "Crossing out" data on the screen is useful to denote data no longer valid but which must be maintained. Dashes or minus signs through each character accomplish this.
8. Conflict indicators are best displayed in the ALTITUDE columns. The controller usually must make an altitude decision, and these indicators draw his attention to the ALTITUDE column. It is easy to scan the aircraft identities once the eye has settled on the conflicting altitudes.
9. Many ideas in menuing for the data display came out of the simulation.
(a) The "readback line" is extremely important. Without it distractions can cause input errors.
(b) Format checking can be done during entry; only after the format has been checked should the ENTER symbol be displayed for execution.
(c) The order in which menu items appear is important since the optimum order for each sector may be different. The menu displays must be versatile enough to allow the reconfiguration of the hierarchywith minor reprogramming.

THE DESIGN SOLUTION

Aspects of the various design proposals which met the basic criteria (described earlier) were subjected to simulation, and in the light of the factors which emerged, appropriate trade-offs were made to arrive at an optimum solution. The design solution takes the form of a terminal consisting of:

A CONSOLE which contains a planning display area (one or more CRT modules), a control display (one CRT module) and a touchwire overlay unit.

A LOCAL PROCESSING UNIT (LPU) comprising a minicomputer and an I/O interface.

Effects on the Controller Task

The effects of computer-organized data maintenance can be far-reaching. Verbal transfer of data and the manual actions involved in organizing, re-organizing, manipulating and performing calculations are greatly reduced since these activities are largely assumed by the computer. In our design solution the results of data manipulation are automatically and immediately displayed at all relevant sectors within the Centre, and are also transmitted to other concerned Centres in an appropriate form.

The net effect on the controller task is to reduce human interaction and verbal communication to a new minimum required for strategic and tactical planning.

For the purpose of quantitative comparison, the communications model previously analysed was compared with a sector aided by computer-based data processing and display. The results obtained follow.

Digital Methods Limited Study; OSD #5; London Low Sector 1615-1633Z

The analysis of the Operational Sequence Diagram resulted in the following interface occupancy percentages:

- Pushbutton Switchboard, 43.15%
- Associated radio/radar controller, 3.52%
- Visual and aural input from radar and speakers, 2.87%
- Other internal voice inputs, 26.11%

This gave a total interface occupancy of 75.66% in flight data maintenance and coordination activity by the procedural controller, in a traffic situation described in the DML study as moderate to light.

This traffic situation was compared with interface occupancy percentages for an automated data processing system, similar to the suggested design solution.

The following assumptions were made:

- A synthetic radar display capable of showing, as minimum information, aircraft identity and altitude.
- Computer-organized manipulation and distribution of data internally.
- Computer/computer data transfer between Cleveland and Toronto.
- Sector Console; as per the suggested design solution, with a minimum capability as demonstrated in the simulation.

On this basis OSD # 5 was re-examined. A large number of actions associated with writing strips, transmitting strip data, moving strips, asking for information, updating data, and coordinating data changes, were eliminated. The only data maintenance necessary in the automated environment was that required for data changes and ATA entry. The assumption of controller/controller verbal communication for strategic planning (the approval request for UA 499 climbing out of Buffalo, for example) was made, since it seemed realistic that this be retained for some time after the phase-in of such a system. Technically, controller/controller communication through the computer could achieve the same end. Whether this would be operationally desirable would need further evaluation.

The comparison produced the results of fig.2.

From this it is clear that the reduction in data maintenance activity is dramatic. The primary factor leading to this greatly reduced figure is the data display terminal; including the fact that all results of data manipulation are automatically displayed, not only at the sector initiating the change but at all other outlets displaying the same data. This eliminates what is, at present, a large percentage of the procedural controller's workload even with the pre-printed strip.

From these figures it is also clear that the procedural controller on the London Low Sector as it then was, aided by the proposed automated data display, could handle an average of several data changes per minute - steadily - without coming close to a 50% interface workload. Traffic would have to be well in excess of the permissible Instantaneous Sector Count for his radio/radar controller before it approached the communications workload recommended by the Montral TMA study. Thus it may be concluded, even from such experimentally limited figures, that the clerical and communications workload of the procedural controller will be significantly reduced. He will be able to take an increasingly active part in the strategic planning of his sector. The longer range consequences might well be re-organization of the "sector" concept.

The primary effect on the controller's data maintenance task will be to shift his workload to the automation interface, and to severely reduce the time involved on all others. Practice, familiarization and an optimized layout will serve to improve performance.

By reducing and simplifying the controller's manual tasks, and eliminating a large amount of verbal communication, the controller's management of his sector is greatly improved.

Effects on System

The effects of such automated data processing and display on the system will be considerable. Sector boundaries can be revised in cases where the data maintenance workload drops drastically. Since this will not affect the limitations set on the radio/radar controller by the Instantaneous Sector Count, several radio/radar controllers may work from a common data display, maintained by one procedural controller. In addition, none of these men need by physically side by side, should this be a useful consideration.

Controller Acceptability of the Design Solution

- The familiar FPS format is retained
- The familiar fix posting format is retained
- In general the format and layout of the visual procedural information is familiar yet concentrated for easier scanning.
- The touchwire input requirements and menu formats minimize mistakes and require no special manual skills.
- Manual handling of data is less than before.
- Voice communications demands are fewer.
- The symbols used are standard in ATC.
- The controllers has almost nothing new to learn or memorize.
- The responsiveness of the machine should present no problems and instill confidence.

The fact that the results of data manipulation can be displayed immediately in other sectors should sharply reduce the amount of verbal coordination within a Centre. This effect is not present in an automated data processing system where the output is paperstrips. It is not suggested that the elimination of all verbal coordination concerned with data maintenance and the organization of traffic sequences would be an immediate result of automated data display, since controller acceptance would probably inhibit such change. It is likely, with the passage of time, that the functions of co-ordinator and supervisor will be re-assessed. Similarly communications and coordination between adjacent sectors of neighbouring Centres will be affected.

Resume and Conclusions

1. The controller's task will change. At present he is not only the manager and decision maker but also the data processor, manipulator and recorder. As well, he is a data transmission device and the organizer of the data flow. As a consequence he is subjected to a high stress level (due to the many conflicting demands on his time) and the system is subjected to constraints often solely dictated by the time factor associated with human data processing and transmission. To the extent that the controller becomes more a monitor and manager and less a data processor, manipulator and recorder, the stress level will decrease and/or the system constraints will be relieved.

2. Because of the new sector to sector communications capability the coordinator's role will change; probably moving up the hierarchy of responsibilities from functions of coordination to those of supervision.

3. The design solution is applicable to any automated ATC environment; radar, procedural, mixed and TMA. The cost/benefits of this design would recommend its use in any of the above.

4. The design solution will permit expansion, contraction or re-grouping of sectors depending on traffic density and/or personnel.

5. The automatic deployment of displayed data throughout the Centre will expedite metering of traffic flow, permitting the gating of traffic up to the maximum runway capacity.

6. We speculate that further development of the concepts embodied in this design will lead to a synthetic dead-reckoning computer-driven display. The computer will compare the updated flight plan tracks with radar data, and display disparities and conflicts; offering a roster of preferred control decisions for the controller's selection. This shift in emphasis fro reliance on procedurally-supported radar control to radar-supported procedural control will accommodate and support the trend toward intermittent positive control, data links and ultimately, machine/machine ATC systems.

7. In exploring the consequences of integrating the LPU into our data terminal (to provide computing autonomy) we concluded that a viable alternative to a CPU could consist of a set of autonomous, yet co-operating LPUs.

8. A further design alternative would incorporate a second touchwire mask over a portion of the "spare" planning display, to which all pushbutton functions would be transferred. Thus the LPU would assume responsibility for finding an operating communications channel to connect the controller to the appropriate sector, centre, R/T channel or other information sources.

9. Our proposal permits the incorporation of a Management Information System. Control decisions are recorded not only on voice tape, but also in the computer memory; subject to retrieval, computer collation, analysis and assessment. The computer can provide information for the analysis of procedures, assessment of obsolence, and identification of trends and directions for change. The MIS is thus a built-in design tool for the ATC system of the future.

GLOSSARY

ATA	Actual Time of Arrival
ATC	Air Traffic Control
CPU	Central Processing Unit
CRT	Cathode Ray Tube
CW	Continuous Wave
DML	Digital Methods Limited
EDD	Electronic Data Display
ETA	Estimated Time of Arrival
FPS	Flight Progress System/strip
HF	High Frequency
IFR	Instrument Flight Rules
I/O	Input/Output
LPU	Local Processing Unit
MIS	Management Information System
NAS	National Airspace System
OSD	Operational Sequence Diagram
R/T	Radio/Telephone
SATCO	Signaal Automatic Traffic Control Organization
SSR	Secondary Surveillance Radar
TAS	True Airspeed
TMA	Terminal Area
VFR	Visual Flight Rules

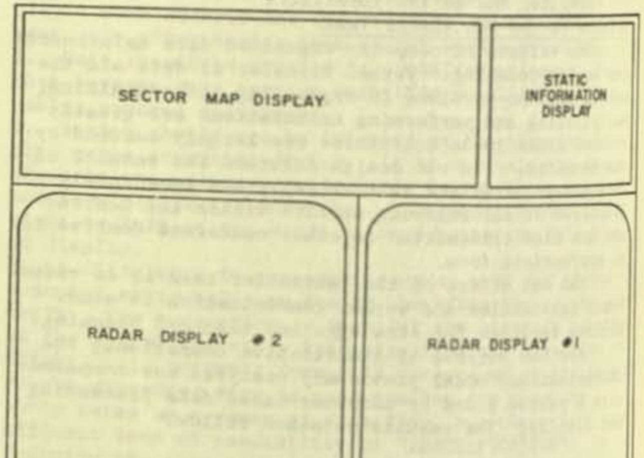
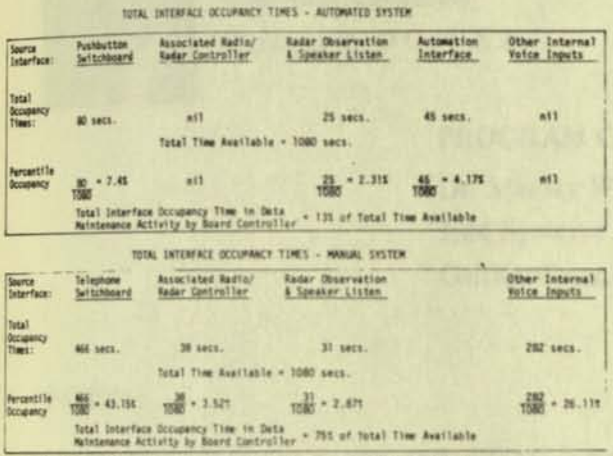


Fig. 1
CONTROLLER'S CONSOLE
(GENERALIZED LAYOUT ONLY)

References

1. Report of a Study: An Automated Display to Replace ATC Flight Progress Strips - For the Canadian Department of Transport. Waterloo Research Institute Project 9022 by Constant and Seeley, University of Waterloo, Waterloo, Ontario.

2. Digital Methods Ltd. Study by G. Phyffe.



TOTAL INTERFACE OCCUPANCY TIMES - AUTOMATED SYSTEM

Source Interface:	Pushbutton Switchboard	Associated Radio/ Radar Controller	Radar Observation & Speaker Listen	Automation Interface	Other Internal Voice Inputs
Total Occupancy Times:	80 secs.	nil	25 secs.	45 secs.	nil
	Total Time Available = 1080 secs.				
Percentile Occupancy	$\frac{80}{1080}$ = 7.4%	nil	$\frac{25}{1080}$ = 2.31%	$\frac{45}{1080}$ = 4.17%	nil

Total Interface Occupancy Time in Data Maintenance Activity by Board Controller = 13% of Total Time Available

TOTAL INTERFACE OCCUPANCY TIMES - MANUAL SYSTEM

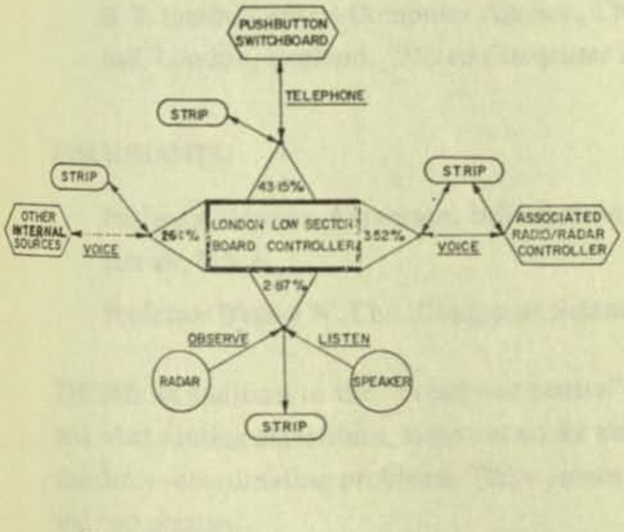
Source Interface:	Telephone Switchboard	Associated Radio/ Radar Controller	Radar Observation & Speaker Listen	Other Internal Voice Inputs
Total Occupancy Times:	466 secs.	38 secs.	31 secs.	282 secs.
	Total Time Available = 1080 secs.			
Percentile Occupancy	$\frac{466}{1080}$ = 43.15%	$\frac{38}{1080}$ = 3.52%	$\frac{31}{1080}$ = 2.87%	$\frac{282}{1080}$ = 26.11%

Total Interface Occupancy Time in Data Maintenance Activity by Board Controller = 75% of Total Time Available

Source Data: Digital Methods Study - OSD#5 - Low London Sector (1615z - 1633z)

COMPARISON OF TOTAL INTERFACE OCCUPANCY TIMES
FIG.2

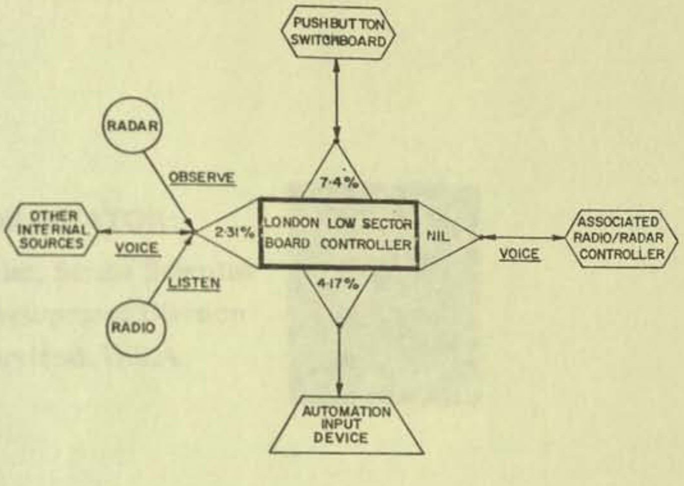
COMMUNICATIONS MODEL
MANUAL SYSTEM 1968



OSD #5 TOTAL OCCUPANCY - 75·65 %

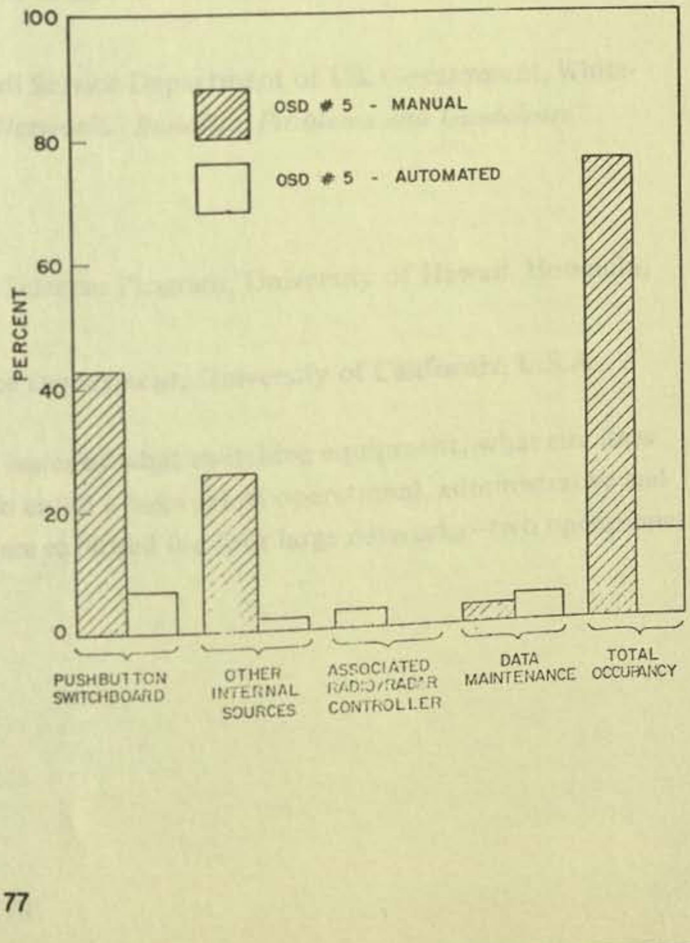
FIG. 3

SIMULATION COMMUNICATIONS MODEL
AUTOMATED DATA PROCESSING AND DISPLAY SYSTEM



OSD #5 TOTAL OCCUPANCY - 13·88 %

FIG. 4.



NETWORKS–RECENT DEVELOPMENTS

CHAIRMAN:
Dr. Jack Moshman
Moshman Associates
Washington, D.C., U.S.A.

PROGRAM COORDINATOR:
Dr. Stanley Winkler, Senior Scientist
IBM Systems Development Division
Gaithersburg, Maryland, U.S.A.

PAPERS:

N. G. Anslow and J. Hanscott, BOAC, London Airport, Hounslow, Middlesex, England, *"Implementation of International Data Exchange Networks"*

A. A. McKenzie, B. P. Cosell, J. M. McQuillan, and M. J. Thrope, Bolt, Beranek & Newman, Inc., Cambridge, Massachusetts, U.S.A., *"The Network Control Center for the ARPA Network"*

D. L. A. Barber, Division of Computer Science, Department of Trade and Industry, National Physical Laboratory, Teddington, Middlesex, England, *"The European Computer Network Project"*

B. T. Smith, Central Computer Agency, Civil Service Department of UK Government, Whitehall, London, England, *"Mixed Computer Networks: Benefits, Problems and Guidelines"*

DISCUSSANTS:

Professor Norman Abramson, Information Sciences Program, University of Hawaii, Honolulu, Hawaii, U.S.A.

Professor Wesley W. Chu, Computer Science Department, University of California, U.S.A.

THEME: In addition to the "bread and butter" issues of what switching equipment, what size lines and what routing algorithms, large networks also entail a large set of operational, administrative and standards–coordinating problems. These issues are explored for four large networks–two operational and two planned.

IMPLEMENTATION OF INTERNATIONAL DATA EXCHANGE NETWORKS

N. G. Anslow and J. Hanscott
BOAC, London Airport,
Hounslow, Middlesex, England

ABSTRACT

The paper reviews some of the problems involved when constructing a data transmission network for use on an international and inter-Continental basis. It discusses the technical difficulties of getting standard voice grade telephone channels suitably conditioned for the transmission of data at medium/high speeds, particularly in the absence of agreed international standards for some parameters which are essential to achieving a satisfactory bit error rate.

The paper also discusses the administrative and control problems of operating a data network, together with the practical difficulties of achieving and maintaining its performance to a consistently high standard. The inter-relationship between medium/high speed data transmission facilities and the conventional low speed telegraph network is noted and some thoughts are offered for future developments.

The implementation of medium speed data circuits on a world-wide basis poses many new problems for the Communications engineer. Most engineers, by their training and experience, are realists, although perhaps perfectionists at heart. This is why I believe one frequently hears that there are no good international data circuits, only some which are just a little bit better than others. Why is this? It is, in my view, because we are asking the networks provided by the P & T Authorities around the world to do a job for which the basic equipment was not designed. In essence, we are using voice grade telephony circuits for the transmission of data which does not have the varying and asymmetric characteristics of the speech waveform and therefore can bring about a different set of circumstances insofar as cross modulation is concerned. The devices that we use to get the data information onto the public network, ie the modems, are sensitive to conditions to which conventional voice telephony is not sensitive. Although a great deal is done to condition conventional carrier type telephony circuits to be able to transmit data in an efficient way, only some of the necessary characteristics are dealt with and not all of them. For example, frequency attenuation distortion and propagation group delay distortion can be confined to within certain recommended limits by the addition of appropriate values of inductance, capacitance and resistance to the circuits by well proved and well known techniques, but such questions as phase jitter, for example, are not so readily dealt with and form at the present time no part of an internationational tariff specification.

2 Our basic problem is that we have to go from a digital form of output from the computer through to an analogue representation of it to transmit it through our telephone systems and then convert it back again to digital at the required terminal. Obviously, life would be far easier if we could go digital all the way. However, this will be a long time coming for international circuits and therefore we must do the best we can to adapt our systems into the equipment which exists at the present time.

3 BOAC's data network radiates from our computer complex HQ in London to the USA, Canada, Scotland, Ireland and the Provinces in the UK, and also to four Continental European capitals. In the UK and North America these circuits employ multidrop techniques and all of them work at this time at 2400 bps. In multidrop working up to four terminal computers may be connected to one line. This means that the data circuits must be equalised not just from end to end but from all terminals to end and to each other. On the Atlantic, however, we use time division multiplexing techniques which permit a data rate of 9600 bps on two trans-Atlantic voice grade circuits. In addition to the foregoing, our computer complex now performs the main electronic switching function for telegraph traffic and on the trans-Atlantic circuits we time division multiplex some sixteen telegraph channels with our medium speed data circuits. The central computer programme is so organized that medium speed traffic originated in, say the US, may well generate a telegraph message going to some more remote location served by our telegraph network. Thus, the conventional global telegraph network operated by BOAC and its pool partners, and known as the OPTS, is now from a functional point of view, integrated with the medium speed network, although from the data transmission aspect at this time the principal integration is on the basis of hardware.

4 Our philosophy so far has been to use cable circuits for the long haul international routes. This decision was taken when the system configuration was being planned because the 'make good' or fall back situation appeared to be better with cable circuits. Also not too much was known in 1966/67 about the performance of satellite circuits as medium speed data channels and the additional propagation time via the satellite was very unwelcome to the overall system planners at that stage. Recently we have used satellites for our purpose - with success. In many ways the circuits are better requiring less equalisation, although at certain times the noise levels are higher. The additional propagation delay is noticeable and somewhat undesirable but the system can manage to operate satisfactorily with a one satellite hop. Almost certainly we will use a satellite

circuit from London to Hong Kong when we expand the service to that area. However, if we wish to further expand from Hong Kong to, say, Bangkok or Singapore, the two hop delay may pose both technical and operational problems.

5 The OFTS network which is administered by BOAC on behalf of all its members, employs a common code, ie the international alphabet No. 2 and most of the circuits operate at a speed of 75 bauds. In association with our member partners, we operate electronic switching systems at London, Hong Kong, Johannesburg and Sydney and, until very recently, in the USA with the co-operation of ARINC. Traffic from any member airline may flow on this network which, in essence, is the network of BOAC and each of its partners all joined together for the common interest and for which agreed cost sharing provisions have been derived. We work under strict limitations in regard to the traffic which is authorised to flow on these circuits; the regulations vary from country to country and some P & T Authorities reserve the right and actually carry out from time to time monitoring of the traffic flowing. The international control of the network is carried out from BOAC's telegraph control centre which, in the event of circuit outage, will organize alternate routings. The OFTS network also interconnects with the SITA network, through which many of the smaller and more remote stations are served. It is therefore possible for airlines to originate telegraph messages and be able to rely on the fact that if it has been correctly addressed and formatted with the agreed start of message and end of message provisions, etc. that such messages will be correctly delivered.

6 Our medium speed network is, however, at this time much more individualistic as far as BOAC is concerned, although we do provide some airlines with whom we have a commercial association with terminal equipment in their offices. Transmission from the central IBM 360 165 complex is made in EBDIC 6 bit code, using a polynomial cyclic check character system for data integrity. The data is transmitted at 2400 bps to the Ferranti Argus terminal computers. Each of these Argus computers is capable of driving up to 72 input/output devices, of which in general the vast majority are CRT type displays with a small number of relatively high speed printers. From the Argus to the CRT terminals or printer terminals a 7 bit code is used which conforms very nearly, but not quite, to the ISO standard. Since the original installation in 1968 further programmable terminals have been added at the same locations where the requirement is small and the installation of an Argus terminal control unit would not be justified.

7 In implementing this network we faced a number of problems. These were the engineering considerations, the operational and control aspects and the question of costs and tariffs. Some unique problems in the planning and technical areas arose, but in general all the problem areas by now are well defined and well known to the various Telecommunications agencies. The geographical areas which provide the major portion of data transmission are Europe and North America. In each case the Telecommunications Authorities have well established transmission systems and operating methods. In North America the Bell System provides total coverage on a Continental basis. In Europe the National Postal and Telecommunications Administrations provide a similar, if more fragmented, system. In neither case can it be said that the national operating practices and regulations are completely compatible with the agreed international CCITT recommendations. The difficulties presented to international users can be ascribed, in part, to the confusion generated by these different practices and to the lack of consideration on the part of some of the operating agencies for the commercial interest to the customer. However, there are signs of improvement and it is hoped that the present round of CCITT studies will lead to more universally useful and applicable recommendations.

8 In regard to the problem of engineering data network, the following arise:-

a) Specifications for international circuits, recommended by CCITT, are really applicable only to point-to-point connections. Multipoint connections are obtainable in principle, but sometimes cannot be implemented in practice due to shortcomings in national systems. This problem is particularly serious for high quality circuits, such as CCITT M102 type (AT & T type C2), where the local circuits serving the user office are provided by a local telephone company. It may be necessary for the user to provide signal processing or re-generation, in order to obtain the overall performance required.

b) Due to differences in the specification of national and international circuits, the interconnection of these by users on their own premises may require transmission equipment and experienced staff to operate it. On Boadicea system, for example, we have included amplifiers and branching equipment at line control centres to provide flexibility and circuit matching. Without these facilities the interconnection and control of the various signals from different sources would be very difficult to achieve.

c) Reliability has to be built into a system. It is important at the planning stage to consider the possible circuit failures and the available fall-back options. It is possible to negotiate particular routes, although some administrations would prefer to have absolute control of this. BOAC have been able to select certain cable routes across the Atlantic and the availability figures are monitored on a monthly basis. Where necessary, duplicate routes are used and alternate paths are specified.

d) Many national P & T Administrations are now imposing restrictions on type and model of modem which may be used on private networks. This situation has always obtained on the switched network, but a more liberal attitude has generally applied to other users. We are finding now, particularly in Europe, that modems which do not conform to CCITT recommended designs are not permitted. This poses problems to

BOAC DATA LINE CONTROL

system designers in those areas where the equipment manufacturers have previously been able to recommend suitable models for particular applications.

e) In addition to (d), some administrations restrict the use of particular types of equipment, for example multiplex equipment. These restrictions vary from country to country and may take the form of surcharges.

f) Circuit rental charges are not always based on known factors. In general, the cost of a circuit cannot be obtained prior to an order or enquiry being made. In the UK circuit rentals include a portion based on the work undertaken by the BPO in provisioning the particular circuit. Some European P & Ts base their charges on notional telephone traffic and this idea is being propagated through CCITT at the moment. The design of a network to provide the most economic service can therefore be a long and frustrating process.

g) It is possible to obtain approval for connection and use of private equipment on telephone circuits in most countries. However, the process does require formal application and processes. This can be difficult over long distances and with different languages. BOAC have been fortunate in having competent local staff who have been able to cope so far.

9 In regard to the operational aspects, it is of prime importance I believe that the user provides himself with the means to ascertain the performance and quality of the data transmission path he is using and to be able to exercise some measure of control over the testing and allocation of data transmission paths. Frequently we have found that our major problem is not the main data trunk highway, but the local end connection from the main terminal operated by the Telecommunications Authority and the BOAC premises. We have found it necessary to equip ourselves with duplicate local end facilities and to provide means of ensuring a rapid changeover. We have also found it necessary, both in London and New York, to equip ourselves well with adequate test equipment for measuring and recording the quality

of data circuit performance. In this way we find we can detect trends when the quality of line performance is deteriorating and, above all, we ourselves could find out whether any trouble is truly ours, ie in the modems or multiplexers or 'in house' equipment or whether, in fact, the trouble is with the lines.

10 Fault reporting and the control of circuits has to be tailor-made for each particular situation. The user may have some choice in the reporting point and it is worthwhile to consider the differences in quality, scale, etc. of the particular authorities involved. Some of our circuits operate through the networks of a number of P & T Authorities, ie in Europe. It is therefore essential that there is one reporting point for fault and circuit control with which our own staff will deal, in order to secure speedy restoration of the circuit.

11 On the question of costs, it has been heartening to note that progressively over the last five years costs of data channels have been reducing. We as customers and who, as good Communications Managers, are always looking for cost reductions, have done our share by the introduction of multiplexing equipment. To some extent I believe various P & T Authorities have been somewhat alarmed at the potential loss of revenue they foresee because of the use of this kind of equipment. On the other hand I maintain that it is just because the airlines in particular have been forward looking and introduced this kind of equipment, that the market will be even further expanded.

12 The advent of the computer, with its ability to process data at phenomenal speeds and thus develop a voracious appetite for data, coupled with the capability of generating a considerable amount of data to line, has brought about a new era as far as Communications engineers are concerned. We have always in the past been most concerned with the true economics of the systems for which we have been responsible. If we can reduce the number of words in a message or codify the information in some way, as the traditional communicator has always done, then economy is achieved and the system is freed for more vital traffic to flow. On the medium speed systems and the way in which they have been programmed, there is not this same devotion to economy largely, I think, because until relatively recently data processing has been an 'in house' occupation. With the advent of 'on line' or realtime systems, and even with the present data transmission capabilities, it is still just as necessary to avoid sending large blocks of spaces, even though it may be easier to programme the system that way. As communicators, we must still continue to ensure that although we have adequate capacity for data transmission beyond the immediate needs, we still need to consider the overall aspect of communications economy.

13 To operate a medium speed data transmission network on an international basis is, to say the least of it, a challenge. There are many technical problems, but it is the administrative problems concerned with the tariffs and with the day to day operation of the circuit that tend to attract most of our attention. At the present time we operate, in effect, as a single user system. Economies could be made if airlines could share data transmission services in such a way that the data stream for each airline was kept separate and thus avoid the complexities and difficulties of developing software interfaces between the various systems which might use the data transmission service. Undoubtedly, there could be economies and improvements in service, particularly in relation to fall-back channels, and economies in regard to equipment and staff to monitor and operate efficiently the data transmission service. These are matters for the future and will require much discussion and negotiation with the relevant P & T Authorities, but is undoubtedly a problem which will have to be tackled in the near future.

Note:

BOADICEA Station Address Plan, BOAC and Associated Airlines Overseas Fixed Telecommunications System Schematic, Typical Group Delay Distortion Graph and Typical Frequency Amplitude Distortion Graph are available on request from the author.

THE NETWORK CONTROL CENTER FOR THE ARPA NETWORK

A. A. McKenzie, B. P. Cosell, J. M. McQuillan, M. J. Thrope
Bolt Beranek and Newman, Inc.
Cambridge, Massachusetts, U.S.A.

Abstract

The ARPA Network allows dissimilar, geographically separated computers (Hosts) to communicate with each other by connecting each Host into the network through an Interface Message Processor (IMP); the IMPs themselves form a subnetwork that can be thought of as a distributed computation system. To detect failures in this system each IMP automatically and periodically examines itself and its environment and reports the results to the Network Control Center (NCC), at Bolt Beranek and Newman Inc., for action. The NCC computer, like any other Host, can itself fail without affecting network integrity; further, the NCC central processor can easily be replaced, in case of failure, by any standard IMP.

The present paper briefly describes the NCC hardware; discusses such software issues as NCC-related routines in the IMPs, data-collection and interpretation mechanisms, line status determination, IMP status and program reloading, and Host and line throughput; details NCC operations (manning, problem-handling procedures, track record); and summarizes overall NCC experiences and future plans.

I. Introduction

Almost four years ago the Advanced Research Projects Agency of the Department of Defense (ARPA) began the implementation of a new type of computer network. The ARPA Network provides a capability for geographically separated computers, called *Hosts*, to communicate with each other via common-carrier circuits. The Host computers typically differ from one another in type, speed, word length, operating system, etc. Each Host is connected into the network through a small local computer called an *Interface Message Processor (IMP)*; each IMP is connected to several other IMPs via wideband communication lines. The IMPs, all of which are virtually identical, are programmed to store and forward messages to their neighbor IMPs based on address information contained in each message.

In a typical network operation a Host passes a message, including a destination address, to its local IMP. The message is passed from IMP to IMP through the network until it finally arrives at the destination Host. An important aspect of this operation is that the path the message will traverse is not determined in advance; rather, an IMP forwards each message on the path it determines to be best, based on its current estimate of local network delay. Since the path choices are determined dynamically, IMPs can take account of circuit or computer loading (or failures) in an attempt to insure prompt delivery of each message.

In three years the network has expanded from 4 to over 25 IMPs and is still growing. Early work on the ARPA Network is described in some detail in a set of papers presented at the 1970 Spring Joint Computer Conference[1-5]. Additional work is described in a paper presented at the 1972 SJCC[6].

An interesting aspect of the IMP subnetwork (i.e., the set of IMPs and communication lines) is that it can be considered a distributed computation system. Each IMP performs its own tasks relatively independently of its neighbor IMPs; nevertheless all IMPs are cooperating to achieve a single goal — reliable Host-Host communication — and in some cases, for example, the dynamic path selection mentioned above, each IMP coopperates with its neighbors in making reliable delay estimates for various path choices.

In any distributed computation system it is likely to be difficult to detect component failures quickly; the difficulty is increased in the IMP subnetwork by the wide geographic separation of components. For this reason we chose at the outset to incorporate automatic reporting functions in the IMPs as an aid to failure diagnosis. Each IMP is programmed to examine itself and its environment periodically and to report the results of these examinations to a central mediating agent. This agent has the function of collecting the (possibly conflicting) IMP reports, determining the most likely actual state of the network and, in the case of failures, initiating repair activity. The mediation function is performed by the Network Control Center (NCC) located at Bolt Beranek and Newman Inc. (BBN) in Cambridge, Mass. The mediating agent is the NCC computer, which is attached, as a normal Host, to the BBN IMP. It should be noted that although the NCC computer is an *important* component of the network it is not an *essential* component; as with any other Host it can fail without disturbing overall network integrity.

The NCC computer is concerned primarily with the detection of line failures and IMP failures. In addition, the NCC computer monitors the volumes of Host traffic and line traffic; these are parameters which can give advance warning of network elements whose capacity may need to be increased and which can be used for site usage accounting. Finally, the NCC computer keeps track of other data, such as switch settings and buffer usage, for each IMP; these data are frequently helpful in diagnosing IMP failures.

The remainder of this paper describes the operation of the Network Control Center. Section II describes the NCC hardware located at BBN and Section III provides details of the overall software operation. Section IV discusses the manual procedures followed by NCC operators and technical staff in diagnosing and correcting network malfunctions. In Section V we have provided typical summaries of the types of information collected at the NCC in recent months, and mention some anticipated changes in NCC operation.

II. NCC Hardware

The central site NCC hardware consists of two packages, a central processor with 12K of 16-bit memory, a real-time clock and a "special Host interface", and a special set of hardware designed specifically for NCC functions. The current CPU is a Honeywell 316 computer; this choice provides two important advantages. First, the "special Host interface" required for connection to the network is exactly identical to the "standard Host interface" already designed as part of the IMP, thus reducing the implementation cost. Second, because all special hardware has been kept modular and external to the CPU package, if the NCC computer goes down for an extended period, it can be replaced by any standard IMP (Model 316, Model 516, or Terminal IMP). This is significant because we frequently have several IMPs on site in preparation for field delivery; thus the potential for substitution of the NCC machine is of practical value.

The special NCC hardware consists of two dial-up line controllers, a half-duplex Teletype I/O interface, and hardware associated with a panel of 32 display lights, a programmable audible alarm, and 16 control switches. All of this equipment is housed in a separate cabinet along with the required power supplies. When necessary, it can be simply connected to the I/O bus of an alternate CPU.

Two Model 35 ASR Teletypes handle most of the input and output functions. One, attached through the Teletype interface in the special hardware package, is dedicated to a print-only logging function while the other, the NCC computer's standard console Teletype, serves both as a report printer and as the primary source of operator input. Input can also be provided through the 16 control switches, and other output is given via the 32 display lights and the alarm. The dial-up line controllers are reserved for possible future use. The external I/O equipment is duplicated at nearby locations for the convenience of NCC personnel.

III. Software Operation

The IMP subnetwork consists of three principal classes of components: 1) a collection of wide-band common carrier data lines, 2) a set of IMP processors, 3) IMP system software. Network performance can be affected by failures in any of the components in each of these classes. Therefore, in conjunction with our construction of the network, we had to develop procedures for quickly detecting and repairing component failures within any of these classes. In this section we will describe the software used to assist in detecting such failures.

NCC-Related Software in the IMPs

A basic assumption, which underlies the NCC effort, is that the most effective way of detecting failures is to have each IMP periodically compile a report on the status of its local environment and forward this report through the network to a mediating agent, the NCC. This agent has the task of collecting and integrating the reports from all of the IMPs to build up a global picture of the current state of the network. The data generation within each IMP is performed by two routines: a timing routine which controls the periodic execution of the report routine, and the report routine itself.

The timing routine used is the IMP's statistics mechanism. This mechanism establishes a network-wide synchronized clock which it uses to coordinate the execution of a set of self-measurement (statistics) routines which have been incorporated into the IMP. The bulk of the statistics routines are concerned with factors such as measuring IMP bandwidth capacity and storage utilization, etc. One of the statistics routines, however, is the "Trouble Reports" routine, which provides data to the Network Control Center.

The Trouble Reports routine, when initiated by the timing routine, interrogates various parts of the IMP system to determine which lines are alive, which Hosts are up, etc. It formats that information into a message which is forwarded to the NCC's data collection mechanism. Since space is at a premium in the IMP system, the routine does no pre-processing of the information; it is merely collected and forwarded.

In addition to the statistics and reporting packages, each IMP contains a small debugging package, DDT. DDT is a simple command interpreter capable of such functions as examining and modifying a memory word, clearing a block of memory, searching memory for a particular stored value, etc. DDT is structured so that it can be driven remotely through the network, returning any responses back through the network. The remote use of DDT is important to many NCC operations.

Each IMP contains several routines which perform such NCC-related actions as "looping" data sets and line interfaces, testing Host interfaces, etc. ("Looping" is the interconnection of circuit elements such that all transmissions from an IMP are returned to that IMP rather than being sent to the IMP at the other end of the line.) DDT is used to initiate and terminate these routines by modifying words in IMP memory which contain their parameters, including an enable/disable bit. For example, one routine monitors a word which, when changed to a line interface number, loops the appropriate interface. This particular ability is vital to isolating the malfunction when a line goes down, so that we know whether to notify telephone company personnel to fix the line, or to notify Honeywell field engineering to repair the interface.

NCC Development

While the data generation scheme and (in large part) the data actually collected have remained invariant during the development of the NCC, the data collection/interpretation mechanisms at BBN have undergone steady evolution. In the first versions, while the network was small, the data were sent as ASCII text which was typed out on the BBN IMP console Teletype; personnel at BBN periodically scanned the typescript to determine if anything noteworthy was happening in the network. Since the collection was being done on a Teletype, a low bandwidth device, space within the message was at a premium; however, since a person was required to read it and make sense out of it, the format had to be intelligible. The only way to balance these factors was to omit the collection of much interesting data.

As the network became larger and more reliable, the proportion of status messages which said anything other than "everything's still OK here" decreased, thus making the location of the messages which required action on our part more difficult. The scheme we developed to make the location of critical messages somewhat easier consisted of having each IMP: 1) send us a status message every 15 minutes and 2) examine its status every minute and send an additional message at that time if it detected a *change* in status. Since these routines were being driven by the synchronized clock of the statistics package, the effect of this scheme was that every 15 minutes we would receive a block of "checkin" reports, one from each IMP; interspersed between these "checkin" blocks on the typescript there would be an occasional "change" report.

This setup functioned tolerably for some time, but eventually several factors combined to make it unwieldy. First, the number of IMPs in the network was constantly increasing, so that the amount of typescript which had to be scanned in order to determine what was happening in the network became overwhelming. Second, outside organizations became increasingly interested in receiving monthly reports on IMP and line performance; the prompt and accurate compilation of these reports by hand became more and more difficult. Third, there was pressure to take statistics on line usage and Host traffic in order to obtain advance warning of network elements whose usage was approaching saturation and to investigate accounting algorithms for network usage. All of these factors led us to install a Host on the BBN IMP which is dedicated to monitoring network performance and doing much of the bookkeeping required for our reports.

With a separate Host dedicated to monitoring the network, we were able to abandon ASCII text format in favor of binary format, and to expand the reports to include more internal status information as well as statistics on Host traffic and line usage. We were also able to increase the frequency of reporting to once a minute for the "checkin", and to send "change" reports as soon as changes are detected. We also worked, and are still working, on the knotty problem of formalizing the heuristics which are used to integrate the (often conflicting) reports from the individual IMPs. The following paragraphs discuss several of the problem areas of greatest interest.

Line Status

For its own routing purposes, each IMP is continually measuring the quality of each of its data lines. Every half-second it sends a thousand-bit status message on each line and expects to receive a similar message from its neighbors. Each status message includes the number of the IMP which originated it. When an IMP receives a status message from one of its neighbors correctly, it marks its next status message to that neighbor with an acknowledge bit. Thus an IMP's receipt of a status message with the acknowledge bit set indicates that the line is in good condition. Conversely, whenever a half-second interval elapses and the IMP does not receive a status message with an acknowledgment of its own previous message, it counts an error on that line.

In conjunction with this acknowledgment scheme, an important system debugging feature is the ability to "loop" lines for test purposes. Each line is nominally a pair of independent one-way circuits, one in each direction. "Looping" is the interconnection of these circuits such that one end is disconnected and the other end receives its own transmissions. A line can be looped in one of three places: either inside the IMP's line interface or at the local data set (under program control), or at the remote data set (manually). The IMP system, by checking the origin of status messages, can detect looped lines.

Using its line error count and detection of looped lines, an IMP can make a simple usable/unusable decision, for its own purposes, for each of its lines. A line can, however, be "network unusable" for a variety of reasons (the IMP at the other end is down, the interface on the local IMP is broken, the line itself is broken, etc.) and at the NCC we must be able to distinguish amongst them in order to initiate the appropriate repair procedure. Therefore we supplement the IMP's report of whether it thinks the line is up, down, or looped with the IMP number of the IMP on the other end of the line, the total number of status messages sent on the line, and the total number of status messages received on the line (whether their acknowledge bit was set or not). The NCC takes the 3-way division from the IMP (up, down, or looped) and incorporates into it a 2-way division (status messages coming in or not) to form a 5-way breakdown of line status as seen by the IMP at one end of a line: up, down without errors (unusable but with status messages without errors being received), down with errors, looped, and "no information" (the IMP has not reported to the NCC recently). Every minute, for each line in the network, the NCC takes the latest status for each end of the line and determines the state of the line according to the decision rules shown in Table 1. Whenever any line's state changes, a message is printed in the log.

The IMPs are essentially synchronized with regard to the generation of status messages; furthermore, status messages constitute a known constant traffic load on each line. Therefore, for lines whose state is declared up, a measure of line quality is given by the fraction formed by dividing the number of status messages correctly received by the number of status messages sent, since only line errors (detected by checksum hardware) will cause status messages to be incorrectly received. This fraction is printed in the log whenever the numerator differs from the denominator by more than one and the fraction is neither zero nor one. Thus we are alerted to line failures before the lines become completely unusable.

Since the IMPs have been designed to infer the network's topology dynamically, they are not directly concerned with the common carrier data lines; rather, they are interested only in which portions of the network they can access through a particular line interface. NCC personnel, however, must deal with the actual lines. A report from an IMP that the line connected to interface 2 has become unusable is not useful unless we can determine which line is actually connected to that interface. Toward this end, the NCC maintains a connectivity table which contains, for each line in the network, the IMP numbers for the IMPs at each end and the interface numbers that that line should be connected to. The NCC types a message in the log whenever it determines that a line has been moved from its nominal interface or when a report for a line not contained in the connectivity table is received.

IMP Status and Program Reloading

The NCC is faced with a difficult problem in attempting to determine that an IMP is no longer functioning. Since a broken IMP can't send us a message indicating that it's broken, we must infer this condition from the absence of its "check-in" messages. In the past, this decision was made after a scan of the typescript and the observation that the IMP had not checked in "for a while". The current NCC system declares an IMP dead when it has not reported for three minutes. Because of the effects of problems like network partitioning, this is an inadequate test for actually determining whether the IMP is up or down, but it does alert our personnel to the need for further diagnosis. For example, all lines to the IMP may

STATUS FROM HIGH NUMBER IMP

		UP	DOWN NO ERRORS	DOWN WITH ERRORS	LOOPED	NO INFORMATION
STATUS FROM LOW NUMBER IMP	UP	UP	IN LIMBO	IN LIMBO	IN LIMBO	UP
	DOWN NO ERRORS	IN LIMBO	DOWN	DOWN ON HIGH END	LOOPED ON HIGH END	UNKNOWN
	DOWN WITH ERRORS	IN LIMBO	DOWN ON LOW END	DOWN	LOOPED ON HIGH END	UNKNOWN
	LOOPED	IN LIMBO	LOOPED ON LOW END	LOOPED ON LOW END	LOOPED ON BOTH ENDS	LOOPED ON LOW END
	NO INFORMATION	UP	UNKNOWN	UNKNOWN	LOOPED ON HIGH END	UNKNOWN

The terms are defined as follows:

HIGH NUMBER IMP - IMP with higher network address
LOW NUMBER IMP - IMP with lower network address
UP - The line is usable for both IMPs.
DOWN - The line is unusable for both IMPs.
DOWN ON ONE END - One IMP can transmit to the other, but not vice versa.
LOOPED ON ONE END - The line is looped as seen by one IMP, but not as seen by the other.
LOOPED ON BOTH ENDS - The line is looped as seen by each IMP.
IN LIMBO - Conflicting reports from the two IMPs.
UNKNOWN - Insufficient data to make a decision on the line's state.

TABLE 1: DECISION RULES FOR LINE STATE

be down, rather than the IMP being down.

In the rare case of network-wide failures it is often difficult to determine which IMP triggered the network failure, much less what caused that IMP to fail. Nevertheless, personnel at the NCC must attempt to make these determinations. To assist them, each status report that an IMP sends to the NCC contains a snapshot of the IMP's environment. The snapshot information is used to determine if the IMP is experiencing a transient or getting into some kind of trouble. This information includes the version number of the program running in the IMP, the storage utilization and the amount of free storage left in the IMP, the state of the sense switches and the memory protect switch (to detect unauthorized tampering), a list of the statistics programs which are enabled, a list of which Hosts are up, and an indication of whether tracing is enabled. The NCC logs any change in reported status and, in the event of a network failure, we attempt to correlate environmental data for individual IMPs with the network failure as a whole.

Since all IMPs run the same program, we built a small "bootstrap" routine into the IMP which, when initiated, sends out a request for a core image on a line selected either by parameter or at random. When any IMP receives such a request it returns a copy of its entire (running) program as a single message. The bootstrap routine then checks incoming messages on the selected line for correct length and checksum; when the core image is successfully received it is initialized and started. If an incorrect core image is received, the bootstrap routine sends another request. This facility provides a quick and easy way to obtain a fresh copy of the IMP system. Since the bootstrap resides in the protected memory sector, and thus is nearly always intact, site personnel are almost never required to handle IMP system paper tapes when an IMP requires reloading.

Program reloading can be initiated remotely by commanding DDT to execute a transfer to the bootstrap. Without the remote reloading ability, the only way to distribute a revision of the system would be to mail out paper tapes of the new program to each site, and then schedule a time, with personnel available at each site, to load and start the new version. With the remote reloading ability, however, we merely load the new version into the BBN IMP, direct BBN's neighbors to reload from us, then direct their neighbors to reload from them, and so on until the new version is propagated through the entire network. In fact, propagation of a new program release can be accomplished by one person in a few minutes, rather than requiring a month of planning and several hours of work by a nationwide "team". Also, since the procedure doesn't require assistance from site personnel, it can be scheduled to occur at a time when network usage is extremely low (typically very early morning, a time when site assistance would be most difficult to arrange), thus minimizing the loss of network availability.

Host and Line Throughput

With the change from ASCII to binary reporting and the consequent easing of bandwidth limitations on the NCC we have been able to take initial steps toward building an accounting facility for network usage. Toward this end, the IMP measures the amount of use each Host has made of the network in eight categories. The eight categories are the combinations of the following parameters: transmissions from and to the Host, inter-site and intra-site transmissions, and packet and message traffic. Thus, the eight categories are:

1. inter-site messages sent
2. inter-site messages received
3. inter-site packets sent
4. inter-site packets received
5. intra-site messages sent
6. intra-site messages received
7. intra-site packets sent
8. intra-site packets received

The IMP counts data transmissions only; control messages and RFNMs (destination-to-source message acknowledgments) are not included. The NCC tabulates Host traffic data from all the IMPs in the network. At the end of each hour it copies this table into a second table and then clears the first to obtain a clean "snapshot" for the hour, which is then printed on the report Teletype. This table is also added into a daily table which is printed at midnight every day.

In order to be able to better predict when lines may become overloaded, we also keep track of the line utilization in the network. The IMP measures the line throughput by counting the number of successfully acknowledged packets. The NCC accumulates these line throughput data for each line and types them out with the Host throughput at the end of each hour, and at the end of the day.

Visual and Audible Alarms

Although the computerization of the NCC virtually eliminated extraneous typescript, it was still desirable to free the NCC personnel from having to regularly check the typescript to determine whether action was required. We therefore attached a set of lights and an audible alarm to the NCC. The NCC maintains two sets of "virtual light" display information: which IMPs are alive and which lines are functioning. The NCC staff can select either of these sets for output in the physical display lights. This provides for a quick visual survey of the state of the network.

Whenever a line breaks, or an IMP stops working, the alarm is sounded and the virtual light for that IMP or line is flashed. This minimizes the time for the NCC personnel to notice and take some corrective action, while at the same time freeing them from having to watch the lights or log to achieve this rapid response to network failures.

IV. NCC Operation

The Network Control Center staff consists of five computer operators and several regular BBN technical staff members. The operators are familiar with the operation of the NCC machine and, to a certain extent, with the diagnosis and resolution of network problems. The technical staff members are both hardware and software specialists, most of whom have participated in the design and implementation of the network from the outset.

As the network developed and the role of the NCC increased and became more clearly defined, a fairly comprehensive scheme for manning evolved. It became clear very early that 9 to 5 coverage with informal arrangements to contact staff members at home was insufficient. A dedicated Network Control Center telephone line was installed, with computer operators acting in a monitoring capacity to direct inquiries and problems to available staff members. This has become the single contact telephone for Host site personnel, the telephone company and Honeywell field engineering.

The present NCC program, with its detailed log, allows the operators to assume first order responsibility for network operation. Operators now man the NCC 24 hours a day, 168 hours a week; technical staff coverage is normally close to 50 hours a week, with additional "at-home" availability of key personnel. Routine chores are handled by the operators and only more complicated situations are referred to staff members. After hours, the operators attempt to contact specific staff members at home in the event that a problem arises or a phone call is received which the operators cannot properly field. In specific rare cases, such as attempting to pin down an obscure hardware malfunction, a problem may be preserved (i.e., not fixed) until a staff member can investigate. Even outside regular working hours, most problems are resolved, or at least under control, within a few hours.

There are several different means of handling problems, depending on severity and type. For routine controlled situations (such as IMP preventative maintenance and scheduled repair, scheduled Host testing, and scheduled line test and repair) the NCC operators coordinate the activities of the Honeywell field engineering, Host site, and telephone company personnel involved. We have established the policy that the state of an IMP or a line is not to be intentionally modified without first seeking the permission of the NCC. We insist upon strict adherence to this policy in order to prevent a deferrable outage from occurring during an unscheduled failure and thereby jeopardizing network integrity.

The alarm calls attention to IMP or line outages. The display lights, in conjunction with the log and the ability to obtain a quick printout of network status, usually make it fairly easy for the operators to determine what has failed. In the case of an indicated IMP failure, the operator on duty calls the IMP site, verifies that the IMP has failed, gathers some rudimentary information as to the type of failure, and enlists the aid of site personnel to bring the machine back on the network. If this is not possible, technical staff members are called in to investigate further. If a hardware malfunction is indicated, Honeywell field engineering is alerted to repair the problem.

At present, IMP maintenance and repair are carrier out under contract by Honeywell field engineering. Coverage is prime shift with guaranteed 2-hour response time. When circumstances warrant, however, the NCC will request extended coverage for repair or for standby backup. Most repairs are completed by the end of the day they are reported.

In the case of an indicated line failure, the operator performs a series of checks to confirm that the line has actually failed. This is necessary since some IMP failures appear in the log as line failures (the converse is also true). Diagnosis is performed from the NCC by using IMP DDT to test the terminal equipment. If a line failure has isolated a site from the NCC, the operator will contact site personnel and direct them in performing the tests for him. When a line problem has been confirmed, the operator notifies the appropriate telephone company office, frequently supplying considerable detail.

Each line is maintained and tested from a private line office at one end. Manned around the clock, these offices are equipped with test facilities for finding and repairing line problems. Unless there is a manpower or access problem related to local facilities, line failures usually are corrected within a few hours of the initial report. Maximum repair time is normally about a day.

For many NCC activities the cooperation of the sites is essential. Site personnel aid the NCC in the diagnosis of a variety of problems, help in recovering from IMP failures, and take local responsibility for the IMP. Their assistance is particularly useful when investigating obscure hardware and software malfunctions.

Our relationship with the organizations involved in network maintenance has been good. Honeywell field engineering, telephone company, and site personnel have a high regard for the conclusions reached in our problem analysis. This believability has been fostered by a good track record; in at least 75% of all failure reports to Honeywell or the telephone company, an actual problem has been detected. Line problems and many IMP problems are usually fully diagnosed and dispatched to the appropriate maintenance group within half an hour. Some more subtle IMP problems, however, occasionally require gathering data over a number of failures before a conclusion can be drawn.

V. Experience and Future Plans

A great deal of additional work is done with the NCC's log and summary reports in order to produce monthly reports on network status and usage for ARPA and other interested parties. Since the NCC machine has no secondary storage capability we are unable to accumulate monthly summary information on that machine; instead the daily summaries and log information must be used as input to manual preparation of Host traffic reports and IMP Down

Month	Average Line Outage	Average IMP Down	# of Nodes	Average Host Inter-site Output (packets/day)
September 1971	.59%	3.27%	18	51,386
October	1.66%	1.77%	18	95,930
November	1.65%	5.50%	18	116,515
December	3.21%	3.95%	19	107,896
January 1972	1.02%	1.92%	19	172,037
February	1.23%	2.73%	19	224,668
March	1.36%	4.00%	23	240.144
April	.88%	2.86%	25	362,064
May	1.11%	2.57%	25	505,639
June	.41%	.97%	29	807,164

TABLE 2: SUMMARY OF NETWORK OPERATION

and Line Outage summaries. A certain amount of judgment is used in the latter two summaries; several outages which the technical staff feels are due to a single cause are normally combined into a single (longer) reported outage. Table 2 provides summary information for an actual ten-month period of network operation.

Until early this year the Host traffic summaries were produced manually from the NCC's hard-copy summary reports. The NCC now punches a paper tape (on the report Teletype) of all daily summary information and this is used as input to computer programs which produce the reports more rapidly and accurately. Eventually, when experience indicates that several of the network's service Hosts are reliably up around the clock, we expect to have the NCC transmit all of the summary information through the network for storage and later manipulation. This will enable us to more easily provide answers to interesting questions such as:

- What are the peak hours of network use and what is the peak-to-average traffic ratio?
- What percentage of network traffic do single-packet messages constitute, and how does this percentage vary from Host to Host?
- What is the ratio of weekday use to weekend use?
- What percentage of line capacity is used during peak hours, on the average, and during weekends?

Although the data needed to answer these questions are available now, the data manipulation required constitutes a prohibitive manual burden. Thus, the installation of an NCC computer lifted one bandwidth limitation only to reveal another. In an attempt to deal with this new problem, we are planning to experiment with an additional Host which was recently added at BBN. This is the machine which is currently being used "off line" to process the paper tapes mentioned above.

Another change which is under consideration is automated single-point reporting of line problems. The NCC program, after appropriate automated line testing, could report confirmed line failures directly to a Teletype at some telephone company central location via one of the dial-up line controllers. Telephone company personnel would then direct this report to the appropriate office for test and repair.

Finally, certain of the NCC command options will be made available to other organizations (such as the ARPA office) via one of the dial-up line controllers. This will be primarily to allow access to information on the overall state of the network, particularly the up/down status of the IMPs and lines.

Acknowledgment

This work was sponsored by the Advanced Research Projects Agency under Contract No. DAHC15-69-C-0179.

References

1. "Computer Network Development to Achieve Resource Sharing," L.G. Roberts and B.D. Wessler, Proc. 1970 SJCC.
2. "The Interface Message Processor for the ARPA Computer Network," F.E. Heart, R.E. Kahn, S.M. Ornstein, W.R. Crowther, and D.C. Walden, Proc. 1970 SJCC.
3. "Analytic and Simulation Methods in Computer Network Design," L. Kleinrock, Proc. 1970 SJCC.
4. "Topological Considerations in the Design of the ARPA Computer Network," H. Frank, I.T. Frisch and W. Chou, Proc. 1970 SJCC.
5. "Host-Host Communication Protocol in the ARPA Network," C.S. Carr, S.D. Crocker, and V.G. Cerf, Proc. 1970 SJCC.
6. "The Terminal IMP for the ARPA Computer Network," S.M. Ornstein, F.E. Heart, W.R. Crowther, H.K. Rising, S.B. Russell, and A. Michel, Proc. 1972 SJCC.

THE EUROPEAN COMPUTER NETWORK PROJECT

D.L.A. Barber
National Physical Laboratory
Teddington, Middlesex, England

In November 1971 the Ministers of eight European nations; France, Yugoslavia, Italy, Norway, Portugal, Switzerland, Sweden, and the United Kingdom, together with the Euratom Centre at Ispra, signed an agreement to start a project aimed at building a European computer network. The network, which will use the store-and-forward packet-switching techniques, will initially join five data processing research centres in four countries; but later is expected to be extended to include centres in other Nations.

The paper sets out the arguments that led to the decision to go ahead with a European Computer Network, and discusses the form it will take, the functions it will perform, and the way the project will be conducted.

Introduction

Early in 1969 a working group was established by the European Economic Community to examine research in science and technology. This working group, chaired by a Frenchman, Monsieur Aigrain, made wide ranging proposals for various advanced projects in science and technology to be carried out jointly by members of the community. In April 1970 the EEC working group was widened to include representatives of other European countries and the resulting group of nineteen nations became known as the COST Group. (Co-operation Europeene dans le Domaine de la Recherche Scientifique et Technique).

The COST Group set up a number of expert study groups to consider, in detail, the various projects. One of these, known as Project 11, was concerned with the problems of communication between computers. The Study Group, chaired by the author, put forward a plan to establish a Pilot Computer Communications Network in Europe, to link a number of data processing research centres. This was accepted by several of the member nations, who signed in November 1972 an agreement to undertake a project aimed at building such a network. Initially, the network will link together centres in France, Italy, Switzerland and the United Kingdom, but later is likely to be extended to include centres in other Nations.

This paper sets out the arguments that led to the decision to go ahead with a European Computer network, and discusses the form it will take, the functions it will perform, and the way the projects will be conducted.

The Need for a Network

It is now well accepted that the proper use of computers is one of the most vital factors affecting the economic growth of a nation. In the mid 1960's the association of telecommunications facilities with computers vastly increased their effectiveness and, since then, the use of teleprocessing techniques has grown at a remarkable rate. There is little doubt that use made of communication networks linking users to remote computers will continue to increase sharply during the next few years, and that the transmission of data across Europe will become of vital importance to the economics of European countries in general. The pattern which is beginning to emerge in national data communications will be duplicated across national boundaries, and there are already several instances of private computer communication networks that incorporate international data links.

However, the growing numbers of private data communications networks owned by Industry, Commerce and Government is a cause for concern, because private networks often under utilise telecommunications resources. And, because the number of connections to remote computers is tending to double every year, various national proposals have been made for more efficient networks which may be shared by many users.

When an attempt is made to connect several computer systems by a shared data network fundamental incompatibilities are usually revealed between them. The development of shared networks, therefore, focuses attention on the need for really effective agreement of standards for communication between, and with, computer systems.

The provision of efficient shared data communications facilities can also have a useful effect on advanced computer system development by allowing co-operation between research centres. An excellent example is the Advanced Research Projects Agency (ARPA) network[1] sponsored by the Department of Defence of the United States. The ARPA network is clearly having a most beneficial influence upon the development of computing in America.

Networks of a similar type have been proposed and investigated by Universities and Research Establishments in Europe and there are various plans under consideration for building such networks in some countries. There could be many advantages if these networks were joined by an International network, and this would be particularly the case if the international networks were compatible with it, and each other.

It is of the utmost importance, therefore, that early and adequate thought be given to the problems of international networks and of standardisation in particular, so that national developments do not produce a set of incompatible systems which cannot be linked without great difficulties requiring special interfacing or even re-design. International bodies like the CCITT (International Telegraph & Telephone Consultative Committee) and CEPT (European Committee of Post & Telegraph) are studying the telecommunication aspects of these networks, but many of the problems concerned with the linking of computers both to other computers and also to various types of terminal equipment, are of a somewhat different nature to those concerned with existing telecommunications facilities; and there is already much debate on how best to deal with these new problems both nationally and internationally. It was partly the conviction of the increasing importance of computer networks on a European scale, and partly a consciousness of the many problems which still need to be solved before such networks can be effectively implemented, that led to the decision to build an experimental computer communications network. This will link various European data processing research centres and will serve two main roles:-

1. It will facilitate the exchange of ideas and the co-operation between data processing research establishments, and will permit the eventual sharing of resources between Centres. It will also highlight the practical difficulties of doing this on an international scale.

2. It will form a focus for discussion and comparison of presently proposed National network schemes. These may be held back at present because most countries are reluctant to embark on a purely National scheme that may be incompatible with the schemes of other Nations; a common meeting ground could well promote the resolution of differences in the various National proposals.

In addition, if widely publicised, an experimental network could serve as a model for new commercial and other networks which are at the planning stage. In this way, incompatibilities between future data processing systems built in Europe could be minimised.

The Nature of the European Network

Experience gained with present day remote-access data processing systems indicates that the data processing aspects and the communications aspects of these systems should be separated. This is especially so in an experimental network where development of the two aspects may be required to proceed almost independently.

A vital feature of any network is high reliability, and this may be achieved by connecting the centres by a mesh network of data links, with sufficient links to provide at least two paths between any two centres in the network. In the event of link failure, certain Centres may have to redirect and forward traffic and this must not affect the normal data processing activities at the Centres. An attractive method of achieving this is to provide a separate computer at each Centre to control the communications links and handle through traffic. This computer need only involve the main computers in activity when local traffic appears.

The European Computer Network may, therefore, be looked at from two aspects: it will be primarily a network of data processing centres co-operating to provide facilities, or achieve a performance, unattainable by individual centres; but within this network of data processing centres, the necessary communication facilities may be regarded as a sub-network, which can be established and managed as a separate entity. This concept has many similarities with that of the ARPA network. However, it is important to remember that the European Network will have to include any special features that might be necessary to allow communication across national boundaries, between data processing centres with, hitherto, largely uncorrelated research programs.

The precise details of the co-operation between the data processing centres is, as yet, not completely formulated, and will, in any case, continue to develop in a variety of ways over the next few years. The design of the communications sub-network must be flexible enough to permit such developments to take place, and it must be capable of being adapted to suit changing requirements. Indeed, the adaptation and development of the communication sub-network will be one important aspect of the future programme of co-operative research between the national centres.

At this stage, then, details of the data traffic to be handled by the network cannot be certain, but it is likely to be of three fundamentally different types. There will be considerable traffic between the operating systems of co-operating computers; this is likely to be an interchange of high speed bursts of data each of a few hundred characters. There will be traffic from manually operated terminals; this may have some similarities to traffic between computers, but the bursts are likely to be much shorter in duration, and will be generated at much lower rates. The third type of traffic may arise towards the end of the experiment, when data collection and telemetry systems may begin to use the network. This kind of traffic is likely to be very infrequent, and very brief, comprising only a few characters in each message.

The experience with the ARPA network, and the work at the National Physical Laboratory[2] in the United Kingdom, indicates that the packet-switching technique offers a number of advantages for handling the expected forms of traffic. The communications sub-network will, therefore, be a high-speed store-and-forward packet-switching network which, in addition to its role of providing communication between co-operating data processing centres, will be used as a vehicle for a number of experiments in communication techniques. Typical of these experiments will be an investigation of different packet routing techniques, and a study of the effect of hypothetical tariff structures. The ability to make such investigations will be a feature of the design of software for the communications computers, while provision will also be made for the collection of statistics about the internal operation of the network, and the way in which it is used by the data processing centres. These statistics will help to show the merits and demerits of packet switching methods in relation to the needs of international data communication.

The Implementation of the Network

The task of building a network to link research centres in different countries is obviously more difficult than that of building a purely national network, for the already considerable problems of linking dis-similar data processing systems are augmented by those of co-ordinating teams of scientists in each country working on various aspects of the project. The situation is further complicated by the fact that Nations may participate at two levels: they may take part in the study and design phase only, or may additionally go on to equip and operate a network centre.

The management of the project will be the responsibility of an international committee with one representative from the Government of each National taking part. Each representative will have one vote, and decisions will be taken by a simple majority, provided it includes the Nations that have elected to nominate a Centre.

The management committee will select a director and a small permanent executive body to be responsible for the day-to-day control of the whole project. This executive body will comprise members nominated by some of the participating countries, and will probably be resident at one of the National Centres.

An advisory committee formed from representatives of the data processing centres will draw up a program of co-operative research between the centres, and also between other establishments that may be able to co-operate with each other using the COST 11 communications sub-network; possibly gaining access to it through any national network that may be available.

The communications sub-network itself will be planned in detail by commerical consultants working to a specification already prepared by the Study Group which formulated the project. The specification covers the design of a standard nodal centre, the use of which is to be encouraged at all centres, otherwise the introduction of the network will be difficult. However, it could prove unduly restrictive, and might inhibit beneficial evolution, if alternative designs were entirely prohibited. For, in the long run, it must be possible to introduce improved versions of a node into an existing network. Of course, anyone contemplating a non-standard design risks being unable to communicate effectively with other centres, though these should still be able to interchange messages among themselves, because the mesh network will allow any faulty node to be by-passed.

Network Configuration

The initial form of the network will be as shown in Figure 1; with Nodal Centres in Ispra, London, Milan, Paris and Zurich. Further links and centres will be added later depending on the way the experimental network develops. There will be two alternative paths between any pair of centres initially, with the likelihood that further paths will be available later.

The functions of each nodal centre (N.C.) will be as follows:-

a. To control a number of communication lines to other nodal centres.

b. To forward transit traffic between adjacent nodal centres.

c. To handle traffic between the network and a certain number of secondary centres associates with the nodal centre.

d. To collect traffic statistics and carry out accounting computation.

Associated with the node there may be a number of secondary centres (S.C.). These may be as shown in Figure 2:

1. Subscribers computers devoted mainly to local data processing tasks; initially these will be the existing computers at the national data processing centres.
2. Terminal processors capable of collecting messages from a number of local terminals requiring access to the facilities of the main network.
3. National network processors - these might be associated with national data communications networks, which would be linked together through the experimental European international data network.

A nodal centre will often be connected to its secondary centres by local connections, particularly in the initial stages of the network; but provision will also be made for connections between a nodal centre and remote secondary centres.

To cater for the expected growth of traffic between the initial centres of the network, and for the anticipated expansion of the network to include centres in other nations, the nodal centres will have to be designed to be highly flexible. It is very likely that considerable alterations may have to be made following the initial operating experience, or arising from a change in the way the network is used. It is, clearly, essential that the choice of both the software and hardware aspects of the nodal centre shall permit changes to be made with relative ease; it is likely, therefore, that both hardware and the software will be designed to be modular in form, and the decision about which to use to perform specific functions will be influenced considerably by the need to retain the utmost flexibility. The possibility of making changes to the operating systems of the nodal centres from a remote location will also be considered both from the point of view of maintenance of the network, and of enhancing its adaptibility to changing requirements.

The Nodal Centres

The transfer of information within the network between the Nodal Centres will be by fixed-format packets, but the interchange of information between the Nodal Centre and the Secondary Centre may be in the form of messages of any length in multiples of a byte of eight bits up to a maximum of 255 bytes.

There are two distinct functional parts envisaged within the Nodal Centre: a 'packet-switch' and a 'data-exchange' (see Figure 3). The packet-switch is responsible for the transfer of packets between the communications lines to other Nodal Centres (transit traffic) and between these lines and the data-exchange. The data-exchange organises the transfer of messages to and from the Secondary Centres (local traffic) and performs the translation between messages and packets, i.e. it divides long messages into packets for transmission to remote centres, and assembles incoming packets into messages for local delivery.

To carry out these tasks the Nodal Centres must be capable of:

- receiving messages from the attached S.C.s;
- receiving packets from others N.C.s;
- forwarding the above messages or packets to its attached S.C.s or other N.C.s;
- providing store-and-forward switching for messages exchanged between one S.C. and another S.C.;
- providing for speed conversions;
- managing message and packet priorities.

In addition to these functions the Nodal Centres will be involved in:-

- the breaking of messages into packets,
- the management of message storage,
- the routing of messages and packets,
- the generation of formatted messages,
- the automatic error detection and correction,
- the coordination of activities with other N.C.s,
- the coordination of activities with its S.C.,
- the collection of network and traffic parameters,
- the detection and location of faults.

The nodal centre will be implemented in the form of:-

- a communication processor (or processors);
- hardware interfaces attaching the communication processor to the data sets;
- hardware interfaces attaching the communication processor to the secondary centres;
- software for message transmission, validation, failure detection and recovery, and data gathering;
- software for message reformatting required to meet the procedural standard interfaces to Secondary Centres.

Each nodal centre will be capable of handling a maximum number of 4 lines to other nodal centres and a maximum number of 8 lines to remote secondary centres. But the addressing capability of computers will provide 16 parts to permit possible future expansion, or the addition of special peripheral equipment.

The Secondary Centres

The National Centres at present use data processing systems of different manufacture, so the subscriber computers that will form the initial secondary centres will, in the main, be different. The European Computer Network will thus be highly inhomogeneous.

Due to the variety of S.C. machines and to the differences in the methods by which they communicate with external equipment it is not intended to make individual modifications to the N.C. in order to interface to each S.C.; instead a single standard access interface will be defined.

The interworking of S.C. machines with the standard accesses interfaces of the ECN demands the introduction of specially designed "adaptation units", involving in the general case both hardware adaptation and procedure modification.

The secondary centre will therefore comprise the following components:-

- The existing hardware and software which is independent of the formation of the network, and will remain unchanged.

- New hardware and software individually designed and implemented for the convenience of the user in attaching his data processing systems to the standard access interfaces of the network.

Communications Path and Interfaces

The general communication path diagram of the network of Figure 4 shows a connection between two secondary centres passing through an undefined number of nodal centres. Along the communications path between secondary centres a number of standard interfaces and procedures have to be considered.

The type of interfaces are as follows (Figure 2, Figure 4):

I1 - "Line Interface" - Interface between the nodal centre and the communications facilities.

I2 - "Standard Access Interface" - Interface between the nodal centre and the secondary centre.

I3 - Interface between the present computer at any particular centre and the "adaptation unit" required to match it to the network.

Each interface has to be defined in terms of:

- hardware,
- procedures.

The hardware interfaces I1 and I2 will use relevant CCITT recommendations. But the procedures necessary to allow two parts of the system to communicate with one another, are still being defined. They are of three kinds (Figure 4):

P_{SN} = "SC - NC communication procedure", this allows a secondary centre to communicate with its nodal centre.

P_{NN} = "NC - NC communication procedure", this allows a nodal centre to communicate with another nodal centre.

P_{SS} = "SC - SC communication procedure", this allows a secondary centre to communicate with another secondary centre through the network.

Supervisory Centre

A "Supervisory Centre" will be provided at one of the National Centres. It will perform the following functions:-

a. The collection of statistics about network and traffic parameters.

   These parameters may include:

   - distribution of network traffic,
   - occupancy of communication circuits,
   - delays encountered by messages,
   - message lengths,
   - queue lengths,
   - message arrival rates,
   - utilization of the nodal centre storage,
   - utilization of the nodal centre CPU time.

b. The collection of statistics about detection and location of faults concerning centres and lines.

c. The transmission of up-to-date information to the nodal centres of the network, e.g., concerning new configurations of the network or new routing strategies.

d. The preparation of accounting computations.

Characteristics of Communications Links

The circuits used to connect together the nodal centres of the network will be private leased circuits fitted with high-speed modems. The modems may be provided by the local PTT, or if permitted, may be proprietory modems of high performance. A transmission speed of 48 Kilo bit/sec will be used where possible, though it may be necessary to accept a reduced transmission speed if a group band circuit is not available between two centres.

Some modems will be provided connected to the public switched network with appropriate automatic calling and answering equipment to enable the packet-switching computer to set-up calls through the public switched network. This facility will be provided to allow an additional circuit connection to be made to any other nodal centre in the event of the failure of a leased line, so that at least some degree of intercommunication may be retained for emergency traffic.

Provision will be made for the looping of circuits by operators or engineers for test purposes, both on the line side and on the packet-switch side of the modems and line control equipment. It will also be possible to reconfigure the connections to incoming lines by manual patching so that in the event of catastrophic failure of a nodal centre alternative direct connections may be made between lines or modems.

The circuits joining remote Secondary Centres to the Nodal Centres may be private leased lines operating at 4,800 bits/seconds. However, there may be greater use made of the switched network to make temporary links with secondary centres that use the international network only occasionally. It is likely that automatic calling and answering equipment will be used for some of these lines.

Nodal Centre to Nodal Centre Communication

The Nodal Centre will be able to handle up to six trunk lines to other Nodal Centres using an extensible communication multiplexer in which channels will only be provided when required. Each channel may operate asynchronously and independently of the others; and can handle transmission rates from 2.4 to 48 kilo bits/sec. In addition consideration is being given to the possible use of PCM links at some future date. Where these are available, their lower noise level, and their much higher speed, make them very attractive. The interface between each channel of the multiplexer and a data set is, of course, the standard full duplex interface as defined in the CCITT V 24 recommendation.

The multiplexer channels will perform by hardware functions such as serial to parallel and parallel to serial conversions, recognition of flags and commands, generation and checking of the cyclic redundancy check sum as well as various test facilities, such as allowing the output of the channel to be looped into its input, for testing purposes.

The structure of the packets used for communication between Nodal Centres comprises a fixed length packet header and a variable length text. The packet header identifies the text and gives some of its characteristics, for example it contains information on:

- the type of the packet, e.g. a control packet carrying NC to NC information, or a user packet carrying part of a user message.
- the origin and destination of the packet; its priority and an order number identifying the relative position of the text in the user's message.
- instructions about the routing of the packet and any measurements that are to be triggered in each Nodal Centre through which it passes.

The text, which contains an integral number of bytes of information as indicated by the packet length character in the header, may be either user text, passed through unchanged; control information destined for a nodal centre or statistical information being collected by the network measurement Centre.

Evaluation of the Network

To ensure that the results from the test with the experimental European Computer Data Network are of practical use for the future, it is necessary to consider not only the technical features which should represent as far as possible any likely final system, but also the commercial, and operational aspects. The statistical measurements and the results inferred from them will not be representative of the likely future commercial use of the network, if every participant in the experiment can use the network without any costs. For example, as no charges are being made for using the network, it will be a most attractive means of data transmission and high-speed intercommunication; so the load of the network will be substantially higher than if some form of tariff were to be charged. There will, therefore, be a notional charge computed to endeavour to indicate whether such a network could be of any commercial value.

During the first stage of the test only the participating national centres will exchange data information. But when tests have proved the reliability and utility of the network for international communications the various nodal centres will be able to invite participation from universities, and other research institutes, but before this is done, a tariff structure will be devised which covers all the features of the overall network, including any services available at the data processing Centres. The tariff will include the use of programs, computer capacity, data banks, etc. The knowledge of the cost of such services over an internationa network will make it possible for potential commercial customers to assess the economic value of joining such a network.

The profit from operating the network - even if only a notional one - will enable an economic case to be made for its enhancement by the addition of new equipment and services.

Tariff Parameters

The details of the tariff structure have yet to be decided, but will probably include factors such as:-

a. Various levels of priority of messages.

b. The quantity of information handled.

c. The connection time.

d. The distance between Nodal Centres.

e. The special facilities that may be offered.

Statistical Measurements

The fullest possible statistical information will be collected from the experimental network. Initially, this will be used for the analysis of the efficiency and operating characteristics of various parts of the network, but later it will help to determine a notional tariff structure as discussed above.

The collection of statistics places an additional load on the network but in the early stages this is unimportant. Later only the more important parameters may be collected - these will be chosen after operational experience. At first information will be gathered both locally at each Nodal Centre, and at a nominated statistics collection centre, whether this procedure will continue is yet to be decided.

There will be two types of measurements to be made: those concerning flow of information across interfaces between parts of the network, and those concerning the internal behaviour of equipment.

Within the Nodal Centre measurements will be made at the packet-switch level. These will concern the trunk line behaviour, i.e. error rates, incidence of line failure, usable bit rate, variation with time, modem performance etc. The distribution of traffic on trunk lines will also be measured as a guide to new circuit provision, as network loading increases. The internal behaviour of the packet-switch will be monitored to indicate where improvements in its software might enhance performance.

A similar series of measurements will also be made in the Data Exchange portion of the Nodal Centre. Here it will be local line and interface behaviour that is of concern, while traffic distributions will reveal the ratio of local to trunk traffic at each centre. The internal behaviour of the Data Exchange software will also be recorded. One final function of the nodal centres will be to monitor the behaviour of the attached Secondary Centres, so the extent to which any malfunction of these centres can cause additional loading of the network may be established.

Conclusion

The European Computer Network project will be of great value in coordinating the work of data processing research establishments in Europe. It will serve as a test-bed for new ideas and as a focus for discussion about National networks. Now that the European Community is being enlarged as more Nations join it, the importance of computers and the means for intercommunication between them can scarcely be overstressed. It will be most interesting, as the project develops, to see if the methods planned for the management of the project will prove adequate for such a complex undertaking involving a number of Nations.

Acknowledgement

The Author would like to record his appreciation of the part played by the following people in planning the COST 11 project:-

Mr. E. Toftgaard-Hansen	-	Denmark
Mr. J. Desfosses	-	EEC Secretariat
Mr. J. Pire	-	EURATOM
Mr. M. Deparis	-	France
Mr. R. Despres	-	France
Mr. M. Elie	-	France
Mr. P. Faurré	-	France
Mr. L. Krüger	-	Germany
Mr. W. Weisser	-	Germany
Mr. F. Höld	-	Holland
Mr. N. Maayen	-	Holland
Dr. P. Masarati	-	Italy
Mr. A. Casoria	-	Italy.

References

Proceedings of IFIPS Conference Edinburgh, August 1968

1. The principles of a data communication network for computers and remote peripherals. D.W. Davies.
2. The design of a message switching centre for a digital communication network. R.A. Scantlebury, P.T. Wilkinson and K.A. Bartlett.
3. Transmission control in a local data network. K.A. Bartlett.
4. The control functions in a local data network. P.T. Wilkinson and R.A. Scantlebury.

Proceedings of AFIPS Spring Joint Computer Conference: Vol. 30, 1970

5. Computer network development to achieve resources sharing. Lawrence G. Roberts and Barry D. Wessler.
6. Topological considerations in the design of the ARPA computer network. H. Frank, I.T. Frisch and W. Chou.
7. Analytical and simulation methods in computer network design. Leonard Kleinrock.
8. Host-Host communication protocal in the ARPA network. C. Stephen Carr, Stephen D. Crocker and Vinton G. Cerf.
9. The interface message processor for the ARPA computer network. F.E. Heart, R.E. Kahn, S.M. Ornstein, W.R. Growther and D.C. Walden.

Proceedings of SIGCOM Symposium, Pine Mountain, October 1969

10. Experience with the use of the B.S. interface in computer peripherals and communication systems. D.L.A. Barber.
11. A model for the local area of a data communication network-objectives and hardware organization. R.A. Scantlebury.
12. A model for the local area of a data communication network software organization. P.T. Wilkinson.

Proceedings of Conference on Laboratory Automation, Novosibirsk, October 1970

13. The NPL Data network. D.L.A. Barber and D.W. Davies.

Proceedings of SIGCOM Symposium, Palo Alto, October 1971

14. The design of a Switching System to allow remote access to computer services by other computers and terminal devices. R.A. Scantlebury and P.T. Wilkinson.
15. The Control of Congestion in Packet-Switching Networks. D.W. Davies.
16. Operating Experience with the NPL Network. D.L.A. Barber and K.A. Bartlett. ACM Workshop, L'IRIA, Paris, March 1972.

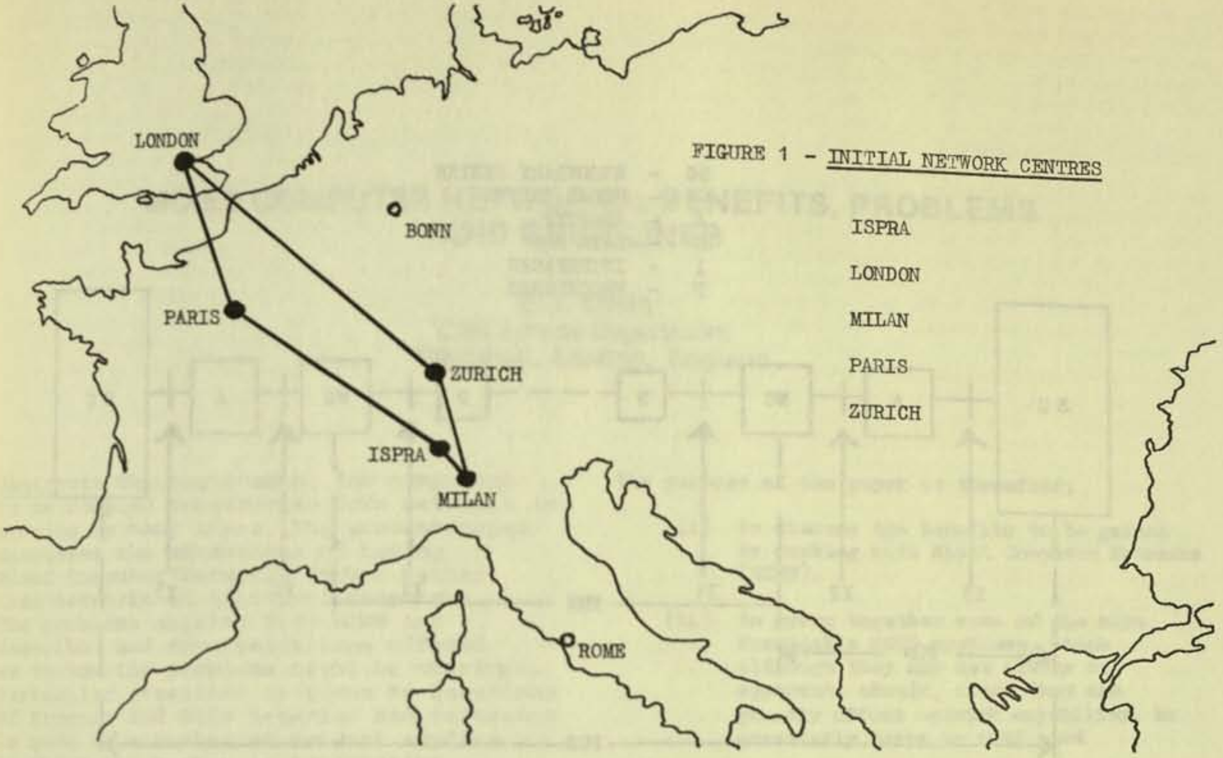


FIGURE 1 - INITIAL NETWORK CENTRES

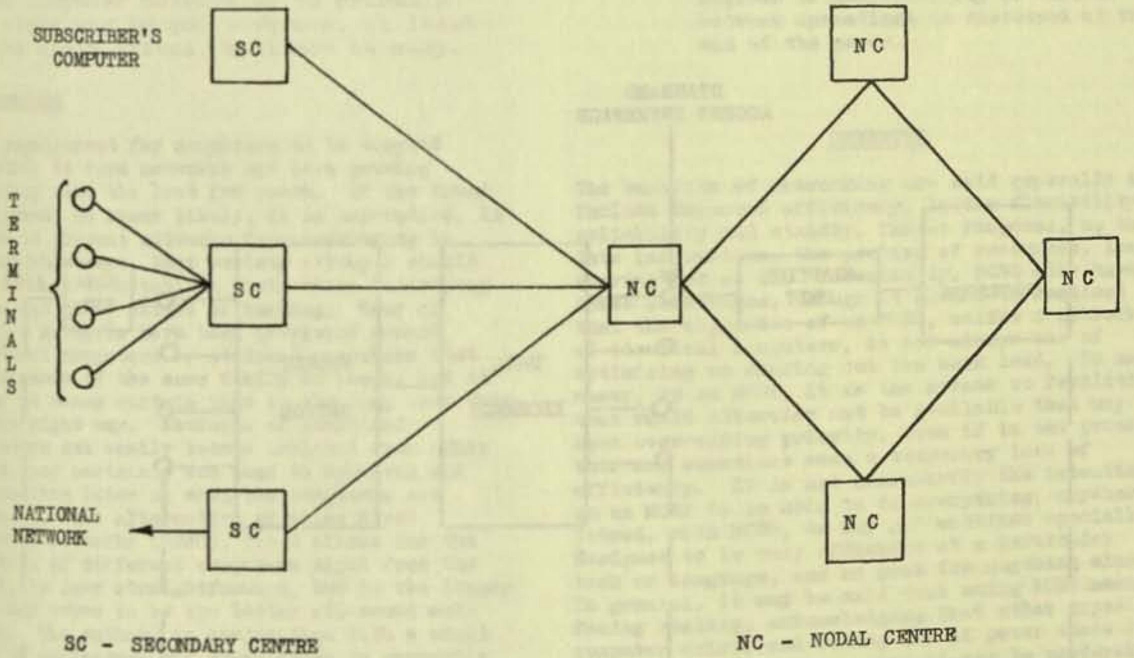


FIGURE 2

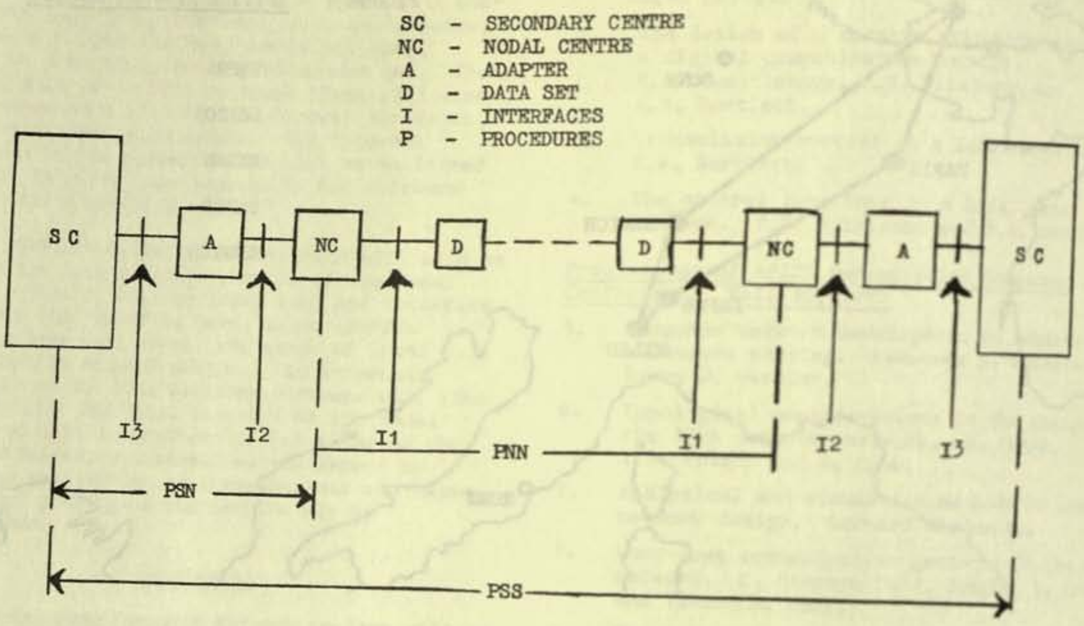


FIGURE 4

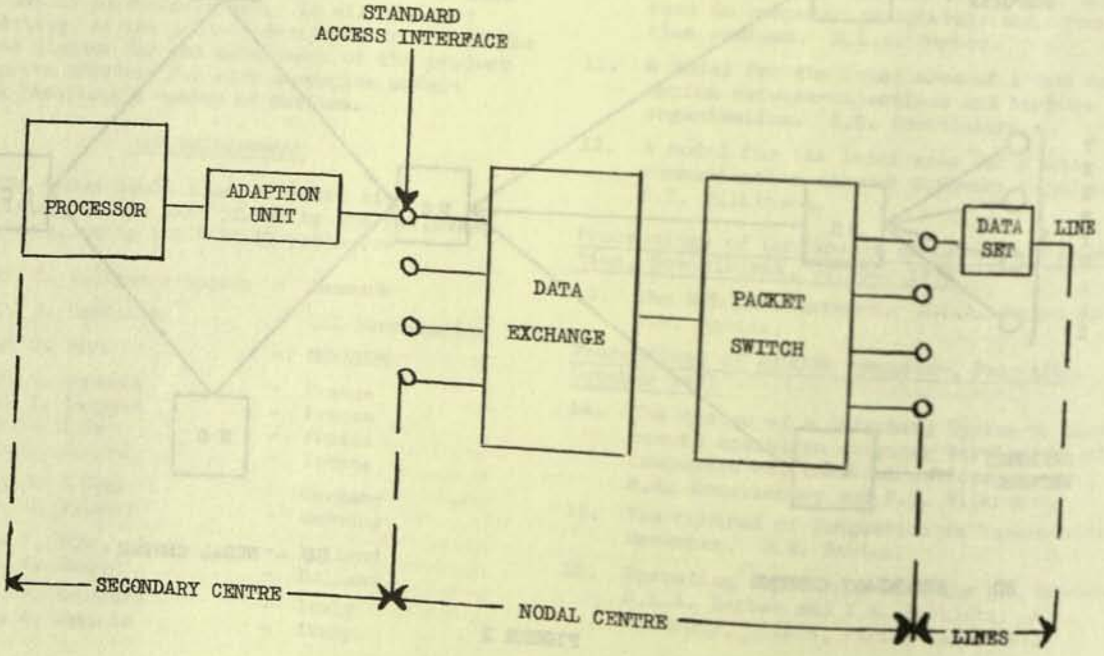


FIGURE 3

MIXED COMPUTER NETWORKS: BENEFITS, PROBLEMS AND GUIDELINES

B. T. Smith
Civil Service Department
Whitehall, London, England

Abstract: The requirement for computers to be coupled together to form networks is growing in many areas. The present paper discusses the advantages of having Mixed Computer Networks (MCNS) rather than networks of identical computers. The problems arising from MCNS are described and some guidelines offered as to how the problems might be overcome. Particular attention is given to questions of Program and Data transfer and reference is made to a number of network studies in which the author in conjunction with consultants is currently engaged. A table for indicating degrees of network achievability is illustrated. The paper concludes with the view that, although mixed computer networking is probably the right way to go, progress, at least in the early states, will not be easy.

Background

The requirement for computers to be coupled together to form networks has been growing steadily over the last few years. If the trend continues, as seems likely, it is imperative, in order to prevent networks from developing in undesirable ways, that serious attempts should now be made to establish their basic underlying facts and their axioms of working. Many of today's networks have been developed around identical computers or at least computers that are members of the same family or range, but it is by no means certain that in the long term this is the right way. Networks of identical computers can easily become isolated from other users, and certainly can lead to upheaval and frustration later on when the computers are replaced. The alternative of using Mixed Computer Networks (MCNS), which allows for the addition of different computers right from the start, is less straightforward, but in the longer term may prove to be the better all-round solution. The author, in conjunction with a small team of colleagues and consultants is currently studying the technical considerations of MCNS, and it is this aspect of networking with which the present paper is concerned. Although the studies are still in progress, it is thought that our approach to the problem may also be of interest to others.

The purpose of the paper is threefold;

(i) to discuss the benefits to be gained by working with Mixed Computer Networks (MCNS).

(ii) to group together some of the more formidable MCNS problems, which although they may not always be apparent, should, since they can greatly affect network capability, be constantly borne in mind, and

(iii) in the light of the team's current studies to put forward a few early guidelines for the would-be designer. A feasibility chart for showing the degrees of achievability of various network operations is discussed at the end of the paper.

BENEFITS

The benefits of networking are said generally to include improved efficiency, better flexibility, reliability and standby, faster response, up to date information, the pooling of resources, load-sharing and so on. Potentially, MCNS also have these advantages, though it should be realised that the objective of an MCNS, unlike a network of identical computers, is not always one of optimising or sharing out the work load. In many cases, in an MCNS, it is the access to facilities that would otherwise not be available that may have over-riding priority, even if in the process this may sometimes mean a temporary loss of efficiency. It is not necessarily the intention in an MCNS to be able to do everything, anywhere. Indeed, with MCNS, we may get machines specially designed to be very efficient at a particular task or language, and no good for anything else. In general, it may be said that using MCNS means facing reality, acknowledging that other types of computer exist, and making use of power where it is available. In some cases it may be preferable to transfer programs to the data; in other cases to transfer data to the program. Even more than this, the use of MCNS should mean a much better chance of keeping pace with technology, better opportunities for re-organising work, for sharing resources and facilities, and more importantly, since network standards are likely to become computer standards, an opportunity to rationalise computing itself. These and other topics are more fully treated below:

Linking different users:

As long as manufacturers continue to produce different computers there will presumably be a need for MCNS. Even if the market place were reduced to a single supplier, and in the interest of everyone, it is hoped that this will never be, there would still be a variety of models, ranges, and special purpose machines that would need to be connected. In theory it should be possible to link in much more easily with other centres and organisations for example with the banks, major customers and suppliers. If properly designed, a network should be connectable, either directly or via a sub-network, to anyone, or any equipment, anywhere, with only minimal adjustment.

Keeping pace with technology:

Computer users often find themselves unable to take advantage of new ideas and developments in the industry because of the cost of changing on-going systems. In an MCNS this can be much less of a worry. The problems of replacement, premature obsolescence, the phasing in and out of equipment, the consequent disruption to operations, and even the problems of implementing a system for the first time, are very much reduced. Problems associated with the forecasting and planning of computer requirements are considerably simplified. Users can try out new configurations without first having to buy a new system.

Use of Specialised centres:

With MCNS, specialised computing services are much more attractive. The use of certain machines for specific functions such as graphics, text editing, matrix inversion etc, should become widespread, and the use of separate computers such as COBOL, FORTRAN, PL/I machines can be expected. This is an example of how MCNS can make use of the best power available.

Data mobility, easier re-organisation of work.

In an MCNS it should be easier to move data between installations irrespective of the computers they have. Alternatively, merely by changing its ownership, it may not be necessary to move the data at all. Because of this, organisations should find it easier to re-group, to respond to changes in legislation, or to other events. Staffing should be less of a problem since these can be concentrated at a single centre or dispersed throughout several, whichever is preferable at the time.

Better use of common services, stand-by, merged data:

Other uses are not hard to find. More widespread use of common services; more alternatives to choose from; better stand-by in case of breakdowns, and the use of merged data from different sources for across the board sampling, modelling and simulation. The possibilities of truly inter-connected computer power are endless.

More rational computer development:

From a purely computing viewpoint, probably the most important benefit from MCNS is the useful rationalising influence that they can have on computing itself. Many of today's big computing problems arise out of the irrational, uncontrolled developments that have taken place in hardware, operating systems, language, data structures and software-writing generally. These are obstacles not only to networking but to computing efficiency per se. Some of the beneficial influences which MCNS users might expect are:-

Better co-operation, better standards:

Whether users are engaged on similar or very different tasks, they will feel encouraged to greater co-operation. The arguments for staying aloof, and working in isolation, will diminish. People will see that it is in their interest to conform to standards and will make conscious efforts to do so. Standards should come naturally and not by edict.

More use of existing systems and an end to one-off software:

In Government, as elsewhere, people are always too ready to develop their own system. Invariably, the reasons given are that, circumstances are different, "our problem is unique, compromise would be too difficult or costly - when possibly the real reason may be that it is too difficult to access an existing system, or to find out more about it. MCNS should therefore see an end to the costly business of duplicated software, the continual re-inventing of the wheel every time a new program is written. In the UK, as in other countries, one still encounters the so-called NIH (Not Invented Here) factor, the desire to stamp one's own name on a product, but in a healthy MCNS environment this should be much less of a nuisance. Programs will become more readily exchangeable and acceptable; in fact users can be expected to look to the network to tell them of other people's products.

More influence on manufacturers:

Dr Herb Grosch of the National Bureau of Standards has said many times that it is a fallacy to believe that manufacturers are in business to save their users money, and this is probably true. As in the case of time-sharing and large computer bureaux, which manufacturers only reluctantly accepted, so with networking. Manufacturers obviously would prefer to deal with customers singly rather than jointly, as they will probably have to do as network users come together and, as a lobbying force, become more powerful. If networks are seen to be inevitable by manufacturers, then the advice they will undoubtedly give is that the computers should all be of one kind - naturally their own. Users on the other hand should take the lead and, if their needs are more general, should not hesitate to connect different computers. In so doing they will induce manufacturers to do what is right for the community as a whole.

PROBLEMS

The principal problem with MCNS from which probably all others arise is that of compatibility, how to bring together into a working harmony machines of possibly very different architecture and structure, which in turn, lead to important differences in operating systems, languages and software generally. It is not of course suggested that each problem should receive equal treatment or that compatibility must be obtained at all costs. Some of the problems encountered are intransigent and it would be foolish to make machines or programs compatible just for the sake of doing so. One of the important advantages of MCNS as indicated earlier is to move the work to where it can best be processed. It is necessary however to identify the problems if only to learn to live with them and minimise their effect. Ultimately it is hoped - and this is also expressed elsewhere in the paper - that users will bring pressure to bear in the right quarter to obtain desirable standards. Some of the important problems which MCNS must contend with are discussed below.

Hardware:

Differences in main processors are concerned mostly with control mechanisms; input, output, and interrupt facilities, different word lengths, byte sizes, character sets and differences in speed and overall performance. There can also be serious differences between peripherals. Magnetic tapes, for example, which in theory should present fewer problems can often be troublesome because of their tape labelling methods or the different ways in which their beginning and end tape marks are sensed. Physical interchange of peripherals (mag tape, disks, printers etc) is often quite impossible. Resort to the logical or software level ie via an MCNS might well be the only way that exchange is possible at all. This of course is an attractive by-product of MCNS.

Operating Systems:

Operating systems differ widely from machine to machine. Some cater for little more than input/output and a certain amount of error checking and interrupt. Others, which have taken literally thousands of man years to write, can administer to the wants of large multi-programming time-sharing and inter-active systems. Particularly important from a network viewpoint, are systems which allow full job and file definition to be made at run time - or even dynamically during the running of another program, features which require the operating system to service long job queues, schedule workloads, and allocate resources on demand. Added to these facilities, there are usually extensive libraries of macros, sub-routines and catalogued procedures. The latter especially are vital in minimising job set-up time. The matching of the different facilities that are offered, in a network of different computers is a difficult problem, especially if it is required, as it should be, to avoid doing work twice, once for the network and once for the Operating System.

Job control:

Job Control Language which closely reflects the facilities offered by the Operating System is also a problem. As increasingly more functions are, quite rightly, taken away from the programmer's control and placed in the care of the Operating System, more and more ambiguities appear. As a result it is not uncommon to find a single command term used for two different functions, and two command terms used for a single function. The effect of this on a network of mixed computers is not hard to imagine.

Language:

The myth that programs written in a high-level language are immediately transferable between machines has long since been exploded. At least one case is known, where in the transfer of a FORTRAN program, there was only one statement, CONTINUE AT, which did not require change. Not all cases are this bad, but it is doubtful if changes of this nature will ever be made entirely automatically. Even if it were possible it can by no means be guaranteed that a program on two different machines will produce identical results. It is also true that some languages are better suited to some machines than others, for example PL/I on System 360 and ALGOL type languages on Burroughs machines. Features that may be extremely efficient on one machine may be deplorably bad on another and in some cases not even permitted. Although some improvements are possible in this area, it is also an area in which great care should be taken not to waste time on impossible tasks.

Data organisation:

The problems here are the many ways in which data can be stored. Files can range from simple serial type files, to inverted files, tree-type and list-type structures, to linked hierarchies, schemas, sub-schemas and so on. When it is realised that data in these forms may be distributed over devices of varying capacity and dimension (bytes per track, tracks per cylinder etc), it can be seen that direct transfer of data in an MCNS can be very difficult indeed. Furthermore as the information explosion continues and data is stored in mammoth libraries, it is unlikely that any but the most simple data structure will suffice, even for storage and access, let alone transfer. In general, the more complex the data structure, the more difficult it becomes to maintain and handle it in a network. A possible way out of the problem is suggested below under Data Transfer, Guidelines for a Solution.

Internal data representation:

The problems here are those of different character sets, byte sizes, and the methods of storing numbers, integers, floating point etc. Numbers are particularly tiresome not only because of the various ways they can be stored or packed, but because of the many degrees of precision that are available.

Collating sequence:

This is the order in which a given machine assigns binary values to characters. In some machines letters appear before figures; in others, figures before letters; and usually special characters may appear anywhere. If the sequence of two machines is different, it may be necessary before processing to substitute different values, and possibly to sort the data for internal purposes and then substitute and re-sort on output. If there is not a one-for-one correspondence between the sequences, there are other problems.

Terminology:

A not insignificant obstacle to communication on computers, and hence to networking, is terminology. Ambiguities and differences in meaning are prevalent in most areas and not least in the documentation of different systems. What for example are the subtle differences between;

ADP and EDP;
Processors, computers, and machines;
Operating systems, monitors and supervisors;
Programs, processes, and procedures;
Data bases, data sets, and files;
Data items, elements, and fields;
Tags, flags, and keys;
On-line, in-line, and real-time;
Time-sharing, multi-programming, multi-tasking, and multi-processing;

and many many more?

Even if we have a clear idea in our own minds as to the distinction between the terms, can we be sure that others think the same? It is to be hoped, in this area especially, that by closer user working better standards in terminology will eventually emerge.

GUIDELINES FOR A SOLUTION

An essential requirement in a task of this magnitude is an initial period of abstraction and conceptual analysis in which ideas can be examined and the problem broken down into separate logical parts. If this is done, then any system that follows will have a correct modular construction (so that its parts may be easily changed or removed), and will be measurable against a total model (so that all its parts can be seen to fit). Both Dijkstra's onion skin approach to operating systems,[1] and the theory of a hierarchy of nested layers of networking put forward by the National Physical Laboratory (NPL) in England,[2] support this view. The latter it is believed is an essential ingredient to network philosophy and is referred to again later in the paper.

The current studies

The above approach is the one adopted by the team in its current studies. The work at this stage is quite conceptual. No attempt is being made to design a system or take into account the requirements of a particular installation. To do so, would inhibit investigation and mean that some problems would not be treated in the depth that they should. The only constraints placed on the studies are those which by definition apply to all MCNS:-

(i) that in any such network it must be possible to link in independently designed systems and hence computers of different manufacture, and

(ii) that, by belonging to the network, the autonomy of any installation must not be impaired; in other words, if it wants and has the authority, the installation should be capable of detaching itself painlessly from the network and resuming its stand alone operation.

Partitioning the problem:

It was decided to look at the problem from the viewpoint of three observers situated at:-

(i) the main operating system of a local installation belonging to the network,

(ii) at some point in the middle of the network theoretically responsible for all the problems of system interface and network control, and

(iii) at the location of the ordinary user at the end of it all, seated at a remote typewriter or VDU.

Some overlap of responsibility is likely as each observer looks further into the distance, but this is intended in order to get a broader view over critical areas.

A Network Model

A simple model of the team's approach is given in Fig 1. The Roman numerals in the diagram refer to the three studies.

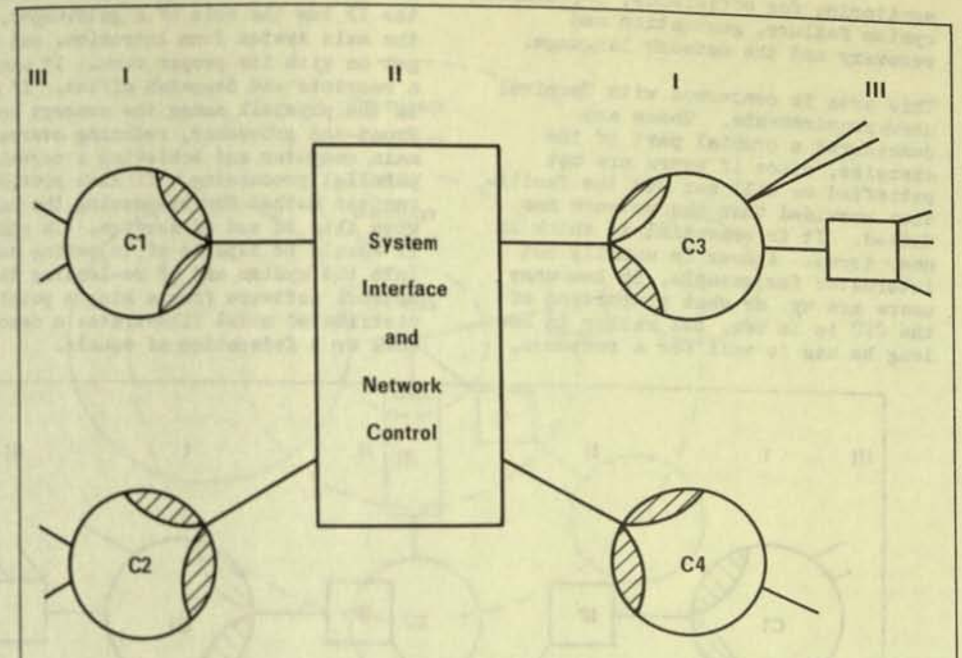


Fig. 1 Network Model

Study I : This area is concerned with the minimum changes to be made to local operating systems to enable them to take part in a network. The shaded portions of the circles, referred to as middleware or pseudo-programs, contain the additional software required, programs for dealing with new types of interrupt, additional scheduling, instrumentation, error handling, system awareness, off-loading of work etc. Hopefully these can run as normal programs with the system hardly aware that it is joined to a network. To detach itself from the network, all that the installation needs to do is to unload the programs.

Study II : This area is concerned with the mechanism of getting across from one machine to another, that is with all the problems of synchronisation, different character sets, data formats and languages and the interchange of programs, data and systems, but excluding pure communications, the area itself being expected to adapt to whatever communications are provided.

The area can be visualised as a large interface turntable on which is placed all the information necessary for the interfacing and bridging of different machines. Notionally all the various parameters and attributes of the machines connected to the network would be stored in a series of look-up tables and lists. From these, whenever the transfer of data from one machine to another is required, an appropriate conversion routine extracts the information it needs and makes the necessary transformation. A message might join the turntable in a UNIVAC format and leave it say in an ICL or IBM format. As new computers and systems join the network new entries are inserted in the tables; old ones may be deleted or changed. The tables would also provide details of network status, availability and connectivity times, user addresses and so forth. The description is a simplification of the process which actually takes place but it illustrates the method of getting onto the turntable in one form, going through a conversion process triggered off by signals in the protocol and leaving it later for a different machine.

The second part of the study (control) is concerned with linkage protocol, monitoring for efficiency, performance, system failure, generation and recovery and the network language.

Study III : This area is concerned with Terminal users requirements. These are considered a crucial part of the exercise, since if users are not satisfied or will not use the facilities provided then the network has failed. It is essential to think in user terms. A user is usually not interested for example, in how many users are up, or what proportion of the CPU is in use, but rather in how long he has to wait for a response.

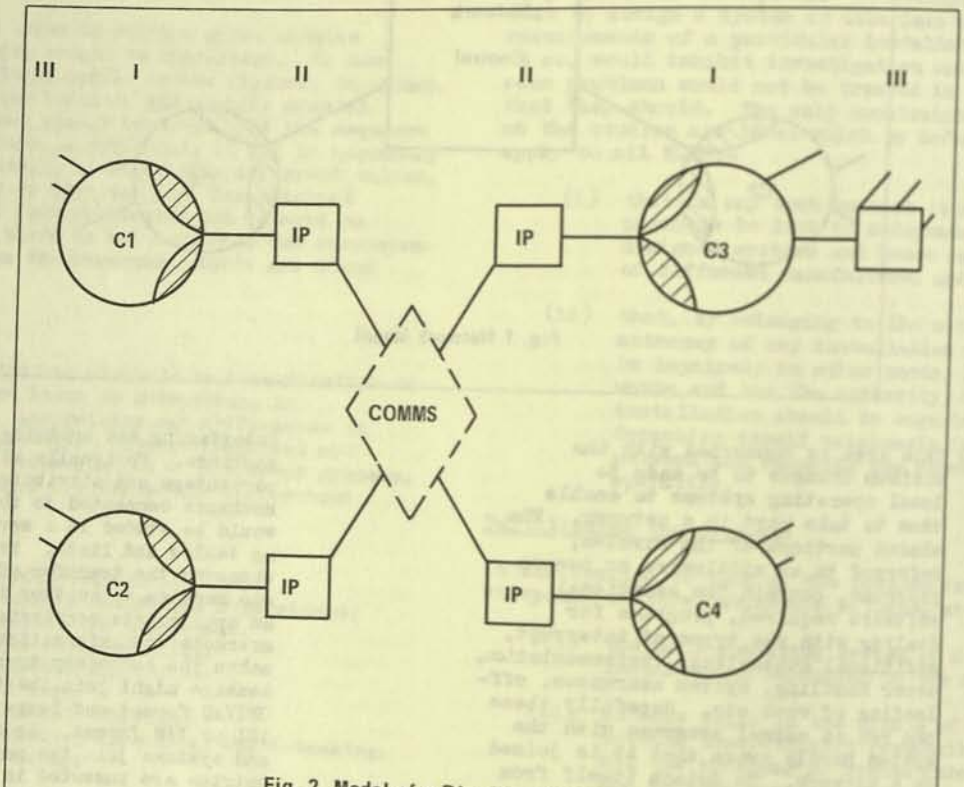


Fig. 2 Model of a Distributed Network

Distributed Model:

The Fig 1 model shows a network with one interface area only and this could in fact be a configuration with the Study II block equating to some actual processor. It is however an easy matter, by "distributing" the block to produce a distributed model as shown in Fig 2. The distributed areas are labelled IPs (Interface Processors) after the style of ARPA IMPs (Interface Message Processors). The function is not however the same as ARPA^3, since communications, which have a major role in ARPA IMPs, are not part of the current study. Conceptually the IP has the role of a gatekeeper, protecting the main system from intrusion, and letting it get on with its proper work. It also acts as a receipts and despatch office. If extended in the physical sense the concept becomes a Front-end processor, reducing overheads in the main computer and achieving a certain amount of parallel processing. It also provides a convenient method for by-passing the main machine when this is out of service. In addition, any IP should be capable of injecting modifications into the system and of re-loading the entire network software from a single point. The distributed model illustrates a democratic network or a federation of equals.

Design matters:

The two models described are of course conceptual. When it comes to designing a system, it is likely that, owing to constraints in the real world, parts studied in one area may have to be moved to another. It should however be possible to keep the different functions logically separate. This is the significance of the NPL scheme of layers of interface referred to earlier and illustrated in Fig 3.

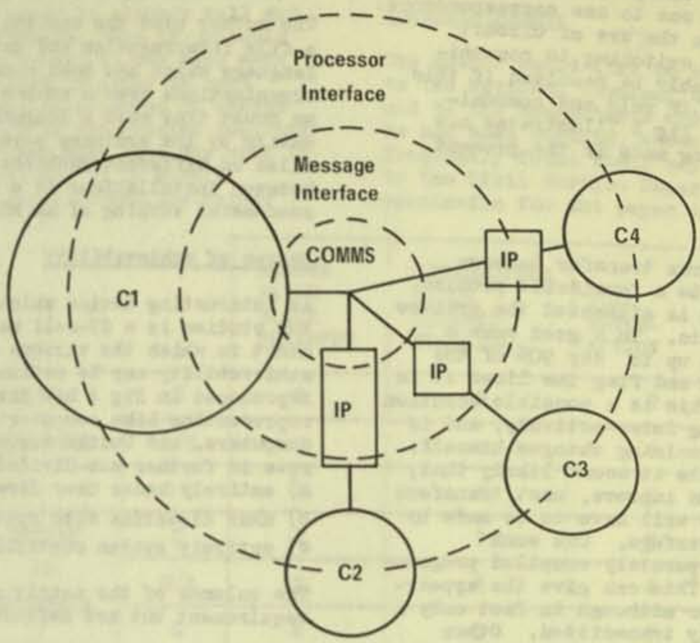


Fig. 3 Main Layers of Interface

Layers of interface (Fig 3):

The diagram shows three concentric layers of the network function. The innermost layer corresponds to the pure communications function, while the outer layers represent network software and user processes. It can be seen that a single processing unit, whether this is a main computer (C1, C2, C3 or C4), a Front-end processor, an IP, or a communications machine (perhaps an ARPA IMP) may straddle several layers of interface or may be completely enclosed by one. This in fact will always be the case, as different machines are called on to perform different as well as multiple functions. The important thing is that the logical blocks or layers are easily identifiable, a) so that they can communicate directly with corresponding layers in other machines, and b) so that their contents may be changed or replaced without affecting the surrounding layers. Communication between layers is by means of a defined interface (eg IP to HOST protocol) and within layers by agreed linking procedures (eg HOST to HOST protocol). There are of course many more functional layers than are shown in the diagram and the isolation of these is an important part of the study. One particular example of this which has proved very useful in Study I is the so-called Network Support Package which has been found to have three functional layers which can perhaps best be described by analogy with the telephone system:-

LCM (Line Control Module): Hardware privileged	equating to the telephone wires leading to a telephone
NCM (Network Control Module): Software privileged code	equating to the dialling mechanism, and
UCM (User Control Module): Unprivileged code	equating to the message or language passed.

Until these layers were identified it had been very difficult to outline the program specifications for this area in any very meaningful way.

Data Processing v Communications

One of the commonest errors in network design is to fail to distinguish sufficiently clearly the function of the data processor and the communicator. It is always tempting especially when using store and forward or message switching techniques to expect the communicator to do code conversion, encipherment, re-formatting, and other data manipulations which more properly belong in

DP. Immediately that changes are made there is trouble, especially for example, when codes are used that do not have a one to one correspondence. Many of the arguments on the use of circuit switching versus packet switching in communications would also probably be resolved if this distinction between the DP role and communications were observed. Fig 2 illustrates how this distinction is being made in the present studies.

Program transfer

As stated earlier, program transfer between different machines can be a formidable problem. In those cases where it is attempted the process is seldom fully automatic. In a good case a filter program can make up to say 90% of the required syntax changes and flag the lines it is unable to deal with. This is a possible solution when the user is working inter-actively, and is prepared to make the remaining changes himself. If this is not acceptable it seems likely that, until language standards improve, most transfers of source code programs will have to be made by ad hoc methods and subterfuge. One such stratagem is to have separately compiled programs at each installation. This can give the appearance of program transfer although in fact only a program name is possibly transmitted. Other methods along these lines are possible. In the future there is also the prospect of program emulation by micro-software. In this case it would however be object programs that were transferred rather than source programs.

Associated with program transfer is the problem of different job control mentioned earlier. As with the transfer of programs themselves, the temporary solution here appears to be to have suitable JCL blocks previously prepared and waiting at relevant installations.

Data transfer

The solution to this problem and those of data organisation are believed to lie in the holding of data in a serial, book-like form (the so-called book concept). In the present state of the art it is of course necessary to hold and use data in other forms as well, particularly for real-time and immediate response systems. The only difference from present practice is that these forms should be considered working files or application files, not data bases. The true data base of an installation should be in book form, and updated periodically or at the end of a working day. If the concept is adhered to, the data exchange problem largely disappears. Data can be referenced using the common standard terms for a book, chapter, paragraph, sentence etc or by equivalent terms and exchanged even more freely (that is in the form of extracts etc) than its real-life counterpart. It can be shown that any collection of data can be represented in this image, and as with ordinary books, there need be no limitations on size of the book, its pages, chapters and so on, or on its contents. This conceptual vehicle is well understood and available and should be used to give us the conformity we need in this vital area.

The author used the concept when writing SPECOL[4], a file interrogation and data manipulation language which has been successfully used by some organisations over a number of years. There is no doubt that such a language which can be used easily by the ordinary person to interrogate files on different machines and to exchange data between installations is a necessity to the successful working of an MCNS.

Degree of achievability

An interesting device which has been of use in the studies is a 27-cell matrix or feasibility chart in which the various degrees of network achievability may be estimated. The chart reproduced in Fig 4 has three basic rows representing Like computers, A family of computers, and Unlike computers. Each of the rows is further sub-divided to denote operations
a) entirely under user directions,
b) user direction with system help and
c) entirely system controlled.

The columns of the matrix may vary according to requirement but are defined here as

(i) Remote Batch working including simple block data transfer,

(ii) Inter-active batch working with dynamic file access and transfer, and

(iii) Fully interactive working in the full sense of the word. Scores for each of the 27 operations have been awarded in the scale 1 to 9. These should not however be taken too seriously, representing only the author's opinion on the state of the art at the beginning of the study. The diagonal pattern in the figure which runs from the top right of the matrix to the bottom left is interesting if also possibly somewhat contrived. Since the scoring is a purely subjective exercise the chart must by no means be taken as a categorical reference of feasibility, but it has proved useful in analysing the problem and in evaluating where greater effort is perhaps worthwhile or in some cases where the situation is better left alone.

Conclusions.

At this stage of the studies while they are still in progress it is not possible to say what the outcome of them will be. Certainly for the industry as a whole, the first impressions are that MCNS are feasible, though it may be some time before they are fully developed and as flexible as one would like. The price for them - the price of overcoming incompatibility - is not cheap, and for those critics that complain, with some justification, that too much computer power is already well and truly silted up with software, this is probably more evidence for their cause. It may also seem something of a paradox that in order to beat incompatibility, one must be prepared to accept and condone it. But such is the complex behaviour of people and institutions, that this may well be the only way. If this is so, it is surely sensible to face up to our problems now rather than later.

Acknowledgments

The author would like to express his appreciation to the National Physical Laboratory in England and to the consultants and colleagues with whom he has been working, whose ideas have perhaps frequently found their way into the paper, and to the Civil Service Department for giving permission for the paper to be published.

		REMOTE BATCH BLOCK TRANSFER	INTERACTIVE BATCH DYNAMIC TRANSFER	FULL INTER-ACTION
LIKE COMPUTERS	U	9	9	7
	U/S	9	8	6
	S	8	7	5
FAMILY OF COMPUTERS	U	8	7	5
	U/S	7	6	4
	S	6	5	3
UNLIKE COMPUTERS	U	4	3	2
	U/S	3	2	1
	S	1	1	1

Degree of Control
U - Entirely user direction
U/S - User with system help
S - Entirely system controlled

Scale of achievability: 1 to 9

Fig 4. Network Feasibility

References

1 DIJKSTRA E W : The structure of the 'THE' - multiprogramming system. CACM Vol II No 5 (May 1968) pp 341-346.

2 NATIONAL PHYSICAL LABORATORY, ENGLAND : Design of a switching system to allow remote access to computer services by other computers and terminal devices. R A Scantlebury and P T Wilkinson. Proceedings of SIGCOM Symposium, Palo Alto, Oct 1971.

3 ARPA NETWORK, USA : AFIPS Proceedings of the 1970 SJCC pp 543-597.

4 SMITH B T : SPECOL - A computer enquiry language for the non-programmer. Computer Journal Vol 11 No 2 Aug 68. Developments in SPECOL - a retrieval language for the non-programmer. ibid Vol 13, No 1 Feb 70.

COMPUTER COMMUNICATION NETWORKS FOR HIGHER EDUCATION

CHAIRMAN:

Professor Martin Greenberger, Department of Computer
Science, Johns Hopkins University, Baltimore,
Maryland, U.S.A. and EDUCOM, Princeton, New Jersey, U.S.A.

PROGRAM COORDINATOR:

Stuart L. Mathison
Arthur D. Little, Inc.
Cambridge, Massachusetts, U.S.A.

PAPERS:

Professor Martin Greenberger, *"Computer Communication Networks for Higher Education"*

R. H. Howell, International Computers Limited, ICNS Project, University of Bristol, Bristol, England, *"The Integrated Computer Network System"*

John deMercado, René Guindon, John DaSilva, and Michel Kadoch, Department of Communications, Ottawa, Ontario, Canada, *"The Canadian Universities Computer Network–Topological Considerations"*

D. D. Aufenkamp, Office of Computing Activities and E. C. Weiss, Office of Science Information Service, National Science Foundation, Washington, D.C., U.S.A., *"NSF Activities Related to a National Science Computer Network"*

DISCUSSANTS:

Professor Eric G. Manning, Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada

Professor Norman Abramson, Information Sciences Program, Univeristy of Hawaii, Honolulu, Hawaii, U.S.A.

THEME: The motivation for regional and national computer networks for higher education will be discussed along with a summary of the major networking efforts underway. Included will be the ARPA network, the National Science Foundation Network, and several presently operating regional networks. Other topics to be covered include network management, usage of networks for computing and instruction, research applications, implications for computer centers in higher education, and future trends.

COMPUTER COMMUNICATION NETWORKS FOR HIGHER EDUCATION

Martin Greenberger
The Johns Hopkins University
Baltimore, Md., U.S.A.

University computer networks are here. They are no longer the wishful thinking of visionaries in the early 1960's or the unmet goals of technologists later in the sixties. Several large networks are now operating and more are on their way. The technology, the incentives, and the know-how have reached the necessary threshold level.

Interesting network projects are underway in many parts of the world. The Universities of South West England are developing a network and so are the Canadian universities. In the United States the Advanced Research Projects Agency of the Department of Defense has a national network operating among many American universities and other research institutions, which also avail themselves of the services of several large time-sharing networks run commercially. The National Science Foundation is mounting an expanded research program to explore possibilities for a National Science Network linking computer facilities and users at educational and research institutions Throughout the country. And computer users in many fields and disciplines are getting together to discuss the possibilities and plan for their part in the project.

What this accelerating activity will lead to is not yet entirely clear. The role that computer networks will play in higher education can be major, or it could conceivably pass in the wind. Much will depend on how the networks are organized and managed and how their users are served. Economic considerations will be paramount.

The following discussion will be concerned with these all-important organizational, management, user, and economic considerations, as well as the technical issues, in the context of several specific case studies here and abroad.

THE INTEGRATED COMPUTER NETWORK SYSTEM

R. H. Howell, International Computers, Ltd.
ICNS Project, University of Bristol
Bristol, England

Abstract

The purpose of this paper is to consider the reasons behind the implementation of a network for the Universities of South West England and South Wales, to describe the overall system as it now exists, and to detail the facilities provided.

The two main reasons behind the project were:-

(i) To improve the computing facilities of the South West Universities.

(ii) To perform research into networks.

The first point can be sub-divided into the various benefits to be gained from the network, over 'stand alone' processing.

(i) To allow all network users to access any facility available on any machine within the network.

(ii) To provide a means to balance the work load of the machines by 'off-loading' jobs from an 'over-loaded' machine to an 'under-loaded' machine.

(iii) To permit access, to all users, to all public data in the network.

(iv) To eliminate the modifications of software to fit individual machine configurations, ie. one machine would hold that particular software and this would be used by the users at 'remote' machines.

(v) To effectively upgrade the smaller machines within the network to the processor with the fastest/largest store within the network.

Networks are still very much in their infancy and the more research into them then the greater will be the expertise in areas such as:-

(i) Dynamic scheduling within a network.

(ii) Job control in a network.

(iii) Transmission control.

(iv) Connection of 'Local' terminals to a 'remote' machine to access 'remote' facilities.

(v) File control of files copied around the network.

(vi) Development of a central file store.

(vii) Fail safe requirements of a network.

When operational, experience can be gained in managerial problems within the following fields:-

(i) Inter-establishment liaison.

(ii) Budgetary control.

(iii) Operator problems - since no single shift of operators has complete control of the environment.

(iv) Fault diagnosis and control.

Introduction

A team of programmers from ICL and the universities involved and one programmer from the Post Office started the project in the summer of 1969. The network is directed by ICL and jointly supported, financially, by ICL and the Department of Trade and Industry. The network environment is provided by the Universities of Bristol, Bath, Exeter and the University College of South Wales and Monmouthshire and the University of Wales Institute of Science and Technology. As the last two colleges share computing facilities the network is comprised of four sites (fig 1).

It was decided to use the Multijob Operating System, produced by ICL as the base for network software. This is a standard Operating System for the System 4 range of machines and is in use at all four sites and provides facilities for both batch and multi-access working.

The hardware of the network consists of a System 4/75 at Bristol with 768K bytes of store, 3 exchangeable disc drives of 7¼ M byte capacity each, a 9 spindle exchangeable disc store of 29 M bytes, a 3 spindle exchangeable disc store and 2 x 2 M byte drums plus the usual slow peripherals. There is a System 4/70 at Cardiff and System 4/50's at Bath and Exeter, all with 256K bytes of store, 3 - 4 exchangeable disc drives and the normal slow devices. Each main processor is linked to a CDC 1700 to form a star network (fig 1), by 48K baud communications lines used in full duplex mode. The CDC 1700 is used as a communications processor, but also holds the network control software for off-loading, statistics and the indexing of public data.

The likely inclusion of machines of different architectures has been borne in mind by the designers, ensuring that ICNS is not introspective and that the network may be enlarged in the future without too much change to the software.

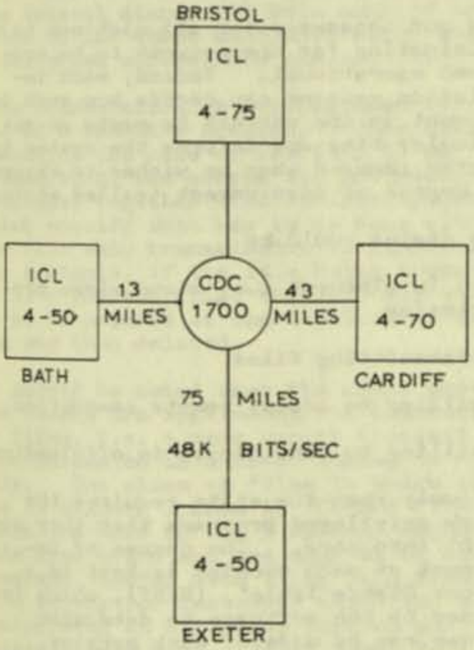


FIG. 1. HARDWARE IN THE NETWORK.

Network Software

Philosophy of Implementation

The two broad aims of the project are:-

(i) to allow access to all facilities available on all machines in the network to all network users, and

(ii) to balance the workload of the installations in the network by off-loading jobs from one machine to another when appropriate.

The project team had to investigate, design and implement the software required to satisfy these aims.

It became clear at an early date that it would be impossible to determine the use which would be made of the network because the universities at that time had no experience of

(i) using and operating a service based on a multi-access operating system,

(ii) the operating system itself and

(iii) the advantages to be gained from a network environment.

For this reason, it was decided that, instead of spending an excessive amount of time arguing about the use, to make some educated guesses and to implement a system as quickly as possible, accepting the fact that changes would have to be made in the light of a study of the actual use made of the system.

Another point which had to be borne in mind during the design of the system was that it was quite likely that machines of other makes and types might be introduced into the network at a later stage. Also, if, as is hoped, the system is a great success, it would be desirable to be able to transfer the ideas to other networks based on different machines.

Thus, for the above reasons, the system software should be designed and written so that:-

(i) changes and enhancements could be made to it easily, as a result of a study of the use made of it, and

(ii) it is as independent of the hardware and controlling operating system as is possible to facilitate the addition of further machines in the system and to ease the problems of transfer to other types of machines if necessary.

Architecture of Software

The structure of the ICNS software within the Multijob operating system falls into three parts:-

(i) additional console service routines to deal with the new commands provided by the system,

(ii) necessary alterations and extensions to existing supervisor routines. These should be as minimal as possible to satisfy the 2nd of the Design criteria, and

(iii) routines which operate in the user area of core. A program written for this area normally operates in a restricted state and is not allowed to access data outside its program area. For ICNS purposes, however, these routines will be allowed to operate without restrictions and are referred to as "Privileged Programs". The bulk of the network software falls into this category.

There are three main reasons why these routines are written as privileged programs rather than as real supervisor routines.

(i) If privileged programs are used, they can be called in when required and conversely, be removed when no longer needed. This means that the network can be initialised or closed down without disruption to the normal operation of Multijob. By occupying core outside the supervisor area, the network may be regarded as expanding and contracting to meet requirements.

(ii) Supervisor routines may be either resident or overlaid. If the network software were entirely resident, the cost in core occupancy would be prohibitive. On the other hand, the Multijob overlay area is effectively limited and it is desirable to use this area as little as possible, otherwise the performance of the normal operating system would be adversely affected.

(iii) Written as a user program means that the network software can be reasonably independent of Multijob, making implementation and support easier since it will be unnecessary to make amendments to each new issue of the operating system and also satisfies the second of the design criteria stated above.

In Multijob, there is no facility for multiprogramming different functions contained in one program and, consequently, two or more programs have to be written containing functions which can multiprogram with functions contained in another program. While ideally the network software would be split into one program per function, this is impracticable because of the amount of core required and because the number of multiprogramming slots taken by the software would not allow any user work to be processed. Thus the number of privileged programs has been limited to two, which will enable a reasonable degree of multiprogramming to be carried out while leaving an adequate number of program slots for general use. The network programs must multiprogram as so much of the network activity is I/O bound, e.g. the transmission of files.

Each privileged program consists of a control segment plus a number of overlaying segments. These overlaying segments are brought into core by the control segment when it is necessary.

Facilities Provided

Initialisation and Closedown

It is not necessary for all machines to be participating for the network to be considered operational. Indeed, each installation manager can decide how much involvement in the network he wants at any particular time and informs the system by operator command when he wishes to change this degree of involvement (called states).

These states could be

1. no involvement i.e. stand alone processing
2. transmitting files
3. willing to accept remote connection
4. willing to participate in off-loading

It is only when the state requires the network privileged programs that they are brought into core. The degree of involvement of each machine is kept in a "Network Status Table", (NEST), which is examined by the software to determine what use can be made of each machine.

Remote Connection

To access facilities available on a remote machine, a user must first "connect" to that machine. This can be accomplished in one of two ways

(i) from an on-line terminal by specifying the required machine in the LOGIN statement. When the connection is effected his local machine becomes 'transparent' and he is treated as any other user of that remote machine, or

(ii) from the batch input stream by specifying the required machine on a //SEND JOB card. The submission of this card causes all the following cards up to a //ENDSEND card to be sent to the specified machine and the job control statements contained on those cards to be inserted in the job queue of the remote machine. This is called "Remote Job Promotion".

File Transmission

An obvious requirement for any network is the ability to transfer files from one machine to another. This requirement is of special importance in the ICNS network because of the existence of the large amount of on-line storage on the Bristol machine. Since the other machines in the network have limited on-line storage (of the order of 22 M bytes), it is likely that the Bristol discs will hold many of the files of the other machines, which are transferred across when needed.

File transmission is accomplished by the use of a TRANSMIT command. This command specifies the name of the file to be transmitted, the source and destination machines, together with various options which specify what has to be done with the file when transmission is complete. For instance, if the file being transmitted is a results file, the user may want it to be printed at the destination machine and then deleted.

It should be noted that the usual protection checks are applied to the transmission of files, i.e. a user cannot transmit a file unless he is allowed access to that file. One class of files to which anyone is allowed access is 'COMMON FILES'. These are held on the 9 spindle exchangeable disc store and are made up of useful programs, useful data files (the results of a survey for instance) which may be manipulated in different ways by different universities, etc. It is hoped that this will save duplication of work in each university.

It is possible in Multijob to 'MARK' blocks of a file for alignment purposes. The file transmission package takes account of these marked blocks and ensures that they are correctly marked on the destination machine.

'MAIL' Facility

It is possible to 'post' a message to any network user. A 'MAILBOX' file exists on each machine which contains messages left for a particular user and also notices applicable to all users (e.g. notices sent by the system manager). When a user logs in to the system, he is informed if there are any notices or messages for him and he is given the choice of having them displayed or not.

Balancing the Workload

To balance the workload of the machines in the region, it is clear that jobs have to be moved, by some means, from an overloaded machine to an underloaded machine. In ICNS, there are two ways in which jobs are moved:-

(i) Remote Job Promotion

(ii) Off-loading

Remote Job Promotion has been described above. It involves the user (or operator) in making a positive decision that a job or jobs should be run on another machine and he inserts special control cards in the Job Control pack specifying the destination machine. The reason for this decision may have nothing to do with the work loads, but may have been made because of some facility on the destination machine which is required by the program. Off-loading will not be implemented in the initial release as it is required to know the problems involved in running a network before the high level scheduler can be finalised, but the following is a brief description of the facility as it is envisaged at the moment.

Off-loading is a system decision, and the user need not be aware that it has happened or that his job was not run on his local machine.

Before off-loading can take place, three decisions have to be made.

1. Which machines should be involved in off-loading
2. Which jobs can be off-loaded.
3. Which jobs should be off-loaded

To determine which machines to involve in off-loading, a measurement of the load on each machine must be made. This loading could be measured in terms of utilisation of CPU, peripherals etc., but, in the eyes of a user a machine is heavily loaded if it takes longer to get his run back than it did yesterday, i.e. he looks at loading in terms of the backlog of work.

This view is taken by the project and the loading of the machine (called the Machine Loading Factor) is measured in terms of the amount of work still waiting to be run. It is a difficult task to measure the Machine Loading Factor (MLF) accurately, since the system has no knowledge of the amount of work waiting in trays in the off-line area. While it may be possible to devise some scheme whereby the system could determine this - for instance, dedicating a slow card reader to the task of reading a control card specifying the running time of each job as it is brought to the off-line area and the system keeping a running total - it is intended to take the easy way out, at least in the initial release, and the system will request the operator to type in the amount of work off-line.

The MLF is represented on a scale called the "dipstick" (fig 2)

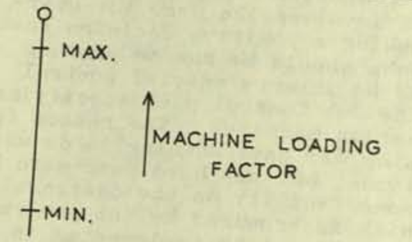


FIG. 2. DIPSTICK.

If the MLF has a value greater than the MAX setting, then the machine is considered to be overloaded, and if the value is less than the MIN setting, then it is considered underloaded. In both cases, the machine is a candidate for involvement in off-loading, one for sending jobs and one for receiving jobs. The values of these settings can be adjusted in the light of operational experience, and to suit local requirements.

It is not possible to move jobs which require the use of raw data, cards, paper tape, magnetic tape etc., so the off-loading is restricted to jobs which access on-line storage only, i.e. input and output files on random access devices. A large number of jobs running under Multijob fall into this category, since it is possible to spool the input and output of a job. Since all machines in the network, at this point in time, are compatible, it is possible to off-load jobs at the binary level.

The choice of which job(s) to off-load is of prime importance, since it would be ridiculous to off-load a job which takes 5 minutes to transmit and runs for 30 seconds. Each program has a Mobility Index associated with it, which is the ratio of estimated execution time to expected transmission time (including all the associated files). The candidate with maximum Mobility Index is chosen for transmission.

Communications

For all the above to be possible obviously requires some communications software to utilise the 48K baud line. This is resident in both the 1700 and the System 4 machines. The communications software within the ICNS network is a message response system, such that every message passed over the communications link has to have its successful arrival reported back to the sending site. As the link is full duplex, these responses can be out of phase with the massages (unlike ½ duplex where you have msg, response, msg, response.....). Because of this, many problems can arise when an error occurs, e.g. a message or response is missed.

To ensure the integrity of message sequence, each message has a number attached to it, which is returned with the acknowledgment (either positive or negative), and if an error occurs, then all messages which have been transmitted after the last one which has been successfully acknowledged are retransmitted.

Summary

Thus, to summarise, the facilities available in the ICNS are:-

(i) Ability to change degree of involvement of each machine in the network.

(ii) Ability to 'connect' to any machine in the network, and then use remote machine's resources.

(iii) Ability to transmit files between any two machines in the network.

(iv) Ability to post messages to users on any machines.

(v) Ability to balance the workload of the region by Remote Job Promotion and eventually, off-loading.

(vi) Ability to access all public data held in the network.

Conclusion

No real operational experience has yet been gained on this network, but by the date of the conference, much more will be able to be said about the success, or failure, of this project and the users' reaction to network use. Since networks are still in their infancy, many useful lessons could be learned even if the users' reactions are not favourable by identifying user problems when using a network and improving the solutions next time around.

Acknowledgment

The author wishes to thank the Department of Trade & Industry and International Computers Ltd for permission to publish this paper which results from work which was supported by a normal cost sharing contract under the Advanced Computer Technology Project.

THE CANADIAN UNIVERSITIES COMPUTER NETWORK
TOPOLOGICAL CONSIDERATIONS

John deMercado, Rene Guindon, John DaSilva and Michel Kadoch
Ministry of Communications, Ottawa, Canada

SUMMARY

In Canada plans are being developed[1] for a Canadian Universities Computer Network (CANUNET). This activity is co-ordinated by the Canadian Government's Ministry of Communications in Ottawa which is also participating directly in the study programs.

This paper reviews the results of the study prepared within the Ministry of Communications on the topological analysis of various possible networks for CANUNET[2]. In particular, simulation results for two possible 18 node topologies for CANUNET are presented. One of these topologies is based on the use of purely terrestrial communication facilities, and the other, a hybrid realization, is based on a combination of terrestrial and ANIK satellite facilities. Further details can be found in the report[2] which is part of the study program.

INTRODUCTION

The Canadian Universities Computer Network (CANUNET) program is a joint undertaking of the Canadian Government's Ministry of Communications and some twenty universities that lie across the 4000 miles of territory joining the east and west coasts on which some 20,000,000 Canadians live.

The present objectives and expectations of this project are to develop a plan for providing effective resource sharing of University computing power and serve as a facility for educating Canadians in all aspects of the design, implementation and operation of computer-communication networks.

The master plan for CANUNET is being developed with the following general constraints. Namely the plan

- must be for a truly national network with a minimum of one campus in each province invited to participate in some aspect of its development;
- should accommodate regional diversity and technological alternatives within a framework of objectives, standards and conventions;
- should permit the network to accept various types of computers and to operate over a variety of lines and be transparent to its computers and terminals;
- should permit the network to be compatible with future general computer networks in so far as the outlines of the latter can be discerned, and preferably it should be a subnetwork of such networks;
- should permit decisions to be made in terms of the ultimate network, rather than in terms of the easiest or cheapest communication between a few computers.

This paper summarizes the analysis of two possible 18 node configurations for CANUNET. One of these configurations is based on the use of only terrestrial communication network facilities and the other on the use of a hybrid combination of terrestrial and ANIK satellite facilities. For both of these topologies curves of total input message traffic versus total average message delay for given message lengths are presented. In addition, as discussed, these simulations have been performed for topologies that use either 4.8, 9.6 or 50 kb/s communication links. The theoretical considerations are given in the appendix.

NETWORK TOPOLOGIES

At the present time it is not possible to settle on a topology for CANUNET and various ten, fourteen and eighteen node topologies have been studied[2]. Three speeds of lines, namely 4.8 kb/sec, 9.6 kb/sec and 50 kb/sec were assumed for each of these networks. Table I summarizes the results obtained for these various topologies. These studies have shown that the performance of a given network from the point of view of message delay will depend more on line speed than on the particular topology, however, the topological configuration directly influences the reliability of the network.

Line Speed kbit/s	100% Network Capacity (kbit/s)	Cost $/Mbit
4.8	1 to 45	.23 to 8.78
9.6	47 to 108	.12 to .26
50	302 to 591	.11 to .20

TABLE I

---

* John deMercado is also an adjunct professor of engineering at Carleton University in Ottawa

In Table I, 100% Network Capacity represents the total input data for which the total average delay per message does not exceed 0.5 sec. Also the cost in $/Megabit transmitted is for a network operating with this average delay of 0.5 seconds, 24 hours per day, 7 days a week. This cost does not include the cost of the local lines and the NCU's (node control units).

An 18 Node Terrestrial Network for CANUNET

A possible 18 node topology for CANUNET based on the use of terrestrial communication links is shown in Figure 1. Curves of the total average message delay versus total input data rate for 9.6 and 50 kb/sec communication line realizations of this network are shown in Figure 2. Further results for routing schemes etc. can be found in the study report[2]. The preliminary common carrier quotations are that the communication lines costs for the network (50 kb/s option) including the cost of the modems and lines from local exchange are of the order of $250,000 per month or $13,900 per node per month. Further cost details for this topology are shown in Tables II and III.

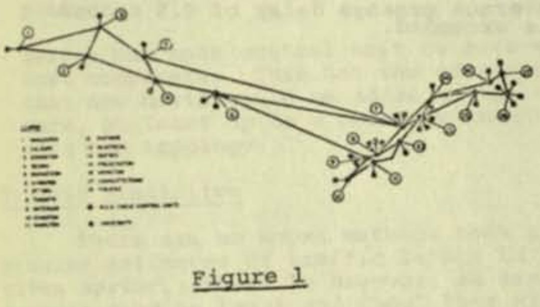


Figure 1

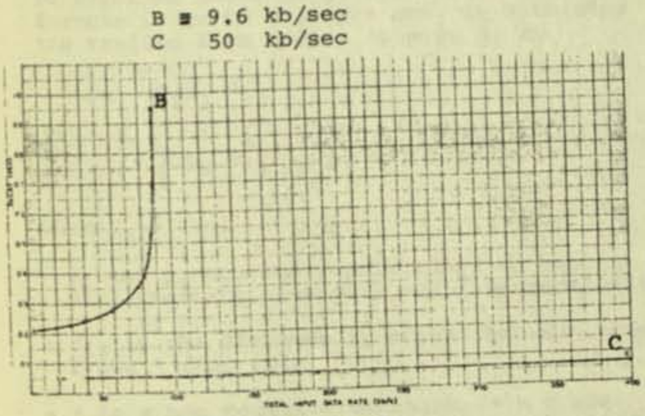


Figure 2

An 18 Node Terrestrial-Satellite (Hybrid Network for CANUNET

A possible 18 node topology for CANUNET based on the use of the ANIK satellite and terrestrial communication facilities is shown in Figure (3). Curves of the total average message delay versus total input data rate for 4.8, 9.6 and 50 kb/s lines are shown in Figure (4). Further results for routing etc. can be found in the study report[2].

A single ANIK radio frequency channel rents for $3,000,000 per year and can support many such 9.6 or 50 kb/sec eighteen node networks. Thus the total communications cost of the single 18 node network using all 50 kb/sec lines and having the topology shown in Figure (3) would be of the order of $310,000 monthly, of which $250,000 per month would go for the rental of the ANIK channel. On the other hand, if four such 50 kb/sec networks with the same earth stations as shown in Figure (3), were being supported by an ANIK channel, then the total monthly communications cost per network would be of the order of $140,000. This compares with the $250,000 monthly cost of the purely terrestrial network. It should be emphasized however that the message/delay performances of both network realizations are different (see Figures 2 and 4).

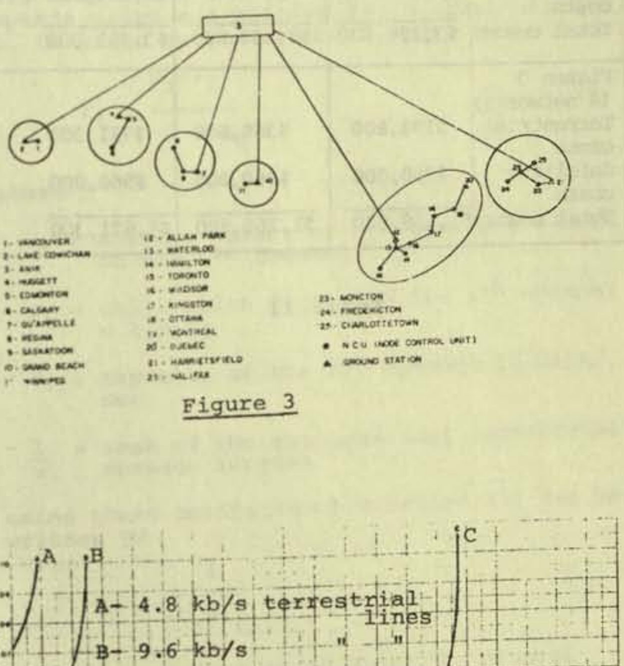


Figure 3

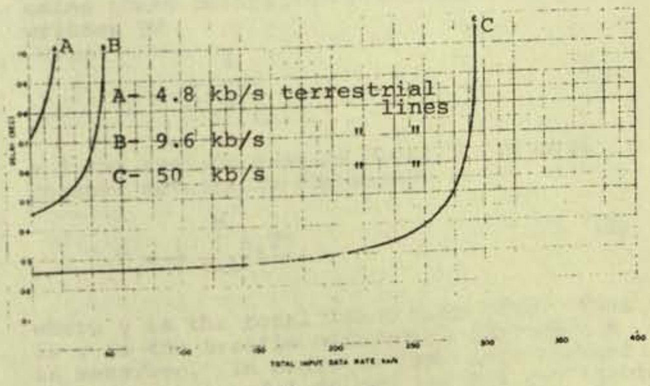


Figure 4

Communication Cost Comparisons

Table II shows the yearly communications cost comparisons for the networks of Figures 1 and 3. Table III shows the costs per megabit transmitted for each of these networks. These costs are for networks operating with a total average message delay of 0.5 sec, 24 hours per day, 7 days a week, and do not include local communications costs. Local communication costs can be expected to increase these monthly figures by between 5% and 10%.

For example, from these two tables it is seen that an 18 node purely terrestrial network with all 9.6 kb/sec communication lines, would cost about $659,000 per year and the cost per megabit of traffic under the operating assumptions listed above would be of the order of $.25.

Network	Yearly Rental Costs		
Line speed	4.8 kb/s	9.6 kb/s	.50 kb/s
Figure 1 costs	$472,164	$659,364	$2,914,200
Figure 3 (single network)			
Terrestrial costs	$198,800	$306,800	$711,300
Satellite costs	$3,000,000	$3,000,000	$3,000,000
Total costs	$3,198,800	$3,306,800	$3,711,300
Figure 3 (4 networks)			
Terrestrial costs	$198,800	$306,800	$711,300
Satellite costs	$960,000	$960,000	$960,000
Total costs	$1,158,800	$1,266,800	$1,671,300

Table II

It is emphasized that no commitment has been sought to date from any common carrier or from Telesat Canada, the owners of the ANIK satellite, to provide facilities for CANUNET. These organizations however, have been most cooperative and helpful in providing their first estimates and other technical assistance.

Network	Cost per Megabit ($/Mbit)		
Line speed	4.8 kb/s	9.6 kb/s	50 kb/s
Figure 1	.607	.246	.196
Figure 3 (single network)	*	*	.553
Figure 3 (four networks)	*	*	.249

Table III

---

* No cost is computed because the total average message delay of 0.5 seconds is exceeded.

APPENDIX: THEORETICAL CONSIDERATIONS

Network Configuration

There is no single best topology for a network like CANUNET. Only those topologies were considered that are flexible and adaptable to new demands without incurring too high additional costs. In all cases, the topologies are predicated on the use of existing common carriers or other communication facilities. No attempt was made to optimize the cost of any particular topology at this stage.

Node Selection

The node locations shown in this study do not necessarily represent the actual final form of the proposed network. The flexibility of the model is such that, as the need arises, modifications to the node locations can easily be accommodated by it.

There are two possible choices that govern the selection of nodes in a Province, and these which dictate the basic network configuration as far as the required numbers of node control units (IMP's in the ARPANET) are

- assign one node control unit per host computer
- assign one node control unit to several host computers. This has the advantage that new hosts could be added to the network, at least up to a point, without changing the topology.

Traffic Simulation

There are no known methods that allow precise estimates of traffic levels to be given apriori; there is however, an empirical formula used by Frank and Chou[6] that allows the simulation of various levels of traffic. A variation of this formula has been used in the CANUNET Model Computer program[3] to generate various levels of traffic (in bits per second) at each node. This has allowed a given networks message-delay performance to be examined under these traffic loads. The formula is quite a simple one, it simulates the traffic from node i to node j, as

$$K \frac{p_i p_j \, d_{ij}^{\alpha}}{\sum_k p_k \, d_{ik}^{\alpha}}$$

where

- p_i and p_j are the populations, in university students at node i and j respectively;
- d_{ij} is the distance in miles, between nodes i and j (for $i \neq j$);
- α is a non negative constant ($\alpha = 0$ was used in CANUNET studies; that is the traffic was independent of distances between nodes);
- k is a numerical constant which determines the traffic level. The value of k used in the simulation varied from .1 to a maximum of 2 for the case where 50 kb lines were being used.

Routing Procedures

Route selection was one of the most important design parameters for CANUNET. The selection of routing schemes in a computer network is of prime importance because they affect the average time that a message spends in the network. In order to minimize this average time delay the route selection should be adaptive[8], that is a function of the traffic level. The ARPA network has in a sense optimal routing in that it provides a minimum of two physically separate routes for each message. The ARPA network is also capable of adaptive routing, that is routes can be chosen according to the traffic level and availability of the lines.

In the CANUNET network simulations fixed routing schemes were adopted. Under these schemes messages are routed between two given points along the shortest path that has the fewest number of nodes that joins these points.

Queuing Theory Considerations

The queuing theory model used in the simulation of CANUNET's topological performance is based on Leonard Kleinrock's work[9]. It has been shown that the average time a message spends using and waiting for a given channel i, is given by the standard M/M/1 queue model

$$T_i = \frac{\rho_i/\lambda_i}{1 - \rho_i} \qquad \ldots (1)$$

where

- λ_i = average number of messages per second on the i^{th} channel
- ρ_i = utilization factor of the i^{th} channel $= \lambda_i/\mu C_i$
- C_i = capacity of the i^{th} channel in bits/sec
- $\frac{1}{\mu}$ = mean of the exponentially distributed message lengths

using these coefficients equation (1) can be written as

$$T_i = \frac{1}{\mu C_i - \lambda_i} \qquad \ldots (2)$$

and the average message delay T for an M channel network is therefore

$$T = \frac{1}{\gamma} \sum_{i=1}^{M} \lambda_i T_i \qquad \ldots (3)$$

where γ is the total input data rate. That is γ is the traffic offered to the network in mess/sec. In order to take into account the fact that nodal delays are not negligible, a constant 10^{-3} is added to the average delay on the i^{th} channel, and the total average message delay averaged over all the channels in the network becomes from (3)

$$T = \frac{1}{\gamma} \sum_{i=1}^{M} \lambda_i (T_i + 10^{-3}) \qquad \ldots (4)$$

The time that a message spends waiting for a channel is dependent upon the total traffic (including acknowledgements) whereas the time spent in transmission over a channel is proportional to the message length of the real message traffic flow. Equation (2) is therefore not correct, and the correct expression for T_i, derived[9] from the POLLACZEK-KHINCHIN formula is

$$T_i = \frac{\rho_i}{\lambda_i} + \frac{(\rho_i)^2}{\lambda_i (1-\rho_i)} \qquad \ldots (5)$$

where

$$\rho_i = \frac{\lambda_i}{\mu C_i}$$

The first term on the right side of (5) is the average service time per message on the ith channel, and the second term is the average waiting time per message on the ith channel. Since this service time depends on the average message length $1/\mu^1$, and the waiting time of the average of the message plus acknowledgement $1/\mu$, equation (5) can be rewritten as

$$T_i = \frac{1}{\mu^1 C_i} + \frac{\lambda_i/\mu C_i}{\mu C_i - \lambda_i} + PL_i \qquad \ldots (6)$$

where the term PL_i has been added to take into account the propagation delay associated with the ith channel. Thus the total average delay T for a message in the network obtained by substituting (6) into (3) is,

$$T = \frac{1}{\gamma} \sum_{i=1}^{M} \lambda_i \left(\frac{1}{\mu^1 C_i} + \frac{\lambda_i/\mu C_i}{\mu C_i - \lambda_i} + PL_i + 10^{-3} \right) + 10^{-3} \qquad \ldots (7)$$

The last term in equation (7) accounts for the delay introduced by the final destination node control unit in delivering the message to its host. This figure of 10^{-3} was used in the case of ARPA network where each host is located only a short distance away from its IMP. In the CANUNET studies, equation (7) was used. It should be emphasized however, that should a host be located far from its network control unit, then the delay of 10^{-3} sec. could be increased. Furthermore the total average delay as given by equation (7) takes into account only short message (packet) traffic.

Message Lengths*

In the model for CANUNET, the fact that acknowledgement traffic affects message throughput was taken into account. Therefore for the CANUNET simulations the following packet lengths were used:

- average message length = $\frac{1}{\mu^1}$ = 640 bits
- average of (message length plus acknowledgements) = $\frac{1}{\mu}$ = 400 bits.

It has been shown[4] that by giving priority to the acknowledgement traffic over the message traffic in particular for line loadings exceeding 50%, that the overall average message delay is decreased. It is worthwhile pointing out that further improvement in the message/delay performance could be achieved if more than two priority classes are introduced. However software and storage costs associated with routing would increase.

Program Description

The input/output for the CANUNET model Computer program[3] is

Input

N → Number of nodes
{P} → Vector of the Population at each node
{D} → Distance matrix
{C} → Branch capacity matrix
$1/\mu$ → Overall average message length
$1/\mu^1$ → Average message (packet) length

Output

{T} → Traffic matrix
γ → Total input data rate
{S} → Shortest distance matrix
{ρ} → Network utilization matrix
{R} → Routing matrix
{A} → Average delay matrix
T → Total average message delay

Reliability Considerations

Preliminary reliability analysis using some new techniques [5,7] of the two 18 node topologies considered in this paper, indicate that the probability is less than 5% that any two nodes will not be able to communicate with each other within 2000 hours. Furthermore for this worse case situation the mean time at which no communication would be possible between a given pair of nodes for a period exceeding 10 hours is of the order of 12000 hours. The probability that this event

---

* These lengths are variable in the simulation program and the studies could have been done using the corresponding ARPA lengths which are 560 bits, and 350 bits, respectively.

will occur within this time is 98%.

ACKNOWLEDGEMENTS

The authors wish to acknowledge many stimulating discussions with Dr. D. Parkhill, of the Department of Communications, Prof. J. Reid of the University of Quebec, Dr. D. Cowan of Waterloo University, Dr. B. Holmlund of the University of Saskatchewan, Dr. J. Kennedy of the University of British Columbia and Mrs. E. Payne of Dalhousie University. Many other people, too numerous to mention individually, are participating in the CANUNET program. The authors would like to thank Miss Gail Widdicombe for expertly typing the manuscript in record time.

REFERENCES

1 University of Quebec — A Proposal for a Canadian University Computer Network (CANUNET), prepared by the CANUNET advisory committee for the Department of Communications. Available from the library of the Department of Communications, 100 Metcalfe Street, Ottawa. March 72.

2 J. deMercado, R. Guindon, J. DaSilva, M. Kadoch, 1 — Topological Analysis of CANUNET. Report of the Communications Planning Branch. Available from the library of the Department of Communications, 100 Metcalfe Street, Ottawa. April 1972.

3 J. deMercado, R. Guindon, J. Da Silva, M. Kadoch — The CANUNET Model Computer Program.

4 J. DaSilva, J. deMercado — Priority Assignment in a Network of Computers. Internal Report of the Communications Planning Branch. Available from the library, Department of Communications, 100 Metcalfe Street, Ottawa. April 1972.

5 J. deMercado — Reliability Prediction Studies of Complex Systems Having Many Failed States. IEEE Transactions on Reliability Theory, Vol R-20, No. 4, November 1971, pp. 223-230.

6 H. Frank, W. Chou — Cost and Throughput in Computer Communication Networks. Network Analysis Corporation, Beechwood, Old Tappan Road, Glen Cove, N.Y. 11542. 1971.

7 J. deMercado — Minimum Cost-Reliable Computer Communication Networks. To appear in the Proceedings of the 1972 Fall Joint Computer Conference, Anaheim, California.

8 H. Frank — Research in Store and Forward Computer Networks. Fourth semi-annual Technical Report, Network Analysis Corporation, Beechwood, Old Tappan Road, Glen Cove, N.Y. 11542. 1971.

9 L. Kleinrock — Analytic and Simulation Methods in Computer Network Design. AFIPS Conference Proceedings, Vol. 36, pp 569-579. May 1970.

10 Science Council of Canada — A Trans-Canada Computer Communication Network. Report No. 13, August 71.

NSF ACTIVITIES RELATED TO A NATIONAL SCIENCE COMPUTER NETWORK

D. D. Aufenkamp, E.C. Weiss
National Science Foundation
Washington, D.C., U.S.A.

Summary

The National Science Foundation is mounting an expanded research program which could lead to the development of a National Science Computer Network linking universities, colleges and other institutions in support of research and education. This expanded research program is focusing on the many issues and considerations related to developing the resource-sharing potential of such a network for the major functions of research and educational computing and science information services. Foundation-sponsored research, developmental activities and special studies have been underway for several years in regard to computer-based resources and facilities for research and education, associated user services and network technology. This paper provides an overview of several of these projects as they relate to the many dimensions and facets of needs and desirability, planning, establishing, operating and evaluating a National Science Computer Network.

Introduction

The National Science Foundation has for many years been sponsoring activities to improve the use of computers and computer-based resources in research, in education and in science information services. The coupling of recent developments in computer and communications technology is adding a new dimension to these efforts. Computer communication networks which are national in scope are in being, and the establishment of a National Science Computer Network linking users at academic and other institutions to specialized resources for computing and science information services would have profound implications for resource sharing in research and education.

In some cases, a network would facilitate the pooling of resources, including the collaboration of researchers independent of institutional affiliation or the establishing of facilities for shared use, which might be beyond the financial reach of any individual institution. In other instances, special computer languages, computational algorithms, data banks and other computer-related developments which were created in a local environment could have direct and immediate application for others in the research and educational community. In particular, such a network would offer possibilities for a much closer integration of computing and science information services than has been effected previously.

The problems of creating and utilizing a National Science Computer Network for research and educational computing and science information services are not only in technical areas but very much in organizational, political and economic areas as well. Issues and considerations that arise immediately are those of specialized resources and services, common interest user groups, network management, impact on campus computing centers, network financing, and the overall need and desirability.

The Foundation's Office of Computing Activities (OCA) and the Office of Science Information Service (OSIS) are cooperating in mounting an expanded research program relative to a National Science Computer Network.[1] A comprehensive and interrelated set of project activities is envisioned to explore and evaluate the many dimensions and facets involved. Research, development and special studies in connection with resource sharing, user services and network technology are already underway, and this paper provides an overview of several projects recently sponsored by the NSF as they relate to this expanded research program.

Research Resources

One of the arguments for a National Science Computer Network is that specialized facilities and resources could be available for given disciplines or classes of computations. The National Center for Atmospheric Research (NCAR) with its large-scale computing facility is well known. Although the principal use of the computing facility is in support of NCAR's own research, the facility is available to academic researchers with appropriate research problems. The Inter-University Consortium for Political Research, with a membership of approximately 150 institutions, has both data and programs available. The Chemical Abstracts Service of the American Chemical Society is a computer-based science information resource which is also nationwide in scope.

Recent Foundation-supported projects for which a National Science Computer Network would have special implications include: The Computer Research Center for Economics and Management Science of the National Bureau of Economic Research, a study of a national laboratory for theoretical chemistry, and a study of a national center/network for computational research on language. Another project for which the Network might have a long-range benefit is the Clearinghouse and Laboratory for Census Data, although its establishment was not predicted on the emergence of network facilities. These four projects are described briefly in turn.

Computer Research Center for Economics and Management Science

The Computer Research Center for Economics and Management Science of the National Bureau of Economic Research is a major disciplinary center established with support from the Division of Social Sciences and OCA. The Center was conceived to stimulate new computational technology for research in the disciplines of economics and management science. Advanced research computational techniques systems are created and distributed to the research community. A key element in the organizational structure of this Center is a Policy and Operations Committee. Its membership includes academic scientists with research interests in economics, management science, computer science, mathematics and statistics. The Center is designed to attract researchers in these fields and in computer science from universities and colleges, Government, nonprofit institutions, and the private sector. Researchers, both in residence at the Center and in remote locations, are participating in the research effort. The development of a National Science Computer Network would add materially to these efforts.

Study of a National Laboratory for Theoretical Chemistry

The National Academy of Sciences-National Research Council is conducting a study of the feasibility of a national laboratory for theoretical chemistry under a project supported by the OCA and the Chemistry Section of the Division of Mathematical and Physical Sciences. The study is being carried out by a committee of approximately 30 individuals selected to include chemists from diverse areas of research, computer scientists, mathematicians, and representatives from other relevant scientific disciplines. The main tasks of the committee are to identify important chemical problems susceptible to solution at such a laboratory, identify special problems of computer science involved, and to explore and characterize the structure and operation of such a facility. In addition to these substantive matters, the committee is concerned also with the broader questions of the impact of such a laboratory on university computing centers, development of computer networks, computer program exchange, and interfacing the laboratory with the scientific community.

Study of a Center/Network for Computational Research on Language

A study is being conducted at the University of Kansas with support from OCA and the Division of Social Sciences to develop the concept of a national center/network as an innovative approach for the broadly based interdisciplinary area of computational research on language. The optimal organization structures for any such center/network are being studied as well as the critical questions of priorities and implications as to areas of research and methodologies to be employed. A wide range of individuals with respect to geography, institutions, and disciplines are participating in the effort. A published study of the project activities is forthcoming.

Clearinghouse and Laboratory for Census Data

Another resource with national import is the recently established Clearinghouse and Laboratory for Census Data (CLCD) supported by the Division of Social Systems and Human Resources. The CLCD is operated by DUALabs under contract with the Center for Research Libraries. Assistance to researchers includes census use orientation and training, research proposal review and cost forecasts. Training programs are provided, needed computer software is developed, and a central information source is maintained of selected census research projects and of available data files. The activities of the CLCD are augmented through the cooperation of other organizations to whom the programs, catalogs, registers and materials of the CLCD are made available.

Research Services and Uses

The development of user-oriented services for network application is fundamental to a viable National Science Computer Network. Although the research resources described in the previous section would all have associated service functions, the development of some kinds of services and facilities need not be associated with a major regional or national resource. Several projects are highlighted which are illustrative of possible kinds of research services and uses within a Network.

Software Development for Theoretical Chemistry

A discipline-oriented software development project in theoretical chemistry is underway with OCA support at the University of Utah under the direction of Dr. Frank E. Harris. The software is being designed in a way to facilitate the use of theoretical calculations by workers in all fields of chemistry including those not expert either in details of the theoretical analysis or in digital computation. The project is also novel in that it provides for consultation regarding distributed software, pilot use of remote computing facilities, and preliminary development aiming toward eventual public use of computer networks. The use of the ARPA Network will be explored in support of these project activities.

Quantum Chemistry Program Exchange

Another kind of prototype network service is that represented by the Quantum Chemistry Program Exchange (QCPE) at Indiana University. Members of the QCPE submit programs together with documentation and test data and outputs. The QCPE provides a uniform testing environment for the programs, completes the documentation and makes the availability of the program packages known through appropriate newsletters and other reports. Foundation support has been provided by the Chemistry Section of the Division of Mathematical and Physical Sciences and OCA. The operation of the QCPE is expected to be eventually self-sustaining on the basis of membership and user charges. The programs being developed under the theoretical chemistry software project noted previously will be distributed through the QCPE.

Hierarchical Computing

Hierarchical computing in which access is provided to a range of levels of computing and information services in support of research or instruction relates directly to resource sharing via a national network. In one OCA-supported project in hierarchical computing, the use of minicomputers in scientific experiments may be improved as a result of a prototype interfacing support system under development at the University of Chicago. The purpose of the project is to provide the means for enhancing scientific methodology for experimental research. This experimental system consists of support for the researcher at three levels: (1) a monitoring subsystem to which minicomputers can be interfaced, (2) a subsystem providing high-level support services such as compilation and bulk data storage, and (3) a subsystem which would facilitate access to major computing facilities on a national network.

Use of Remote Graphics

One of the projects supported by OCA which can be viewed as a prototype of network use in research is being conducted by Dr. D. Raj Reddy at Carnegie-Mellon University. The primary long-term objective of this research is to develop geometric models of structure and organization of the neural networks of the swimmeret system of the lobster. The functional requirements for the research computer system include getting visual data into the computer, reducing it to compact form, displaying the resulting three-dimensional structure on a CRT display to permit researchers to formulate theories, and devise algorithms for comparison of neural networks. One expectation of the project is that the feasibility and utility of such a network modeling system can be demonstrated using an Evans-Sutherland LDS1 graphics system remotely through CRT facilities linked to the ARPA Network at Carnegie-Mellon University.

Use of Remote Data Banks through "Intelligent" Terminals

The use of a large-scale computing facility with specialized data banks and program library through "intelligent" remote display terminals is being addressed in a collaborative project involving Brookhaven National Laboratory and Texas A&M University. A "Protein Data Bank" has been established at Brookhaven National Laboratory in cooperation with the Crystallographic Data Center in Cambridge, England. The prime purpose of such a computer-based, macromolecular structure library is to facilitate dissemination of raw or refined data in a variety of forms to permit users to ask easily chemically interesting questions. One goal of the project is to develop a crystallographic computing system which provides to the researcher in an individual laboratory all the computing facilities necessary for molecular structure determination. A relatively inexpensive minicomputer-based display facility has been constructed which is capable of three-dimensional graphics and interactive use. In addition to the developmental activities supported by OCA, the system will be refined through the experiences of a third group in a remotely-located research laboratory.

Networking Activities Relating to Science Information Systems

The Foundation's Office of Science Information Service has also been supporting system development efforts which are primarily based in the scientific professional societies and the university sector. The society-based systems produce the information bank (i.e., abstracts and indexes of the world's scientific literature) and supply tapes recording its content. The university-centered systems provide the distribution outlets to the user community.

The problems which will pose the major obstacles to successful efforts in networking of information resources are not judged to be primarily technical in nature. The areas which require immediate attention and where it is intended that initial support be focused are concerned with administration, user requirements, economics, and legal matters. For example, the optimum mechanism for the administration of a network which may consist of a loose consortium of universities cannot at this time be specified with any degree of precision. When this problem is compounded by the factors of geographic dispersion and regional interests on a nationwide basis, the ramifications of this concern only begin to emerge.

Thus, the communication system of science and technology is a melange of procedures, media, production techniques, services, centers, and people. The system has evolved by way of necessity and expediency, and in its current stage of development exhibits many different types and levels of complex interactions. Many services and institutions, which constitute centers of experience in various segments of the information transfer community, are operational at the present time, and these will certainly function as nodes in whatever network will evolve. Such a network undoubtedly will be evolutionary, in the sense that it will develop as circumstances permit, rather than being systems engineered a priori. The present resources in operational components of the network represent an investment too large to scrap in favor of an optimal design even if it were possible to design the optimal network. Because of this, it becomes necessary to plan a course of action flexible enough to accommodate different strengths and relationships among the existing nodes which include the tape suppliers and publishers, as well as the distribution centers and libraries.

A natural outgrowth of these efforts has been a start toward embryonic networking activities. Moreover, the costs associated with system development and operation are rapidly making it clear that such activities are likely to become an economic necessity in the future. Much work remains to be done to foster cooperative arrangements, but much has already been accomplished. The interface of OSIS-supported information systems with OCA-supported regional computing networks in projects at the University of Georgia and Lehigh University is illustrative of these cooperative efforts. The following are examples of current activities:

University of Georgia

The system which is under development at the University of Georgia will provide twenty-six Georgia universities and colleges access to a broadly-based information service via the OCA-supported regional computing network. It also services many external users, including Federal agencies. It is a very broadly-based system, but the primary focus is in the areas of chemistry and the biological sciences. The system offers batch-mode current awareness searches based on the concept of selective dissemination of information as well as in-depth retrospective searching of large information bases such as those produced by the Chemical Abstracts Service of the American Chemical Society.

University of California at Los Angeles

The system which is under development at UCLA will eventually service the nine campuses of the University of California System. They currently plan to experiment with providing current awareness searches over the ARPA Network to users at one of the nearby universities that is also on the Network. While the system presently employs information specialists at the user-computer interface for batch-mode service, there is also an interactive mode whereby queries can be made from any of forty rapid response terminals which are strategically located around the UCLA campus.

Lehigh University

The on-line, interactive system at Lehigh University permits the user to query full text articles and abstracts in a natural language mode for current documents via remote CRT terminals. This service primarily reflects the demands of a user group which is comprised mainly of engineers in the various research centers at Lehigh University. In addition, it provides information services to seven other academic institutions in the Lehigh Valley Regional Computing Network whose development was supported by OCA. They are also conducting an information networking experiment with the University of Georgia via a phone line, dataset, and UNIVAC Uniscope 100 CRT terminal. The purpose of this experiment is to provide an operational service as well as to establish key roles for future network nodes.

For example, the Lehigh system does not have extensive document files, but it does possess considerable software for full text analysis, automatic document characterization, file generation, data base construction, and interactive search negotiation. It is conceivable that Lehigh as a network node would maintain no document files but would instead act as an interface message processor, accepting requests from remote terminals, negotiating them, and dispatching them for execution to other nodes. It could also act as a network monitor for information flow.

Georgia, on the other hand, has the most extensive data files of any university information system, so that the determination of its role in the network will be of primary importance. Clearly, all will benefit from networking interactions provided the allocation of functions occurs in an optimum fashion.

The Environmental Protection Agency

The Environmental Protection Agency has just announced a program of cooperation with science information centers developed through grants from OSIS. This cooperative agreement was undertaken with the firm conviction that existing systems and services should be used whenever possible, rather than generating new ones.

Under this new program the information service centers of three universities will be coordinated for EPA use. Lehigh University has been selected to provide on-line services, the University of Georgia will be responsible for batch and retrospective services, and Ohio State University will produce selected SDI programs on a trial basis. This program will permit the Environmental Protection Agency to have access to over twenty-five data bases, while determining usage requirements of the various files and services.

This arrangement will serve a network of some forty EPA libraries and information centers, with the library at EPA's National Environmental Research Center in Cincinnati, Ohio, designated as access point to the information centers. A series of terminals will soon be installed in Cincinnati to provide on-line and batch search and retrieval capabilities. In this manner each EPA facility will be able to search over two million documents and receive a printout with a minimum of delay. Eventually this network will be expanded to improve data base coverage and to provide additional services to EPA personnel interested in searching data bases of environmentally-oriented literature.

Georgia/Lehigh

A simulation study of an information dissemination network is now being supported by a joint grant to the University of Georgia and Lehigh University. Alternative configurations and modes of operation for a computer-based information dissemination network will be studied through the use of simulation and mathematical modeling techniques. The operational characteristics of the network system and associated costs will be examined with respect to time and anticipated growth rates. Computer hardware requirements, communications media, anticipated usage, and manpower requirements will be taken into consideration in design of the model. Several types of dissemination centers and functions for nodes within the network will be studied. The technical plan calls for the development of an adequate model of an information dissemination network and the simulation of network activity via model manipulation. Emphasis will be placed on design criteria assumed in the model and the alternative configurations derived during the course of the simulation model as well as the administrative problems associated with the implementation of such a network.

Instructional Uses

One aspect of the National Science Computer Network initiative is to explore the extent to which such a network could be used in the support of instructional computing on an individual campus. The NSF has already supported activities which have addressed many of the problems involved. The Institutional Computing Services Program, for example, provided support not only for the development of the campus computer center but also for facilities to be shared among institutions. The OCA Regional Cooperative Computing Activities Program also provided support for projects to explore the development of shared computer resources for educational use. The first regional projects under this program were established in 1968. Approximately 30 regional computing projects have been supported involving about 300 institutions of higher education.

Projects have been initiated to study the problems of computer-based curricular developments, documentation and transfer. One of these projects which could have a bearing on the use of a nationwide computer network for instructional support is described below:

The CONDUIT Project

Five regional computer networks are collaborating with OCA support to test the feasibility of transporting computer-based materials for the purpose of enhancing curricular development in higher education. The participants are: Dartmouth College, the North Carolina Educational Computing Service, Oregon State University, and the Universities of Iowa and Texas. The experiment which goes under the name of "CONDUIT" involves forming a central organization and staff at the five regional networks to transport curricular units; testing the transported materials in classrooms in each network; producing a catalog (with standards for material quality and transportability) of curricular materials for national dissemination; and documenting the feasibility study as to procedures, cost considerations (including billing and accounting procedures, user feedback, and adequacy of the process).

Software Testing and Distribution

The National Science Computer Network initiative has special implications for the analysis, testing and distribution of computer programs. A nationwide network could do much to facilitate the transfer of programs from machine to machine and at the same time reduce the need for doing so. It is envisioned, for example, that some sites on a network could assume responsibility for maintaining certain software. This is the case, for example, with some of the systems being developed at the Computer Research Center for Economics and Management Science. OCA and the Division of Engineering are already cooperating in exploring approaches to improving the quality of engineering software. The special attention given by the NSF to problems of improving the quality of mathematical software for research use is highlighted by two projects which have implications for a National Science Computer Network.

National Activity to Test Software

The National Activity to Test Software (NATS) is being conducted with OCA support by Argonne National Laboratory, the University of Texas and Stanford University in cooperation with various field test sites. This project is a collaborative prototype effort to evaluate, certify and disseminate mathematical software. Selected mathematical software, including a set of special function routines and a package of matrix eigensystem routines, is tested, first at the three principal institutions and then at 16 field test sites. The certified products are distributed by the Argonne Code Center. A key aspect of the project is the emerging of a collaborative testing methodology.

Study of Alternative Approaches to Testing and Distributing Software

In a related project supported by OCA, the University of Colorado and Argonne National Laboratory are collaborating in a study to explore approaches to the problems of making high quality mathematical software readily available to the computing community. The study is examining the possible roles of academic institutions, the National Laboratories, the National Bureau of Standards, and the private sector. It is intended that the study produce source material and a blueprint for action. A National Science Computer Network could have special relevance for these efforts.

Network Technology

The OCA and the Division of Engineering have supported several projects which are leading to new developments in computer communications for networks. Two projects supported by OCA which have a special bearing on advancing network technology are the MERIT project in Michigan and the Distributed Computer System being developed at the University of California at Irvine.

The MERIT Computer Network

The MERIT Computer Network, Michigan's Inter-university Computer Network, is an experimental project to link computing systems at the University of Michigan, Michigan State University, and Wayne State University for the purpose of enabling the three institutions to share and extend their computing resources. The MERIT project is concerned not only with the technical problems of computer-to-computer communications but also with the political, organizational and economic problems of sharing major resources. Special emphasis is being given to developing the user services to allow students, faculty and staff of these three universities to use the computing resources of all three systems. The development and testing stages involve Michigan, Michigan State and Wayne State only; however, the MERIT Network does serve as a pilot project to be analyzed in connection with exploring the resource-sharing potential of a nationwide computer network.

Distributed Computer System

An experimental computer network called the Distributed Computer System is being developed and constructed at the University of California at Irvine. This research project is designed to explore those issues in distributed architecture useful in serving mini and midi scale computers which are appearing in large numbers on university campuses. Communications within the system is based on a digital ring topology to which computers are interfaced. The communication protocol is unusual in that messages are addressed not by the location of the receiver but by means of the name of the receiver. Aims of the project are: low cost, reliability, easy addition of new services, modest initial costs and low incremental expansion costs. Such a network could also be considered as a subnetwork in a hierarchy including access to a national network.

National Bureau of Standards Activities Relative to a National Science Computer Network

The Center for Computer Sciences and Technology of the National Bureau of Standards is carrying out studies and other special tasks with OCA and OSIS support in connection with the Foundation's Network initiative. The current efforts focus on the areas of network communication technology and network management and operation and computerized information systems as related to this emerging Foundation thrust.

Alternative technological approaches are being evaluated by the National Bureau of Standards within a framework of possible network structures and functional requirements and constraints for an eventual National Science Computer Network. Included in this effort is a compilation of all relevant literature on network technology. Another thrust of the Bureau's efforts includes considerations of network management. This function must be emphasized strongly in achieving a viable network operation. The numerous tasks associated with setting up and operating a national resource-sharing network are being identified, and an evaluation is being made of alternative strategies for implementing and operating a network utilizing various managerial structures.

Criteria are also being developed for measuring the performance of the proposed Network. Included, too, is a detailed measurement plan containing accounting mechanisms to satisfy overall management objectives. Another important part of the current effort is the development of a set of terminology to support adequately communication among network planners, designers, implementers and users. This terminology will be based on state-of-the-art usage of terms in the computer networking area.

Implications for Colleges and Universities

The establishment of a National Science Computer Network would provide new options to academic institutions for computing and information services, but would also raise organizational, political, and financial problems. Several studies are in progress which are addressing some of the issues and considerations involved. The discussion of activities relating to science information services made note of some of these considerations. Three other studies are also highlighted: A study by the University of Denver is addressing the overall problem of managing and financing academic computing centers; the University of Southern California, California Institute of Technology, and the University of California at Los Angeles are exploring new approaches to computer resource sharing, and the University of California is examining the role and impact of computer and information sharing for a multi-campus institution within the context of a national network.

Alternative Approaches to the Management and Financing of Academic Computing Centers

The University of Denver is conducting a study of alternative approaches to the management and financing of academic computing centers. The purpose of the study is to determine how to optimize insofar as possible computing capabilities and operations on individual college and university campuses, and to identify conflicts between this optimization and established institutional goals, policies and operations. The sharing of facilities by means of a network is one possible alternative. Case studies will be developed in depth for selected alternatives reflecting experiences of particular institutions in responding to specific problem areas and in the effects and consequences of such actions in that institution. Conclusions of this study should be applicable to most four-year institutions of higher learning in the United States and should assist college and university administrations in dealing with the complex and growing problems of the role, management, and financing of the computing center on campus.

Computer Resource Sharing

The University of Southern California, California Institute of Technology, and the University of California at Los Angeles are collaborating on a feasibility study of computer resource sharing. This study is addressing questions of sharing computer resources within the framework of a nationwide network with special emphasis on implications for a subnetwork of institutions as these three institutions might represent. Areas of concern include responsibilities of the network to institutions and, conversely; management considerations relating to subnetworks; staffing at local and network levels for optimum use of network resources; alternative strategies for coping with risks and liabilities of sharing network resources; and approaches to network financing.

Network Implications for a Multi-campus University

The University of California is conducting a study on the academic, administrative, and economic implications of a sophisticated data communication network on a decentralized, multi-campus university. The results of the study are expected to have applicability not only to the University of California but also to other educational organizations or federations in evaluating alternative network developments.

The issues and considerations being examined fall into the general categories of (1) the impact of increased access to both internal and external services on internal and external users and suppliers; (2) the requirements of instruction, research, administration, hospital and library computing and the effects on them of various current network technologies; (3) the interconnection of diverse networks and the impact of interconnection on all participants; and (4) the isolation of technical and administrative criteria to define a viable network.

Perspective on the NSF Network Initiative

The preceding discussion provides an overview of some of the projects supported by the Foundation which are intended to explore the resource-sharing potential of a nationwide computer network. There are, of course, many dimensions and facets to such a thrust and the Foundation's expanded research program relative to a National Science Computer Network is designed to permit exploration and evaluations of issues and considerations involved.

Guidelines are anticipated in regard to other activities including a proposed trial National Science Computer Network to support these activities. The projects highlighted in this paper should, however, be indicative of the scope of this Foundation initiative.

Reference

1. Aufenkamp, D.D. National Science (Computer) Network, Proceedings of the EDUCOM Spring Conference, Networks for Higher Education, April 13, 1972, EDUCOM, The Interuniversity Communications Council, Inc., Princeton, New Jersey, 1972.

TELEPROCESSING - THE UTILITY OF THE COMPUTER UTILITY

CHAIRMAN:
Gerald H. Brody, Manager
Telecommunications Marketing Systems, Engineering, Raytheon
Data Systems, Norwood, Massachusetts, U.S.A.

PROGRAM COORDINATOR:
Dr. Philip H. Enslow, Jr.
Office of Telecommunications Policy,
Executive Office of the President
Washington, D.C., U.S.A.

PAPERS:

Max P. Beere, Tymshare, Inc., Cupertino, California, U.S.A. *"Teleprocessing–The Utility of the Computer Utility, New Problems? New Challenge!"*

George J. Feeney, General Electric Company, Bethesda, Maryland, U.S.A. *"The Future of Computer Utilities"*

Peter T. Kirstein, Professor of Computer Systems, University of London Institute of Computer Science, London, England, *"On the Development of Computer and Data Networks in Europe"*

DISCUSSANTS:

John C. LeGates, Executive Director, Educational Information Network (EDUCOM), Princeton, New Jersey, U.S.A.

Charles Dalfen*, Legal Advisor to Canadian Department of Communications, Ottawa, Ontario, Canada

William M. Zani, Associate Professor of Business Administration, Harvard University Graduate School of Business, Cambridge, Massachusetts, U.S.A.

THEME: The existence of Teleprocessing networks and computer utilities is an accepted fact, both on a national and global scale. This session will examine the effectiveness of today's operations, what direction they may go towards in the future, what benefits, both social and economic are possible, and what impediments may exist in the technical and legal domains.

---

*Now at Faculty of Law, University of Toronto, Toronto, Ontario, Canada

TELEPROCESSING—THE UTILITY OF THE COMPUTER UTILITY NEW PROBLEMS? NEW CHALLENGE!

Max P. Beere
Tymshare, Inc.
Cupertino, California, U.S.A.

The telephone was designed by a gentleman who understood speech in all its ramifications and the physiology of the ear – probably without peer for his time. Alexander Graham Bell knew relatively little about electricity, electronics, or transmission. He designed a mechanical ear to transduce acoustic sound waves to an approximating sine wave, and, for the receiving side, he designed a mechanical voice box to re-create the acoustic sound waves, and, thus, by happy accident: "Mr. Watson, come here, I want you!" That was March 10, 1876.

Computers were in their gestation period about then. It was not evident 100 years ago that computers would become increasingly dependent on communications. Though, tracing the lineage of computers and communications back to their inception, one now sees a definite and inexorable convergence of the two diverse but interrelated technologies. Today, it is difficult to separate one from the other.

It was a fantastic breakthrough for man on that day 96 years ago, when it was possible for him to project his voice over a wire via a vehicle designed to meet his need. Where was the "Mr. Watson, come here, I want you!" for the data machine, the computer? It never existed. Samuel F. B. Morse laid the groundwork in 1837; Guglielmo Marconi in 1895 made wireless telegraphy possible – but somehow these two data language-oriented breakthroughs never really germinated and matured. At least not as far as machine-oriented communication is concerned. The telegraph in a digital sense has just aged! The wireless has taken a left hook away from its original direction and has become very useful as a carrier of analog communications.

What, then, must be done today to associate computers with a communication system? The digital output of the computer must be made to approximate analog voice signals. This is not difficult to accomplish in feat, but is difficult to do well. Why we must accomplish computer communications this way is easy to understand: At present, *it's the only game in town!* Originally, when computers were made to communicate, there were so few of them and the idea was so farfetched that rather than design a communication system to meet the computer need, the computer output was modified to meet the existing voice communication capability. This method of meeting the data need has continued – and thus the imperfect computer communication technology that exists today.

It is time we looked around. We can solve our immediate computer communication problems consistent with the tools at hand. A lot of people are working on this problem in an effort to find a solution. Many have promised much. A few are becoming reality. The Bell System Digital Net, MCI, Datran, Western Telecommunications – each has an answer according to its understanding of the problem, and each is striving to get its solution underway. But, in each case, it will be many years before the service is available to the extent it is needed today. It takes time, money, political acumen, and good business sense to get a commercial data communication system devised, installed, and operating. In the companies mentioned, with the exception of Datran, each has a start on its systems, but in every case, including the Bell System, it will be 5 to 10 years before the computer communication need can be met on a nationwide basis. More power to them, and God speed!

What of today? The need is today! The business is here today! Our computers and their attendant business machines have the need to communicate today, and that need is growing. It demands decisive attention. We must use available communication capability to answer that need because it is *available now!* We must build our system out of what is available today to bridge the gap tomorrow! It is fraught with problems – No, it offers an exciting new challenge! A challenge the entrepreneur will not let pass. Let's get back a little of our frontiersmanship. Where is our "blood and guts" approach to solving problems? What happened to "damn the torpedos," "remember the Alamo," and the charge up San Juan Hill? Must everything be done with specially built pieces that go together in a precise manner? Did the frontiersman, standing in the middle of a forest, bemoan his fate because he did not have nicely finished planks to build a house? Did he give up because the land he intended to farm was densely forested and covered with brush, no fence in sight? Did he scream foul when it rained on his head? The answer I suspect for the most part was "No!"

We are in the same spot today in the interactive computer communication network arena. We are knee deep in weeds and though help is in sight, it is not available at present. And there are Indians in the woods too – the FCC, PUC, OTP, and the Bell System – to name a few of the more noteworthy tribes, and they have many scalps hanging in their lodges. We can run and lose. We can fight and possibly win. The Carterfone decision is a win, so is the Special Service Common Carrier, but after a lengthy and debilitating fight. We can lose, too, by fighting. You may win a fight with a skunk – but!!! Another way is to befriend, to cooperate, to negotiate. We can read in history that our pilgrim fathers used these methods with the Indians and won. We have been celebrating that win every Thanksgiving.

The entrepreneur must look around himself for whatever raw products he can find, which, by using his wit and strength, he will be able to build with. The frontiersman cuts down trees to build his house and clears his land for planting. He uses scraped animal skins for windows to let in soft light, and sod to keep the roof water tight; builds fences, digs ditches, and deals with the weather and the Indians in whatever way he can to achieve his objectives. If fight he must, he will. But he would rather deal.

Our business need is to communicate *today* – to survive. We can stumble along for a while, but we need adequate computer communication at a reasonable cost as soon as possible. The computer technology is becoming completely dependent on the ability to communicate. The computer technology has far outstripped the capability of the offered communication service to meet the need, mostly because of the voice orientation of the existing system.

But, today's computer communication needs can be met by the existing communication *facilities*. The need can be met by augmenting the existing communication capability with one's own expertise.

Simple error detection and correction schemes applied by the entrepreneur to the common carrier facilities can alleviate the problem of service that is too error prone for adequate computer communications. Catastrophic failure of common carrier facilities, such as severed cables, fire, flood, noise, sun spots, or poor workmanship, can be eased by a redundancy of facilities and automatic alternate routing – again applied by the entrepreneur to the common carrier facilities. The plethora of terminals that are available to the vendor of interactive computer service is amazing, each with its own characteristics: differing speed, code, and character set. To meet the differing service demands from a mix of customers in the general interactive computer environment, many different terminals must be allowed for. Since businesses purchase terminals suited to their individual needs, the computer service vendor can no longer dictate which terminals he will interface. If he is going to stay on top, he must accept them all.

Cost of service is another aspect that cannot go undetected. While the computer service vendor is trying to lower, or at least maintain, his price level, we find the communication costs slowly rising. To meet this challenge, the computer communication network must increase in efficiency faster than the cost can rise. This, too, can be done by the company that is oriented around solving its problems regardless of their origin.

Tymshare, Inc. has taken these entrepreneural steps, and has proved that a quality interactive computer network can be built using the offerings of the common carrier and the technical expertise of Tymshare people. Tymshare management could see that in order for them to achieve the company goals they had set, they would have to solve problems previously considered to be in the realm of the common carrier – and, also, previously considered unsolvable.

TYMNET®, Tymshare's international interactive computer communication network, is the result of this effort. TYMNET® is a move that demonstrates the plausibility of removing the myth from today's computer communication mythology by substituting tech to establish the ***computer communication technology***. By maintaining a cooperative, no nonsense relationship with the common carrier and by resolving problems as they arose, Tymshare was able to achieve its objective of creating a viable computer network. The network now exists. The cost has been high. What of the payoff?

In any evolutionary system, serendipity plays an interesting role. We, and others, have achieved networks that work for us. As a matter of fact, it would work for any vendor of interactive computer service, and it will work anywhere – anywhere there is a communication facility available. The question now is: Is computer network service a salable commodity? Can it be a product from which revenue can be expected? A profit gained? Certainly, a lot of technical genius and money have gone into creating these networks, just as in any new product. It has gone through a gestation period and has proved its worth. But, are the Indians pacified? Or are they waiting in the bush, ready to pounce as soon as a weak spot is detected, an infraction of a uniquely interpreted rule? Why must this be? Certainly this "added value" makes the common carrier service upon which it is based more viable, and indeed creates a new market where one previously, if existed, was decaying. The regulatory agencies should feel few pangs from lack of control, since they do regulate the vehicle upon which this computer network technology is based. The Office of Telecommunications Policy should be happy, since the American public is getting a better return on monies invested in creating a vast national communication network, because new use means new revenue. The special service common carriers seem a bit concerned, but again, why? This added value service is something they, as of now, have not indicated the desire in offering. In fact, they stand to gain, like any other common carrier, because their service, when it is available, can be used equally as well if not better.

It appears that we now stand on the threshold of a new business venture. A new need is becoming very apparent – computer communications. A solution to that need is available today, and it can be made available to those who need it. That solution is in the form of the sophisticated computer networks that exist today for intra-company use. Some inter-company sharing is allowed, enough to nurture this infant along. There exist no known bounds for this undefined new business area. It can be defined as something we already have – and squelched – or it can be looked at for what it is – a new approach to solving an old problem but a new need. Computer network service is standard communication capability with added value put in operation by the entrepreneur who has the capability and the foresight to do so.

THE FUTURE OF COMPUTER UTILITIES

George J. Feeney
General Electric Company, Bethesda, Maryland, U.S.A.

ABSTRACT

This paper argues that computer utilities will provide a superior alternative to the fragmentation and inefficiency of today's decentralized computing even as electrical utilities replaced established decentralized power generation seventy years ago. Economic, technical and institutional pressures toward centralization are examined in the context of the rapidly fading glamor once attached to "having a computer". The technical feasibility and operational implications of highly centralized computer utilities are illustrated through discussion of the history, current status and future plans of a large scale computer utility network.

The fact that we are here today discussing computer utilities as a reality rather than a vision indicates the phenomenal advances that have occurred in the past few years in both computer technology and user sophistication. One might argue that these network utilities represent an adaptation to some unique, highly specialized requirements. Instead, I think they portend some fundamental changes which are starting to take place in the way computing is sold, produced, and used throughout the world.

To better explain this change, let's look at the parallel development of electrical utilities 70 years ago. In 1905, approximately 50,000 local stations owned by private companies produced about one-half of the total U.S. power, largely for their own internal needs. Today only about 6% of the total electric power is produced by local firms for internal consumption. The bulk of this country's power is produced by about 200 large utilities. Innovative work by men like Steinmetz and Stanley in the technologies of power generation, transmission, and distribution made the utility approach possible. Highly concentrated power generation facilities resulted in major economies of scale. High voltage transmission technology made it practical to transmit power over long distances. Local distribution facilities then provided the means to deliver power efficiently to the end users.

We believe that the computer industry is on the verge of a similar change. Today, computing is decentralized and fragmented. There are over 60,000 general purpose computers in the United States, most of them providing strictly local in-house service. This is largely the result of the growth and evolution of the computer industry itself. The advances of second and third generation computer technology brought computers within the reach of not just corporate headquarters, but every major plant and office location. Even though this process of proliferation is still taking place, a major counter trend toward centralization is emerging. This is largely due to three factors.

First of all, there are the economic pressures of significant savings that could be realized through improved coordination of computing. In principle, a few very large machines can perform a given amount of processing more cheaply than many small ones. In addition, substantial savings can usually be realized through the standardization of procedures and the pooling of resources. Second, the technical obstacles to unification are gradually disappearing. Communication costs and technology are becoming more attractive. Relatively wide band data transmission is now feasible over voice grade telephone lines. High performance remote entry terminals are now available, and modern operating systems are making networking relatively manageable. And third, there are strong business incentives for large national firms to centralize their computing. More and more multi-location companies are discovering that you can't manage vital national and international operations like order service, inventory control, field sales, factory scheduling, and financial planning with a scattered collection of cottage computers.

There is another related trend accompanying the move toward centralization. Even though the value of computing continues to grow, the glamor surrounding computers themselves has started to fade. Many companies are looking for relief from the "computer rat-race"; that unending cycle of system upgrade, over-capacity, under-capacity, and then yet another system upgrade. Even larger firms, which could afford to provide their own centralized computer networks, are questioning why they must operate their own systems to enjoy the benefits of efficient computing, any more than they must operate their own generators to obtain the full benefits from electricity. The computer utilities offer a services alternative that makes computing a discretionary cost rather than a fixed burden and ends the computer rat-race.

Some years ago we became convinced that the computer utility represented not only a probable future direction for computing, but also the major opportunity for our Information Services business. I would like to discuss our response to this challenge and some of our experiences in implementing the largest and one of the first general purpose information networks.

In 1969 we operated 17 separate computer centers located in the major cities of the United States. Each center provided time-sharing services to customers in its immediate vicinity. This had suited the market well since the primary demand was for engineering and scientific problem solving support. The need for a change, however, was becoming apparent. Our major customers were large corporations that typically had many offices scattered across the country.

Although we offered information services in nearly every major city, our fragmented facilities prevented us from responding to these companies on a coordinated basis. Even more important, time-sharing had been identified as an effective tool for on-line business applications and network teleprocessing. We established a single nationwide network system in 1969 to capture the vanguard network applications. This was an effective first step; however, it was clear that the time had come for unification.

After considering several alternatives for consolidation, we decided to implement a highly centralized, nationwide information network. Our chief reasons for doing so are the following. Since computing is our sole mission rather than merely a support activity, we did not face the severe organizational challenges that another organization might in going to a centralized facility. Second, a series of regional facilities would not solve our capacity problem. On a decentralized basis, our computer capacity came in relatively large chunks. If Seattle was operating near capacity on a single system, a 10% increase in business entailed doubling the computer capacity. Centralization solved this problem in that the increments were spread over a much larger base. Third, it made good sense from a strategic standpoint. Regional computing is something that smaller services companies and in-house systems could do well. We had the national load base and the technical and financial resources to develop a full scale information network. Taking the rather dramatic step of going straight to a highly centralized facility provided an opportunity to lead rather than follow the requirements of the business world.

Having made this decision, we proceeded to develop and deploy an advanced telecommunications network consisting of programmable communications concentrators and frequency division multiplexors. General purpose telecommunications networks with local editing, store and forward logic, error control and retransmission, and redundant data paths, may now be considered relatively commonplace. However, implementing one in 1969 posed many technological challenges. In 1970, we moved from seventeen centers down to three. And last year we consolidated those three centers into one supercenter located near Cleveland, Ohio.

Let's examine some of the implications and opportunities associated with operating a large information network.. First of all, super reliability is not only feasible; it is absolutely essential. One of the most important elements of reliable service is a dependable power supply. One can't expect customers in Los Angeles to be terribly sympathetic about power failures in the mid-west. Consequently, it was essential that our supercenter be protected from brownouts, blackouts, and short-term power fluctuations. Early last year, we installed a full-scale battery-protected power supply that shields the processors from any disturbances in local power. In addition, the concentration of equipment has allowed us to provide an unparallelled form of backup. An extra system is kept in "spinning" reserve, running internal work. It is immediately available to take over for any commercial system which fails. The incremental costs of this power supply and backup capacity are relatively small when spread across a large system base but would have been prohibitive in 17 locations.

Since a centralized computer utility is inherently oriented toward remote processing, it is possible to operate an extremely tight security system at the central site. In addition to 24-hour guard coverage, specially-designed windowless facilities and electronic monitoring and detection systems, we have a stringent no-visitor policy which prohibits all entry to the center except for bonded employees and vendor personnel directly concerned with operating systems.

A computer utility must be widely accessible. The GE network now offers local service in over 250 cities in North America. Even though none of our customers would be interested in all those locations, each application calls for linking together a unique subset of them. In 1970, we also used our new communications technology to link Europe to our network. A communications computer installed in London is linked to our center in Cleveland by both satellite and underseas cable, each operating at 9600 bits per second. Access is then distributed to most major European cities by a combination of concentrators and multiplexors. Operating on a global basis has provided us with some unique challenges. To be useful to the international business community, we must supply service from 8:00 a.m. in Rome to midnight in Los Angeles which is 8:00 a.m. in Rome. The network operates 24 hours a day, seven days a week. Operating across a dozen time zones makes it possible to increase daily system utilization and spread peak usage periods. The savings from this increased utilization more than covers our entire transmission cost.

We have tailored our software so that each customer may select the time zone in which he wishes to operate when he runs his programs and updates his files, so that all time is local to his major operations.

If computer utility is to be successful, it must be established with a clear path for growth and evolution. Although our network consists of over 75 interconnected communication and central computers, it is really just one system. Our facilities for load balancing, adding incremental capacity and accommodating failed components have progressed to the point that customers are no longer aware of or concerned over which processor may be running their programs on any given day. Shutting down the whole network to run preventative maintenance, modify hardware or install new software has become unthinkable. Continuous service is essential.

So far we've been talking about the formal structure of the utility networks as seen by the designers and operators. From the applications developer's standpoint, it consists of a command system, a series of languages, and many processing and communications alternatives which he can utilize to establish his own private computer network. There have been some major improvements in these capabilities in the last few years. Within the GE network they tend to fall into four main areas.

First of all, there has been a tremendous increase in the scope of system capability. As the orientation has changed from local engineering problem solving applications to network business applications, require ments have developed to handle larger programs, to be able to process magnetic tape files, and to be able to talk to the business system programmer in his own language: COBOL. In addition, while network applications tend to be time critical, they frequently involve a significant amount of work that is not interactive in nature. We recently introduced a background batch processing capability which is fully integrated with our interactive service. This will offer customers the combined advantages of a convenient interactive interface for job entry and data base inquiry combined with the full capabilities and economies of a batch processing system for executing major jobs.

The next major area is remote terminal capabilities. Most time sharing systems were developed around the speeds and characteristics of the teletype. During the past two years, however, there has been a major growth in the use of 30-character per second terminals, both because of their three-fold increase in speed, and because they employ the full ASCII character set. In 1971, we upgraded our entire network to operate interchangeably at 10, 13.5, 15 and 30-characters per second. While the teletype will undoubtedly prevail for some time to come, terminals which operate at higher speeds will predominate in business applications to meet the requirement for volume printing.

We recently introduced a high-speed service which provides 2000-4800 baud dial-up service across the United States. When coupled with the new modem technologies and advanced character compression techniques, this will provide remote printing at local computer speeds over voice grade lines.

The third major area of development is manageability. Subscribing to a time-sharing service used to be like signing a blank check. The costs were often neither predictable nor controllable, and it was difficult to assess whether or not the funds had been well spent. An administrative control capability has been implemented which places the whole matter of monitoring usage on a more business-like basis. A designated individual within the customer's firm can control who uses the network and assure that it is used only for authorized, funded projects. Month-to-date expenditures can be checked daily on a projected basis to assure that pre-established budgets are being observed. This is better control than most firms have over the use of local computers.

The final area is a particularly exciting one. We have provided an alternative to costly hardware upgrades for firms wishing to establish or expand corporate computer networks. The alternative is called INTERPROCESSING and it involves linking a customer's in-house batch system to our information network by a high speed communication channel. Our network becomes the interactive computational front-end to the user's in-house batch processing system. Current, time-critical inventory data from a production control application can be transmitted to the network file system where it then becomes instantly available to sales offices nationwide. Conversely, sales and operating data can be captured accurately and efficiently on a daily basis and fed back to the in-house system. Even to firms with the resources and intention to establish a private network, INTERPROCESSING may represent a reasonable first step, since it is a fast and essentially painless way to go on-line with major applications.

The future promises some major developments in the areas of communications, processing and storage. Digital data networks should provide major cost/performance improvements in wide-band transmission, although we will still be dependent upon the current analog switched facilities to reach smaller communities for many years to come. In the processing area, we are moving toward an integrated system cluster architecture. Each cluster will consist of perhaps a dozen main-frames sharing a pooled file base. The foreground-background dichotomy will be further specialized into additional processing roles. We plan to incorporate IBM systems in the background domain and are seriously studying the addition of Control Data 7600s. The cluster will, of course, be able to accommodate hardware failures automatically and will react to changes in demand by dynamically reassigning both work loads and processing roles. Large information networks such as ours will also provide the first practical environment for a cost-effective replacement for the magnetic tape. There are several promising super storage devices on the horizon which, when perfected, will provide a quantum increase in storage efficiency for those centers large enough to absorb their high initial cost.

On the basis of our own experience and developments that we have seen elsewhere, it seems clear that large-scale information networks will ultimately emerge as the dominant form of information processing. Over the shorter range, interconnected regional facilities will provide useful interim solutions for many firms to the institutional problems of transition from decentralized processing.

Two years ago, we demonstrated the practicality of providing interactive processing on a global basis from a centralized facility. We are on the verge of demonstrating the same practicality for genuinely large scale data processing operations. We are about to enter the exciting era when there will be computer utility networks providing a broad spectrum of processing and communications capabilities which are accessible from virtually any place in the world.

ON THE DEVELOPMENT OF COMPUTER AND DATA NETWORKS IN EUROPE

Peter T. Kirstein
University of London
London, England

Summary

In Europe a number of different computer networks are springing up. Amongst the more successful up to now are those for special purposes such as banks, airlines, and national electricity generation and transmission control and billing. Examples of these are compared with the rather halting attempts to start general purpose distributed computer networks, and the more successful centralised ones.

The reaction of several of the European PTTs has been to propose special data networks specifically to handle low and medium speed data and message traffic. The utility of these services for computer networks are discussed, and the potential impact of service bureaux run by the PTTs are considered. Finally some comments are made on the probable way data and computer networks are developing in Europe.

1. Introduction

The subject of this session is 'Teleprocessing - The Utility of the Computer Utility'. The 'Computer Utility', if it could exist, cannot be thought of as the same sort of Public Utility as water, electricity or gas. These utilities each provide services of one universal kind. One may require electricity to be piped into the subscriber at a higher voltage than normal or with greater current capacity, but that is almost the only variation in the basic commodity supplied. The Telephone Utility is somewhat different; now a variety of transmission facilities, a variety of terminal equipment, and a variety of types of service are desired. There is not space here to analyse the exact nature of the telephone utility; suffice to say that in most countries a variety of terminal equipment is being attached to telephone lines, and the range of sources provided is becoming more varied. Progress in the variety of transmission facilities is coming also. In the U.S. these facilities will be available also from suppliers other than the local telephone company. In most other countries the communications authority, hereafter called the PTT, will still supply it.

The 'Computer Utility', as the expression is used in the U.S., is quite a different beast. I first heard the expression from one U.S. private company who wished to promote the idea that the service they were offering could be considered as a utility which met all data processing needs. While the services offered by such bureaux are very useful and expanding in scope, they certainly are far from being universal in their coverage. The fact that very different data processing services are being offered by different service companies to their customers demonstrates this point.

Many of the services do not even compete; to name a few: airline reservation, theatre ticket reservation, current stock exchange prices, bank account enquiry, preparation of invoices, cargo handling documentation, computer aided circuit design, inventory handling, search of records of bubble chamber film data, remote job entry for scientific programs, and prediction of electricity demand over the next hour. At the moment it would probably require at least nine different suppliers to obtain all these services in the U.K., and I doubt if the situation would be different in any other country.

There is no doubt that the service companies offering generalised on-line facilities will increase the range of their offerings. Some will concentrate on offering special packages in particular areas; others will concentrate on those applications which are data base oriented and require access from a wide range of distant places. The signs are that specialised services gradually widen their scope. It is clear that just as the telephone utility in the U.S. consists of a number of companies operating parts of the telephone system but fully interconnected, so airline reservation and bank transfers will be made through specialised data transmission systems owned and operated by an airline (SITA) and banking consortium (SWIFT) respectively. In some cases one airline may run the computer services for other smaller ones, and a consortium of banks may operate together an on-line system; it is not likely that any utility outside the airlines or banks could run the data processing for all of them.

In §2 we will mention briefly the present status of some of the general purpose on-line computer service bureaux in Europe. The survey is not intended to be exhaustive. It does describe many of the longer operations. In §3 and §4 we survey the development of two particular large on-line networks to show how they have developed over the last few years.

Just as the airline networks may broaden their scope to include neighbouring fields like ticket or hotel reservations, so the PTTs providing telephone service may broaden their offerings to include data processing services. The way in which this trend may develop will differ in the different countries depending on legal and financial limitations and inclinations of the PTTs. In several European countries the PTTs are going slowly in this direction. The present activities of the British and German PTTs in providing bureau services are sketched in §5, and the European plans to provide data networks mentioned in §6. In §7 some general purpose computer network activities in Europe are described. Finally, in §8, some conclusions are presented.

2. On-line Computer Bureau Developments

Many of the U.S. Service Companies have extended network operations to Europe. I will say little about these operations; other members of the panel are more qualified to comment on the U.S. activities, and the European operations are never technically in advance of the U.S. The General Electric Network has wide-spread local distribution in many countries using concentrators and leased lines to its large center in the U.S. UCC has a large number of remote batch entry stations and terminals attached to their duplexed 1108 in London. CDC has a number of Cybernet 6000 centres in Sweden, Germany, Italy and the Netherlands; all of these have remote batch links, and there are remote concentrators also for some centers. The centers themselves communicate sometimes over medium speed lines; there is not a fixed wideband network as in the U.S. There has long been discussion of linking the European computers to the U.S. Cybernet, but this link has not been made. IBM and ITT have widespread on-line service bureau operations. Tymshare has a joint operation with the French firm CEGOS, and runs an XDS 940 in Paris. It is planned to connect this machine into the US TYMNET.

The largest European owned commercial service company is that operated by the French firm SEMA. It runs a duplexed CDC 6400-6600 center in Paris, and a CDC 6600 in London. There are a number of batch and low speed terminals attached to these centers. In two cases there are even remote multiplexing arrangements; in one, two batch stations are multiplexed on one line, in another one line is multiplexed for a batch terminal and a number of low speed lines. A number of special applications programs exist in these centers of course. To give a perspective on the size of business carried out by such centers, the total revenue from the two centers in 1971 was about $8M.

The most powerful on-line general purpose service bureau operation in Europe, with widespread communication links, is that of London University; its customers are only research workers, staff and students in the Universities. The central site now has a CDC 6600 and CDC 6400 in one center, and has on-line links to serve 20 batch terminals at medium speed, fifteen CRT multiplexers, four wide-band links, and a number of teletypes. Two of the computers attached to the central site are powerful in their own right - a medium sized ICL 1900 at one college, and a CDC 6400 at another; both support substantial on-line operations. In July a further CDC 7600 will be installed at the center, and medium speed batch links provided to eight other universities. An experimental link between that system and the ARPA network in the U.S. is being considered. In spite of that system having such powerful processors, it would not classify as an information processing utility even for its London University users. The system does not cater for most interactive computing requirements at the central site; nor are there facilities for easy interchange of jobs or files between the other sites supporting interactive work and the central site computers, even though communication links exist. Finally, the general standard of file reliability and fall-back in case of failure of one of the 6000s is not at a level which would be considered acceptable in a commercial utility.

A number of commercial service bureaux offer time sharing and remote batch services on various computers, and some even have multiplexing communications networks to enlarge the caption area of customers. None of them approach the scale of the U.S. companies mentioned earlier. The one attempt at an 'Engineering Design Utility' is in the CAD field. The U.K. Department of Trade and Industry support a Univac 1108 at the National Engineering Laboratory at Glasgow and an ATLAS 2 at the Computer Aided Design Center at Cambridge. These machines support terminals, a remote concentrator, and satellite computers; it is proposed also to link them together. However the actual CPU power of these two is so small, compared to the potential requirements, and the applications supported are so few, that no real CAD utility can result from the combination. Again there is no fall-back facility in case of failure of one of the systems.

Two other service bureaus operations started in the late 60's, with extensive on-line services. The late International Data Highways (IDH) ran two facilities. One, on a duplexed Univac 418, provided extensive stock market services. It held stock exchange prices, and also provided stock portfolio evaluations and a number of related application packages. It provided also a seat reservation system and a stock control system. The other facility was a CTL Modular One, located at a different place, on which an interactive Cobol and BASIC were provided. The two systems were connected to each other and three remote towns by a packet switched system with local number calling in these towns. IDH has recently been taken over by another firm ICFC, which already had a subsidiary IRC which ran a duplexed XDS Sigma 9 and Sigma 5. IRC provides a scientific retrieval service, an advertising analysis package, and a general purpose time sharing service. As a result of the merger, all the applications of both companies are being put on the duplexed Sigma 9 with only some TDM concentrators to several cities. Thus one of the few attempts to set up a computer network by a European service bureau operation has not been able to survive.

The second such service bureau is Centrefile. It started as a real-time stock market quotation system, and then was taken over by a number of companies in turn, now being under the UK National Westminster Bank. It runs a number of application packages; amongst others are a stock broking system, a record keeping and data processing service for building societies (savings and loan associates in the US), and a payroll service. The centre also runs a computer output to microfilm and other services related to banking activities. This bureau is very conscious of the need for reliability and fall-back.

Thus while there is no sign that any European investment is being made on a scale which could be commensurate with a total computer utility, a number of bureau operations are carving out and developing specific application areas suitable for on-line work.

3. Banking Networks

The banks in Europe, and in particular in the U.K., have a strong demand for network services, and we will describe one typical example, Barclays Bank in the U.K. They have two separate centers to which their 3000 branches are being attached, a center for credit card clearing, one for check clearing, and one for a combination of purposes including on-line branch accounts, foreign mail payments and foreign branch accounts. The connection of the branch account is by a centralised system directly to the relevant computer via concentrators and multi-drop lines - the branch terminals being buffered and polled. All these systems have grown up gradually. In two centers the on-line branch accounts are on IBM 360 computers; in another two they are on Burroughs machines; some use IBM and some Burroughs terminals attached to the IBM computers, others Burroughs terminals attached to Burroughs central processors. The foreign mail payments use IBM CRTs attached to an IBM machine. Others use GEC displays attached to a GEC process computer front-ending an IBM 360.

All these and other applications have developed on the different computer systems. To date information transfer between the different centers has been manual, by means of magnetic tape. Some months ago Barclays decided to install a store and forward system, based on six store and forward communication computers. These would be over-connected, and serve to route traffic from the present concentrators to the new communication computers and then to computer centers. Details of the network have not been released at the time of writing, but will be presented at the conference.

Information between banking networks in the U.K. has been transferred, up to now manually, by means of exchanging magnetic tapes in London each day. There is no short-range plan to change this mode of transfer, partly because of the magnitude of the data involved. Between different countries, such a mode of transfer is very slow. After a study financed by a consortium of 73 banks in 11 countries, a decision has been made by the banks to set up a store and forward data network (SWIFT) between the different countries. There will be two store and forward switches, in Brussels and Amsterdams, with programmable concentrators in each of the participating countries (including the U.S.). Initially the lines between these concentrators and the switches will be capable of 9.6K bps, but this will be extended as required. The network will handle a variety of terminals; keyboard/page printers, magnetic tape units, minicomputers handling a number of termials, and front-end communications processors of bank computer systems, such as the switches of Barclay's bank mentioned earlier.

The Central Reserve Banks will also participate. The systems will be operated by a separate organisation owned jointly by the banks much on the same lines as the airlines own SITA. This organisational structure is felt essential because the banking industry feel they must supply their own forms of reliability and security to the network. Discussions are proceeding at the time of writing of this paper with the Telecommunications carriers and the Nort Atlantic telecommunications carriers for approval to set up the network.

Thus Barclays started with a number of separate applications systems, which are being interconnected to form an Information Processing System for the bank. The same procedure is being followed in the other banks. The present development of an interconnecting data transmission system between banks is thus leading to the formation of a banking data processing utility.

A further development, so far unconnected to its banking activity, is that Barclays has gone into partnership with the computer manufacturer ICL to form the jointly owned BARIC service company. So far this company offers few on-line services, and there is no shared use of the Barclays data transmission network. Such shared use is probably not permitted under British regulations. However, this is one way in which a truly sizeable network could provide, for example an accountancy service on a wide scale with little investment in new communication hardware. The example of Centre File given in §2 shows also that in the U.K. at least the bureaus related to banks show signs becoming the nuclei of extensive on-line services to outside customers.

Banks in other European countries, for example France and Italy, have also been active in offering bureau services. In Sweden they have gone much further. An organisation called SIBOL has been formed with the intent of developing a data network specifically for all financial services. This organisation is having an important impact on the plans of the Swedish PTT[4].

4. The French Electricity Utility Operation

Another example of the growth of a network operation may be taken from the development of computing in the French Electricity utility 'Electricité de France' (EDF). A number of special systems have developed over the last few years in EDF. An on-line billing system was developed in one area around Univac 418 and some 23 Remote Entry stations (Univac DCT 2000 and 9300 in type). Three more such regional centralised systems are being established now based on IBM and CII machines, and others based on CII computers will be established later. These centers will be interconnected by a network and attached to the Scientific network at Clamart (see below).

A second type of network is for control of Electricity Dispatching. Here a CII 90-40/90-80 predicts short and medium term load predictions, and passes them on to five regional centers with CII 9300 machines. These regional machines are themselves attached to smaller ones in some power stations by slow (50 or 200 bps) lines. As yet the individual power stations are not controlled on-line through this network. For longer term prediction, an on-line link between the transmission network and the Scientific network (see below) is planned.

The main scientific computing network at Clamart is based on a CDC 6600 and 6500 front-ended by a CDC 3500. Some 16 hardwired remote batch entry stations and five or six faster remote computers are attached to this center. Some links go straight into the 6000s, most enter first via the 3500. The remote computers contain two line printers, a card reader, disk, two magnetic tapes and sometimes teletypes; thus they act also as concentrators.

Schematics of the three networks will be shown at the conference. All are interconnected through the Clamart scientific center. Further enlargement of the network, on the lines of the ARPA network, is being studied.

Although the different application systems at EDF were developed quite separately, it has been shown how the pressure of circumstances has forced an interconnection between the networks. It may now be said that an information utility is developing for the French Electricity supply.

The British Electricity board and the French Atomic Energy authority now sell computer time on a service bureau basis. If the French Electricity Authority wished, and became authorised to supply on-line computer services to customers, it would be in a powerful position to do so.

5. European Post Office Computer Service Activities

In a number of countries (Germany, Netherlands and the U.K.) the Post Offices have started service bureau subsidiaries - either wholly owned or in partnership with private companies. At first sight it would seem that such subsidiaries could be in the forefront of network activity in Europe. The PTTs have large sources of capital, powerful computing facilities, control the data communications services and have large need for network services internally. In practice in no case have these been adventurous in this area yet. The U.K. offshoot, the National Data Processing Service, is particularly large; it operates 2 Burroughs 5500, 12-15 Leo 326 and 10-14 ICL 4/70 and 4/72 (some of these are actually the program compatible RCA Spectra 70s). It is responsible both for the internal processing of the Post Office and external Bureau work. In spite of this large investment, the on-line activities supported are so far few. The Burroughs 5500 are used for terminal-based work in scientific computing and other activities inside the GPO; one of the 4/72 is dedicated to a terminal-based cargo handling system for London airport (LACES), and a little other internal use of the System 4 is terminal-based. No real networking is under way, and none of the external services (with the exception of LACES) is on-line.

The German picture is similar. There the service company, Datel, is owned jointly by Siemens, Nixdorf, AEG and the German Post Office. They now run six Siemens computers with no on-line working, and a GE 265 with time-sharing. In 1973 they will install a larger Telefunction TR 440 with remote remote-batch and terminal facilities. While there are plans for network-based activities, these are proceeding very slowly.

6. European PTT Data Network Activities

The main activity of the PTTs is the supply of communications facilities, and here they are, of course, very active with data activities in mind. Many of the European PTTs are installing or planning separate data networks. They are not constrained, as ATT, to integrate these with the telephone network. There is not space here to enumerate the European activities; a brief survey is given elsewhere[1]. Suffice to say that special networks relying on digital transmission are planned in the U.K.[2], France[5], Switzerland[6] and Scandinavia[4], to be in operation in the late 70s. These will allow switched data access at speeds up to at least 48K bps, and may include packet transmission. A number of preliminary experiments are under way in all these countries. The Germans are installing an asynchronous digital system (EDS), initially partly to deal with telex traffic, at speeds of up to 300 bps[3,7]. Its technology would be valid for much higher speeds, and there are plans to use it in a synchronous mode at up to 9.6K bps. This system is scheduled to start operation in the summer of 1972.

A study of the various papers presented at the Seminar on Integrated Systems for Speed, Video and Data Communications (in which Refs. 4-7 appear) leads to the conclusion that up to the middle 80s separate data networks will be developed in Europe, but these will be increasingly integrated in the normal telecommunications system. There seems little doubt that, in the major European countries, the facilities required for efficient switched data transmission and network operation will be made available. The cost of these services is, of course, still unknown. The time scale is also so long that the computer networks being installed over the next five years will undoubtedly use the present analog telecommunication services. The only exception is Germany, where the EDS system will be used increasingly over the next few years.

The question of restrictions on private or consortium data networks has not been resolved. The airlines (SITA) have set up, and the banks plan (SWIFT) an international industry data network. The Swedes (SIBOL) plan a financial data network, and some airlines and hotel interests had planned a data network (CITEL). On the one hand these networks are being scheduled at a more rapid rate than the plans of the PTTs allow for themselves; on the other they account for a substantial portion of the base load the PTTs would like for their new data networks. Probably the networks now in the detailed planning stage will be allowed to proceed, but in the near future the PTTs will reimpose their rights to limit the introduction of new ones.

7. European Network Activities

The banking, electricity and airline networks mentioned in this paper are not those to which the term 'utility' is normally applied. The European progress on general purpose computer networks is much more halting. Several European PTTs are setting up store and forward computer based data networks to give experimental services, and to help them in the design of the data networks of §6; France, Sweden and the U.K. have announced such experiments. A French computer network on the line of the U.S. ARPA network is being planned, but the exact details of it are not available yet. A number of European governments have also ratified an agreement to set up an international packet switched experimental service, to link initially nodes in France, Italy, Switzerland and the U.K. None of these networks will be operational before 1974.

Two general purpose scientific distributed computer networks have been set up recently on an experimental basis. In France a project jointly funded by IBM and a number of French Institute (Ecole de Mines, CNRS, CEA and Grenoble U) has been set up to connect a diverse set of IBM 360 computers over machine speed lines[8]. The programs allow remote interactive computing and transfer of files. A somewhat similar experimental service has been set up connecting some five ICL System 4 University computers in the South West of Britain[9]. Both of these are very preliminary experiments; in many ways the type of networks mentioned in §2 and §3 have developed further even if they were developed for special purposes. It is significant that the most powerful civil network in Europe to date, that at the Center of European Nuclear Research in Geneva (CERN), has grown entirely in an ad hoc way. The network is different from all others mentioned in this paper, because it is restricted only to the CERN site. That network at my last count contained about 20 computers, and by the end of 1972 will include a CDC 7600, 6500, 6400, 3300, CII 10070, 6 CTL Modular Ones, a duplexed Argus 500, IBM 1800, a PDP-10, PDP-9, many PDP-8s .... . Unlike the London University network, it genuinely distributed. There are at least four separate functional networks as part of that system - one for scientific computers, one for controlling an accelerator, one for interactive computing, and one for running a complex experiment; these, together with some computers for single experimental functions are all interconnected[10].

Finally the British Defence Ministry is planning to set up four large computer utility centers (GRID 77) to offer utility services inside its establishments. This system is being designed to offer adequate fall back in the case of interruption or failure of part of the network.

8. Conclusions

There are a few conclusions to be drawn from the brief examples of European Network Activity presented here. First, there is no sign that the general purpose Computer Service Bureaux will develop so as to provide a universal Computer Utility Service. Their growth is too dependent on internally generated capital, and the external customer base is too difficult to develop; however they will grow and take an important share of any market. The special purpose networks for large industries, such as airlines, banks, electricity and similar fields will grow into real utilities for those fields. Their powerful processing base may also permit some of them to reach over into neighbouring fields. On the whole, however, problems with marketing, demand and restrictive legislation, will hinder this development from proceeding too rapidly. An additional problem is that European governments are not accustomed to making much use of outside bureau services, nor is there too much encouragement for government bureaus to enter into direct competition with commercial ones. For this reason the substantial revenue from this source in the U.S. is largely absent in Europe.

The attempts by the PTTs to develop bureau services are developing much more slowly than those of other bodies. One reason seems to be the rate of decision making in the PTTs, and another the complete difference in marketing force required for their internal and external sales. For this reason the fears of, and restrictions on, their operations imposed in the U.S. by the FCC, are largely absent in Europe.

The European PTTs are developing adequate data communication facilities; present services will have to take the load until 1980, but thereafter an increasing proportion of new applications will be carried on the new networks.

Distributed networks for general purposes are developing very slowly in Europe. It is extremely unlikely that any entity will have developed by 1980 in Europe which can be truly called a total information processing utility.

References

1. Kirstein, P.T., Data Communications, Proc. Jerusalem Conf. on Information Technology, pp. 166-190, 1971.
2. Dell, F.R.E., Features of a Proposed Synchronous Data Network, Proc. Second ACM Symposium on Problems in the Optimisation of Data Communication Systems, Palo Alto, pp. 50-57, 1971.
3. Gabler, H.G., The German EDS Network, Proc. Second ACM Symposium on Problems in the Optimisation of Data Communication Systems, Palo Alto, pp. 80-85, 1971.
4. Lindberg, G., Swedish Plans for a Public Data Network, Seminar on Integrated Systems for Speech, Video and Data Communications, Zurich, paper E5 1972.
5. Deprès, R., Packet Switching in a New Data Network, Seminar on Integrated Systems for Speech, Video and Data Communications, Zurich, paper E2, 1972.
6. Wuhrman, K.E., Systems IFS-1, on Integrated PCM Telecommunications System, Seminar on Integrated Systems for Speech, Video and Data Communications, Zurich, Paper B3, 1972.
7. Bartil, W., et al., A Method of Synchronous Data Transmission between EDS exchanges, Seminar on Integrated Systems for Speech, Video and Data Communications, Zurich, paper E6, 1972.
8. Girardi, S., SOC Project, An Experimental Computer Network, Proc. ACM International Computing Symposium, Venice, 1972.
9. Williams, A.J., The Integrated Computer Networks System, Datafair Nottingham, March 1971.
10. CERN staff members in various articles of the CERN Courier, March 1972.

PUBLIC DATA COMMUNICATION NETWORKS: NEED, TECHNOLOGY AND POLICY

CHAIRMAN:
Lynn Hopewell, Vice President
Network Analysis Corporation
Glen Cove, New York, U.S.A.

PROGRAM COORDINATOR:
Virginius N. Vaughan, Jr.
American Telephone and
Telegraph Company
New York, New York, U.S.A.

PAPERS:

Lynn Hopewell, *"Public Data Communication Networks: Need, Technology and Policy"*

Gordon B. Thompson, Manager, Communications Studies, Bell-Northern Research, Ottawa, Ontario, Canada *"Potential Impact of User/Author Relationships on Public Data Network Design"*

Dieter Kimbel, Organization for Economic Cooperation and Development, Paris, France, *"Planning of Data Communications Networks–Economic, Technological and Institutional Issues"*

August Ohlmer, Ministerialrat, Bundesministerium Fuer Das Post Und Fernmeldewesen, Federal Republic of Germany, *"Summary of the Existing Data Communications Services in Western Europe and Tentative Forecast of New Services for the Next Decade"*

Ken'ichiro Hirota, Director of Total Telecommunication Network, Engineering Division, Nippon Telegraph & Telephone Public Corporation, Tokyo, Japan, *"Public Telephone Network and Computer-Communication"*

DISCUSSANTS:

Jean F. Berry, French Telephone and Telecommunications Users' Association (AFUTT), Marnes-La-Coquette, France

Dr. M. Clayton Andrews, Director of Data Communications, IBM World Trade Corporation, IBM Research Laboratory, Zurich, Switzerland

Lee Talbert, Packet Communications, Inc., Burlington, Massachusetts, U.S.A.

I. P. Sharp, I.P. Sharp Associates, Toronto, Ontario. Canada

THEME: Telecommunications common carrier administrations in many countries are considering strategies for developing new networks for meeting the specialized communications requirements of computer-communication system users. Although all administrations are viewing basically the same user market, substantially different conclusions regarding user requirements and resulting carrier technical strategies are resulting.

The purpose of this Conference session is to provide a forum for the carriers, users and policymakers, to interact and explore this complex environment.

PUBLIC DATA COMMUNICATION NETWORKS: NEED, TECHNOLOGY AND POLICY

Lynn Hopewell
Network Analysis Corporation, Glen Cove, New York, U.S.A.

Radical changes in the potential for information processing are permeating all sectors of society. However, it is clear that the revolutionary changes in computer technology have come into conflict with the evolutionary changes in the telecommunications field. The problems arrising from the interdependence among these two technologies are not because of technology alone, but from the disparity between what is being provided and what could be provided to meet services utilizing both communications and computers. Ordinarily, unfulfilled needs in the market would be satisfied by the entry of new firms. However the conflicts in public policy between a competitive unregulated industry and a regulated monopolistic industry has prevented a rapid transfer of the benefits of new technology into new services.

As a result of these circumstances, the telecommunications environment has experienced increasing pressure for change. The ultimate direction of these changes will depend on the extent to which data users and the computer community participate in and influence public policy.

Nowhere is the lack of such participation more evident than in the common carriers' advanced planning for new data communication systems. First, market studies have been made by the various administrations, but although viewing almost identical market environments, substantially different conclusions regarding user requirements and technical strategies are resulting. Second, little of this advance planning has filtered down, in any useful degree, to the user community. Knowledge of these plans, with the exception of the highly publicized new specialized common carriers in the U.S., remains in the hands of a relatively few specialists closely associated with the carriers. Third, as Kimbel points out in his paper, the carriers' known plans are simply geared to transmit data, and are not applications oriented -"the basic network planning parameters such as the various applications of computers, their telecommunications requirements...are not taken into account.

The primary purpose of this session will be to provide a forum for the interdisciplinary exchange of information which will encourage the participation of more sectors of society in national policy determinations for communications.

Part of the problem is that for many, imagining just how their business might be conducted differently is a difficult task. This difficulty has caused the impact of many dramatic technological advances to have been substantially underestimated by even sophisticated users. We would not want to repeat the error of those in the late 19th century who thought the telephone would have limited usefulness because it provided no written record of the conversation, or of those in the middle 20th centry who predicted a need for no more than a few dozen large computers in the U.S.

Gordon Thompson helps our imagination in the opening paper of the session by suggesting how the communications carriers could recognize the reality of the transactional nature of human interaction and serve as a broker between the consumer and the supplier of services. Although originally a technician, Thompson represents a new breed of communication theorist presently exploring relationships between technology and society.

Dieter Kimbel has recently finished a consulting assignment for OECD to provide recommendations in the telecommunications policy field for member states. He recommends a major "program approach" to identify objectives of social and economic needs likely to be fulfilled by computer communication technology, and to imbed it into the national science and economic policies.

August Ohlmer is in an excellent position to describe the carriers' plans in Western Europe. As Chairman of the International Telecommunications Union's CCITT activities in standards for New Data Networks, he has lead the efforts of the carriers to agree on interfacing standards for new computer communication services.

Kimbel identifys Japan as the only country with a well developed program for national policy determination in communications. Ken'ichiro Hirota brings us up-to-date on Nippon Telegraph & Telephone Public Corporation's thinking.

A panel session will include others well qualified to address our subject. Lee Talbert is President of an organization chartered to use the ARPANET store-and-forward packet switching technology in offering yet another new specialized common carrier in the U.S. Jean Berry is President of the French Telephone & Telecommunications Users' Association and is a consultant in the field as well. Clayton Andrews brings us the enormous experience of IBM as both a supplier and user. I.P. Sharp is involved in a time-sharing service, and is a well known activist on the Canadian computer communication's scene.

POTENTIAL IMPACT OF USER/AUTHOR RELATIONSHIPS ON PUBLIC DATA NETWORK DESIGN

Gordon B. Thompson
Bell-Northern Research
Ottawa, Ontario, Canada

It has been said many times that Public Data Communication Networks will be very important in the future. For this to be true, these networks must have a real utility when viewed by the public. Another way of saying this, is that these networks must recognize the realities that exist in the transactions that the users will engage in. The usual approach to network design relates to the mere hauling of information, and never the transactional relationships between those who will use the public data networks.

Lack of attention to the business relationships between users and suppliers in today's time sharing business can be shown[1] to have produced a system which is far from optimal in terms of generating usage. As a time sharing user, I am obliged to establish a contractiral relationship with the various suppliers of service before I can push the first key. Quite frequently this involves an agreement to accept a minimum monthly billing. It is unlikely that I would have such a contract with more than 1 or 2 suppliers, presuming I am a small user, and so my traffic is confined to only these suppliers. Contrast this with the way today's telephone system handles your long distance telephone traffic which may make use of facilities owned by a dozen or more telephone companies in the course of a single day. Had you been forced to have an explicit contract with each of those individual telephone companies before you could place your first long distance call to a party in their area, you would probably find the telephone far less useful and so would use it far less frequently then you do now.

Clearly problems of standardization exist within the time sharing world that would make the direct application of the implied contract notion used in the long distance telephone case rather difficult to apply directly to the time sharing business. However, it cannot be denied that if the local telephone company were to offer a time sharing computer service to the small user where he could have access to a plurality of suppliers by merely dialing them up, using them, and then paying the telephone company a collated bill, the traffic would be greater than it now is. Such an approach to the time sharing business resembles the way Mr. Sears viewed things when he started his mail order business many years ago. In both cases, Mr. Sears and the phone company would be managing the interface between the consumer and the supplier, bringing them together with the least effort on either's part. Today, in many parts of North America, one can send a telegram via Western Union, and have the phone company bill you for the service. Why not have have it so the small realestate operator, for example, who might wish to use some computer services, could have access to a plurality of software suppliers, each of whom would be rewarded in direct proportion to his use of them, and he in turn would have only to deal with a single retailer, his telephone company?

This seems simple enough, yet unless the legislator is careful, he may deny such an integrative approach through legislative enactments against such things as third party billing. In addition, the network designer may not build in the system sufficient intelligence to be able to handle the complexities of the billing problem. Mere time and distance records are not adequate.

The dispute over whether or not software should be copyrighted or patented has occupied some considerable time and effort. Here again, suitable attention to the transactional or business relationships can resolve the question and provide a more optimal result than either of these two strategies. Copyright is a notion that is much wider than the mere prohibition of copying. When viewed from author's standpoint, it is his life blood. Few authors are as financially endowed as was Lord Byron and can afford to turn all the receipts from their writing activities over to friends and other good causes rather than blemish their aristocratic standing. The Republic of West Germany has recently altered their copyright law to provide a reward for an author every time one of his books is withdrawn from a public library. Copyright is designed to protect the creative author so that he can receive a reward in proportion to the way the public at large values his contribution to their culture.

We who work for large corporations do not understand the vital significance of this to the man who lives by his wits on his royalities. Those of us who are connected with publishing business also know that the way to destroy an author's creativity is to put him on permanent staff. Perhaps that is why the level of creativity that one sees in most technological areas is so abysmally low today.

In the world of music, a sub section of the copyright concept is known as the performance right. Every time a musical selection is played over a radio or television station, it is a transaction and money flows through a network, finally ending up in the hands of the author and the composer. Because the network is large and complex, some of this sticks on the fingers of various switching points within that network. Perhaps the notion of the performing right is more relevant in the computer software situation then either the strict copyright or patent notions. Particularly when the response given to the end users is an individualized one, based upon his enquiry or his initializing action, the need to prohibit illict copying is lessened but the requirement for reward to the creator of the software is not diminished. Copyright can be thought of as a stimulus to good authorship in that it rewards authorship directly in proportion to its popular use.

The performance right notion of music would seem to be particularly relevant in the case of the public

:work, the more so because of the increasing ity that the conventional notion of copyright, icit copying will become less significant. This ion of an interactive or individualistic type it also lessens the viability of patent protec- a suitable means for extracting authorship In essence, the notion of the performance ıen coupled with the idea of a large public data :ation network provides an opportunity for a l adequate system to assure a proper reward for ıt authorship.

therefore becomes important that the public mmunication network not only serve as a means viding rapid and wide interconnection facilities : it also provide a stimulus for increased :ional interaction by suitably rewarding those :ribute content that is remarkable in generating The public data communication network becomes electronic market place that has the intellig- amply reward its better contributors. It a positive Skinnerian environment forcing the ment of proper content. Content that is sensi- users' needs by the actions of an economic place. This means that the automatic message ing procedures need far more than time and e or packet counts. Records must be kept of d what material, what the performance rate is s material, what the transmission charges might :., etc. Accounting records need to be kept for er and each author. Presumably, each author e charged an amount for storing his material on rary shelves, but then would be rewarded in relation to the use that is made of that mater- user on the other hand would pay an amount that :over the royalty paid to the author together ıatever might be required to sustain the system

e world's largest time sharing system, General c Mark II rewards many of its authors in this fashion. Although a portion of the authors material on the G.E. system are maintained as Electric employees, a larger number are dent authors who receive only a royalty on teration of their material. Two classes of rewarded authors exist, those whose material gnificantly to the value of the system to the user and so have their material advertised by Electric as part of the basic General Electric , and those whom General Electric permits to do s on their network but does not advertise for This method of authorship payment is natural s work in the real world. It is something that be considered very carefully by the public data cation network designers. Not only are there tions with respect to the billing but clearly, ral Electric has demonstrated, there are resp- ities and opportunties in the area of advert- nd indexing.

day one hears much about privacy and data net- however, an author is concerned with the e end of the problem and really seeks to have erial used. His greatest fear is that the mat- ill in fact remain private and never be red by users who might contribute to his well Our over attention to privacy in the design of communications data networks may make us over- e most significant aspect that these networks rovide our society in the future, that is ting transactions between people. For every at you worry about privacy, perhaps you should n equal amount of time worrying about how t material can be coaxed into the system in such hat the total utility of the system can be an creasing quantity.

Indexing services alone provide an entrepreneurial opportunity that may hold a very significant key to solving the problems of large information retrieval type systems. Bar-Hillel has pointed out in his book, "Language and Information" that large general purpose information retrieval systems are impossible, but we can't afford to stop trying to build them. Perhaps the key to making these kind of systems workable is by providing an opportunity for a plurality of entrepreneurial indexing services to florish. Again those indexing systems that have widest appeal will be the ones that generate the most revenue, and so provide the largest stimulus for subsequent improvements. It is a kind of positive feedback system that works towards perfection of the strain.

It is difficult with our present perceptions of data systems to conceive of the kind of content that might be generated with the wide application of these notions. Most likely, the content would evolve towards a more natural state then could be achieved by any approach involving conventional market analysis and subsequent software preparation. By using this information market place strategy, the process of evolution would be encouraged, enhanced, and speeded up. It would probably move in the direction of interactive type services permitting you to compose a Stephen Foster tune, talk to your great grandson or great grandfather in a simulated conversation, or paint a Van Dyke picture on your CRT terminal. Clearly the software programs supporting these kind of interactive activities would not be open to illicit copying for they are just not visible. Consequently, the notion of patent or copyright is irrelevant and only the concept of performance right has any meaning. Should you think these content notions a bit wild, bear in mind that it is 15 years since Harry Olsen created the program for composing Stephen Foster tunes on a computer.

Our initial analysis[2] of the character of transactions in this kind of environment suggests that one should optimize around maximizing the royalty payments to the authors of the material in the system. If the royalty payments are too low, then authorship will be at a very low level, and there will be little novelty in the material stored in the system. This would result in users becoming disenchanted with the library and tiring rather quickly with the service. If the royalty payments are too high there will be plenty of new material produced initially, but the costs of accessing by many users would be deemed too high, and little usage would be generated and so the total royalty payments themselves would begin to fall and the whole system would again collapse. The details of this relationship are now being studied by some fairly sophisticated modelling techniques in our laboratory. It would appear that to be really effective, systems using this reward strategy effectively have to be quite large. It therefore becomes important that our early steps in the design of public data communication networks preserve the capability for evolving in this direction for it may become the only significant key to making very large systems have real utility, and may only appear as being important rather late in the evolution of these networks.

Should the transaction density of this software kind of transaction ever become significant in terms of the total gross national product, we could have a very fundamental change in the significance of the service industry in our economy. The service industry is singularly devoid of really effective techniques for converting labor directly into capital. One can build a machine tool and so convert labor directly into capital but there are very few ways in which this can be done in the service sector. Gault McDermid, when he wrote the music for "Hair", converted a few weeks' labor into a rather sizable capital holding. He has done quite well from the performance rights for that music. There are not many Gault McDermids. A large size public data communication network that is designed to act as a positive environment for the creation of software, and rewards its authors in terms of the performances they win, will provide the opportunity for a very large number of people to directly convert labor into capital. This could have an economic significance second only to the American Land Act.

Furthermore, the ecological implications of a significant shift of transactions from the hardware type world to the software type world could be really important in terms of fitting an ever increasing world population into a limited planet and still providing them with useful and wealth creating kinds of things to do that do not tax the planet's reserves unduly. Let us hope that in planning our public data communication networks we do not deny this kind of opportunity by over looking the significance of the transactional relationships between authors and users. It could be one of the most important tools we have to change the forecasts of doom that are so common today. Forrester's World II[3] model does not recognize the capability of the service industry to produce capital. The significant production of capital in a non polluting way completely changes the output of his World II model.

Public data communication networks must then incorporate a sensitivity and capability to do complex billing both rapidly and easily. If this is what we mean when we talk about electronic banking, then this is very important. If on the other hand, we are talking about extending today's banking services, then that is probably relatively trival. It is furthermore important to recognize that the policies established within these public data communication networks with respect to author payments, and other related issues may suplant the conventional copyright legislative activities. Since very little is known in these areas it may be essential that they be freed of legislative hinderance until sufficient experience has been generated to permit policies to be set in stone. The public data communication network could well be the area requiring the most careful consideration in the patent/ copyright field. Essentially the kinds of policies adopted in these networks ought to align positively with the objectives of a proper copyright program, that is, they be aimed at the maximum social good for both today and tomorrow.

REFERENCES

1. "Two Models for the Relationships Between Data System Users, Telephone Companies and Computer Service Suppliers." Unpublished Memo, Bell-Northern Research, Dept. 3H00, P.O. Box 3511, Ottawa, Ontario, Canada

2. "Greening of the Wired City", G.B. Thompson, Reprint, Public Relations Dept, Bell-Northern Research, P.O. Box 3511, Ottawa, Canada

3. "Limits to Growth", Meadows, et al. Universe, 1972

PLANNING OF DATA COMMUNICATIONS NETWORKS—ECONOMIC, TECHNOLOGICAL AND INSTITUTIONAL ISSUES

Dieter Kimbel, Organisation for Economic Cooperation and Development,
Paris, France

Summary

This paper attempts the identification of the technical, economic and institutional aspects arising from the interaction of computers and telecommunications and their promises for the future. In particular those functions or services in an economy related to the collection, processing, storage, manipulation and distribution of information and data. Consequently, the paper differentiates between the applications of computer/telecommunication systems (production factor considerations) and the hardware production aspect of the various components (key industry considerations). In order to realize some of the promises of these blurred technologies (i.e. productivity and qualitative improvements in the rendering of such public services as information retrieval, health care, education, transportation, etc.), one must remember that such systems depend entirely, in technical and economic terms, upon telecommunication facilities. Additionally, the paper highlights the particular and very distinct market environment of the computer and telecommunications industries. Whereas the former is highly competitive (even though there is considerable leadership), the latter is characterized by a monopolistic type attitude both on the service as well as the equipment supply side. Consequently, as telecommunications have become a dominant part of the computer/telecommunications concept, fears have been expressed that the telecommunications industry might become the limiting factor, both for the exploitation of the promises of the merged technologies and for the industrial growth of the systems. To overcome these constraints, experts advocate the abolition of the vertical policy concept and look more towards an integrated horizontal policy. This has been advocated in order to guide the development of both technologies from an applications point of view rather than a mere upgrading of existing telephone networks, which is, at present, common in most countries.

The paper concludes that there is an opportunity for an excellent case study in terms of a Major Programme Approach (i.e., to identify the major objectives of social and economic needs, such as critical public services which are, at present, causing tremendous deficits or bad services). These services could be improved by the integrated use of computer/telecommunication technology. To realize these promises, however, a "horizontal policy approach" seems to be necessary among all partners involved - users, hardware producers, suppliers of services and software). This approach should be under the auspices of a national/international science, technology and economic policy. Finally, as an example, the paper refers to the institutionalized framework of the Japanese government's Informatics Promotion Programme.

Introduction

The emerging concept

A new resource of strategic national and international importance has been developed: computers and telecommunications. The exciting developments of these information technologies may be seen in the fact that the formerly disparate technologies of computers and telecommunications merged to create a new class of combined computer-telecommunication systems.

Due primarily to high potential organisational advantages there is likely to be an increasing trend away from single standing computer systems towards integrated computer telecommunication resource sharing systems. This trend is being fostered primarily by an awareness of improved cost effectiveness, in terms of new technological possibilities and the opportunity this offers to reorganise information flows, particularly in the so-called service sector.

These emerging "interactive concepts" rest on the three following components - computers, telecommunication facilities and terminals. The telecommunication link brings together "raw" computer power and the information in data banks to the millions of different geographic locations where they are needed. The system's overhead costs are shared between many widely dispersed users. This can potentially provide each user with a private computer capability as powerful as the current technology permits, but at a small fraction of the cost of an individually owned system.[1]

The application of such systems, however, extends far beyond the field of computation. In addition to making computer power available in a convenient, economical form, the computer-telecommunication systems can be concerned with almost any service or function related in some way to the collection, processing, storage, manipulation and distribution of information. Consequently, these systems cut across all sectors of an economy: - as a production factor (similar to capital and labour); - as an incentive to further economic industrial growth (key industry approach).

Economic importance

Indeed many authorities emphasize the economic importance of merging computer telecommunication systems. Their economics may be realized on three levels:

(a) through the industrial manufacturing of the various components, which make up such systems, such as computers, telecommunication facilities, terminals (key-industry consideration);
(b) the applications of those systems, particularly in the so-called service sector, which is concerned primarily with the collection, processing, manipulation and distribution of information (infrastructure consideration);
(c) the revenues that might be achieved through the utility mode of operation of such systems.

Key industry considerations

Many authorities predict that the computer-telecommunication industries will become one of the three key growth industries by the end of this decade. For the U.S. it is predicted that some $260 bill. will be spent, before 1980, to build and expand data-processing and telecommunication systems.[2] Of this total - estimated on current prices - capital expenditure for telecommunications alone will be at least $100 billion. The remaining $160 bill. might be required for computer systems and services.

In 1971 the Canadian Science Council noted that the "electronic computer may well be the basis in the 1970's of the world's largest industry after petroleum and automobiles". Their study estimated that the total value of computing telecommunications and software will account for 2 per cent to 5 per cent of GNP in 1979.

In France the computer industry is expected to overtake the automobile industry in dollar volume by 1976.[3]

In Japan, and I will return to this country later, the industry will account for 2.6 per cent of GNP by 1977 and 6 per cent by 1985.[4]

An analysis of the British market estimates that total expenditure for computing in the United Kingdom[5] will approach 4 per cent of GNP by 1980.

Infrastructure considerations

No less exciting are the economies predicted to be achieved through the integrated application of the blurred computer-telecommunication technologies. This in the provision of new "tools" in the service sector of an economy, i.e. including the private industrial divisions of trade, finance and real estate, personal and professional services and general government services; their planning and operative functions being basically the collection and manipulation of information and data. With respect to the automation and productivity improvements - in all industralised economies more than half of the civil employees are engaged in this sector - with the integrated use of computer telecommunication systems, one must remember that:

- it is now technically feasible to bring the full-scale computer complex to anyone in the world served by suitable telecommunication facilities;
- the interaction between the central computer and the remote user is essentially instantaneous, so that the user receives services indistinguishable from those he would receive if he were physically present in the same room as the computer;
- the cost to each user is only a small fraction of what it would be if the same services were provided by individually owned computers;
- the separate achievements and data collections of many individuals can be pooled in large public files so that their contents become simultaneously available on demand to all subscribers;
- the technique of time-sharing has made direct dialogue between man and computer economically practicable;
- the interaction between man and computer allows a harmonius blending of the capabilities of each;
- computer-telecommunication systems have been successfully applied to many fields, such as engineering design, information retrieval, medical diagnosis, problem solving and computer programming.

Consequently, many reputable experts recommend the immediate establishment of plans for computer telecommunication networks offering both traditional services (telephony, telex, etc.) and new broadband consuming telecommunication services. The IED/EIA [6] for example, predicts savings on public services, which at present are either "bad" and/or report tremendous annual deficits, through the application of such broadband networks in the following field:

Public Service	Most Likely Savings
Domestic Air Travel	over $6 billion
Highways	over $6 billion
Police Protection	over $3 billion
Post Office	nearly $6 billion
Fire Protection	over $1 billion
Recreation	over $28 billion

The Interdependence of Computers and Telecommunications

In elaborating plans aimed at realising the promises of telecommunication based computer systems, i.e., proliferating its capabilities to the widest possible range of individuals a number of basic considerations must be borne in mind. These include; the technical, economic, and institutional interdependence of the computer/telecommunication systems.

The Technical Interdependence

Earlier an on-line computer-telecommunication configuration was defined as the merger of the previously disparate technologies of digital computers and telecommunications - primarily, the switched telephone network. Consequently, these systems depend entirely upon the telecommunication lines.

To understand the challenge in telecommunications it is necessary only to remember that the basic design of the existing telecommunication network goes back thirty or forty years and was exclusively tailored for the transmission of the human voice. Consequently, this network, i.e., its main segments the circuits (lines) for transmission and the switching centre to select the path from numerous possible combinations of circuits for the signal to follow, is designed to transmit the frequency range of 300-3400 cycles in analogue wave forms. Distinct from the existing telephone-network-techniques the digital computer "works" on binary coded strings of bits, which cannot be transmitted directly over the public network.

Thus to introduce a computer system into this public network is to couple two different techniques and two different user characteristics. The negative cost effects which the use of the existing telephone network causes for telecommunication based computer systems could be summarized as:

- inappropriate pricing philosphy;
- the need for costly modems to transfer the digital signals coming out of the computer to make it suitable for the analogue network, and to provide the reverse process, i.e., from analogue back into digital form so that it can be received from the recipients terminal;
- the low transmission speeds if compared to the data flow rates of which the central processing unit is capable and which most applications of such systems require;
- the saturation of the existing telephone network;
- the disturbance of the traffic of traditional telecommunication services such as telephony.

Thus, if the promises the computer-telecommunications systems hold for society are to be realized it is necessary to provide adequate telecommunication facilities which meet the quality, quantity and band-width characteristics of these applications.

The Economic Interdependence

Besides the technical deficiencies of the existing telephone network for data transmission and thus for the proliferation of the computer utility concept, the pricing, i.e., the tariffs for data transmission do not seem to be responsive or adequate for the characteristics of these new telecommunications services. They are generally considered prohibitive for most telecommunication supported computer systems and applications. As a result of the technical deficiencies and pricing philosophy of the telecommunication network some negative trends are becoming evident, throwing doubt on the economic viability of the computer utility concept. Unfortunately, as many considerations such as desired terminal equipment, required bandwidth, distance of the remotely-sited user, expected usage, mode of operation and the myriad applications are involved in determining the system costs, it is impossible at present to establish a typical cost-performance calculation. However, it seems possible to identify certain trends within the structure of system costs.

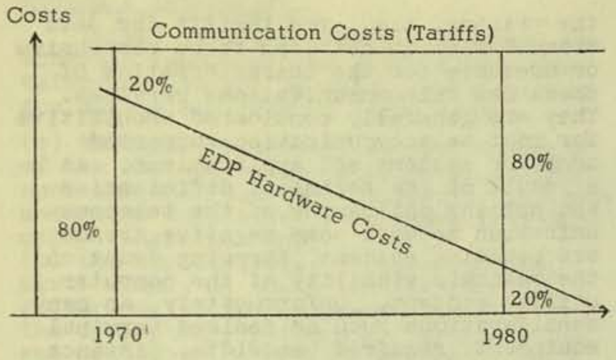
Trends in Data-Processing Costs

Depending on its components the telecommunication-computer system costs might be expressed as the sum of three major elements: costs of CPU's + storage devices; costs of terminals and modems; cost of data transmission and switching. The costs of these three components have followed quite different trends. The computer and the terminal industries have realized massive gains in productivity due to rapid technological changes and the highly competitive environment. Thus, the cost of raw computing power has declined by an order of magnitude every four years and this trend looks like it will hold for a while. With respect to information storage technology, there is even serious talk of "zero-cost" memory development.

Trends in Telecommunication Costs

Communication costs or better the charges of telephone channels to purvey raw computer power and services, did not follow the trends of CPU costs. Computation costs decrease faster (50% every two years) than communication costs (2% per annum, as predicted for the U.S. and Canada). In most other developed countries rapid increases in telecommunication tariffs are reported. Thus the latter become an increasingly significant factor in large telecommunication computer systems. In fact, some present systems already divide costs between communications and data processing equally.[7]

The following graphic is based on the described trends with hardware components which comprise such systems and tries to visualise these trends in orders of magnitude.



Development of Computer and Telecommunication Costs.

Institutional Interdependence

Whatever weakness the previous considerations may contain the fact remains that, due to productivity increases and a highly competitive environment, the EDP segment of the computer-telecommunication concept is experiencing rapid technological change and considerable cost decreases (LSI and large scale production).

By contrast, the telecommunication segment, i.e., the vehicle of this resource-sharing-concept, is characterised by slow innovation and the absence of tariff reductions although, as other contributions of this Conference show, considerable technological progress in transmission and switching media have been developed.

As a consequence, communication costs and the technical performance of the network may prove the limiting, dominant economic factor in computer (information) utilities and thus the strategic variable affecting the speed of its development and the shape of this escalating industry. Therefore, the growth rate of teleprocessing and the quantitative and qualitative economies through its application in the service sector will depend on the rate at which new or better telecommunication services are offered and telecommunication costs fall. Thus the governmental administration which operates telecommunications as a monopoly, and governmental agencies, which have to ensure that the franchised common carriers respond in good time to users' needs, face a tremendous responsibility.

On the other hand, as the technologies are still in an initial stage of future growth, these Institutions, in North America the Federal Communications Commission and the Canadian Department of Communications, in Europe and Japan, the postal and telecommunication administrations, have a unique opportunity of guiding this possible key industry into desired directions.

The conditions for the successful assessment of the computer/telecommunication and related technologies are favourable particularly if we look at the relevant market parameters:

- integrated computer/telecommunication systems are still at a relatively early stage in their development;
- telecommunications are the strategic parameter for all the components which comprise these systems;
- telecommunications in most countries are either directly under a Ministry or are regulated through governmental agencies;
- the legal and institutional tradition of regulation already exists;
- the common carriers represent a considerable macro-economic lever both on the supply side, in terms of services, and on the demand side as buyers of equipment.

New telecommunication technologies and techniques and their economies

Under this heading I will summarise and identify some of the economic effects of technological advances in the transmission and switching components of the public telecommunication network. This might indirectly give an idea of their likely influence on the overall costs of transmitting information and data electrically and/or electronically.

Transmission Technology: Within transmission circuits which, in general, make up 60% of the total plan investment costs, the following developments are pertinent: digital modulation (space and frequency), broad-band circuits, two-way distribution systems and satellite technologies. Their basic economic characteristics are:

- they match bandwidth limitations;
- they improve the feasibility and capacity of existing networks, especially in the local loop network.

Concerning the trunk-network new microwave and satellite technologies promise to eliminate distance as a significant communication cost factor. In a competitative environment this could lead to cost and thus tariff reductions for all telecommunication services, such as telephony, facsimile, T.V. and data transmission, and thus open up fascinating prospects for the information utility concept.

Switching Technology: As prominent advances in switching may be mentioned, the elimination of operator assistance in establishing communications (this resulted in productivity increases in the range of 15-45%[8]). This was followed by the development of electronic switching facilities. Some of the network economies brought about by electronic switching computers are of an operational and functional character. These allow better use of the existing network and will reduce operational costs of stuff and plant. With the introduction of the

DS-Switching computer, the German Bundesost expects network savings will result rom the possibility of operating existng network facilities more efficiently ithout capital expenditure, for at least years, on the existing network.[9] There re, additionally, the economics of space nd maintenance. Another stimulus for ntroducing computers to provide switching unctions and to divide the enormous apacity of new telecommunication linkages nto usable channels (i.e. required bandidth defined by the various applications nd user requirements), is the future osts development of electronic components hemselves.

The Wired City Concept

It was argued earlier in the report hat there are basic technical constraints ithin the existing telecommunication etworks which prevent new telecommunicaton-based services and the handling of he ever growing demands of the traditionl telecommunication services.

To overcome these bandwidth limittions, a seminar discussed the possibiliies of a "wired city" being characterised s having the capacity "for total commnications".[10] According to this study t seems possible, at least conceptually, o increase the telephone system's capaity significantly, by replacing its opper pairs with coaxial cables, thereby orming a switched coazial cable system hich would also allow the reallocation f the overcrowded electro-magnetic pectrum.

It is reported that the British Post ffice is convinced that this is a viable onept and have an experimental system perating in their laboratories at Wembley. t is considered as the forerunner of ulti-service coaxial cable systems to be nstalled throughout Britain during the ext 20 years.

Distinct from existing C.A.T.V. netorks which employ coaxial cable equally i.e. broadband facilities) this configration is two-way oriented and switched. hese experts believe that such a broadand system might be possible within 10-5 years, and described the wired city otal information system as follows.

" A switched coaxial cable system ould have the same philosphy of operation s the existing telephone system", and t could accomodate such services as:

- Advertising; - Pictorial consumer nformation; - Alarm (burglar, power ailure, fire, etc.); - Banking; - Facsimile (documents, newspapers, etc.); - Emergency Communication (hospital beds); - Communication between subscribers and omputers; - Meter reading (utilities); - Distributing of radio programmes; - Shopping from home (see experiment in an Diego); - T.V. (originating and distribution); - T.V. (stored movies available on demand); - Educational T.V.; - Telephone; - Computer-aided instruction; - Picture phone; - Voting, etc.

In Japan, the concept of the wired city will become reality in Tama New Town - a new satellite town of Tokyo. In a $19 million project - to be started in 1972 - some 300 apartments will be interconnected through a coaxial cable network with access to schools, hospitals and other public services.[11]

Similarly there is a recent statement from the FCC which for the first time had publicly noted that "the expanding multichannel capacity of cable systems could be utilised to provide a variety of new communications services to homes and businesses in the community ".

It listed, among the possibilities of a "wired city" concept, such information utility services as "facsimile reproduction of newspapers, magazines, documents, etc.; electronic mail delivery; merchandising; business concern links to branch offices, primary customers or suppliers; access to computers, e.g., man to computer communications in the nature of inquiry and response (credit cheques, airline reservations, branch banking, etc.); information retrieval (library and other reference material, etc.) and computer to computer communications". It also referred to the possible feasibility for C.A.T.V. to develop "capability for two-way and switched services", and, through high capacity intercity communications and computer technology, to become an element in "new nationwide or regional services of various kinds".[12] More recently, in a proceeding involving telephone carrier C.A.T.V. relationships, the Commission again referred to the variety of potential services involving data transmission which could be provided over the broadband cable in addition to C.A.T.V., and to the "real potential that such services will be furnished over regional and national networks consisting of local bradband cable systems interconnected by intercity micro wave, coaxial cable and communications satellite systems".[13]

State of the Art and Planned Activities

On this background information it may now be asked whether and how the relevant authorities in the communication field take up their options in these economic and socially vital spheres. Thus the question at stake is how far individual, industrial-commercial and government needs for telecommunication based computer systems have been elaborated and lead to a balanced strategy, i.e., a communication policy or better a permanent communication policy process which would allow a continuous adjustment due to new developments with the various components and applications. Ideally, such a comparative analysis should also reveal whether the organisational structure of the traditional common carriers and their relationship to the suppliers are appropriate or insufficient to introduce these technologies within a reasonable time frame.

Unfortunately, and this is particularly true for most European countries, there is little information available about what common carriers intend to do, and particularly on what parameters their plans are based, in the years ahead to meet the complex telecommunication requirements. What is available are some terminal modem forecasts and planning parameters which are shown in the following table.

Country	1970/71	1975	1980
Belgium	400	5,000	50,000
Denmark	700	16,000	56,000
France	4,000	50,000	-
Germany	4,300	8,000	68,000
Italy	2,200	13,000	-
Japan	27,000	120,000	-
Netherlands	600	-	-
Norway	300	8,000	-
Sweden	1,100	20-30,000	-
Switzerland	500	5,000	10,000
United Kingdom	12,000 (14a)	57,000 (14a)	234,000(14a) 434,000
United States	105,000 (14b)	820,000 (14b)	2,425,000(14b)

Common Carrier Plans for Digital Data Terminals and/or Modems (14)

Consequently, these forecasts must be interpreted as the quantified demand for telecommunication-based computer systems as the respective telecommunication administrations are considering it.

Some common carriers' approaches to the satisfaction of the telecommunication requirements arising from the terminals enumerated above - which at present show annual growth rates of up to 100% and which promise to be a highly lucrative market[15] - are now discussed.

As it is impossible to give a precise description of the planned services of common carriers and PTT Administrations, I would like to concentrate on the following points:

- Those who will profit from this technology and their number (access-issue);
- What performance characteristics do the planned networks show ?

Although those forward planning approaches are debatable, some criticism is already possible. This even more so as shortcomings with those parameters may be very serious.

The absolute figures of data terminals included in the preceding table look quite promising. They look less so, however, in conjunction with the number of people for whom they are planned.

Socio-economic criticisms

There are good reasons to believe that this approach is too "small" under the various aspects previously discussed; it is to "small" particularly

- with respect to the potential applications and users of telecommunication-based computer-systems,
- in the social context. If access to information and its manipulation capabilities were reserved for a privileged class of people only, this could lead to severe social unrest,
- under industrial considerations, since such a small number of systems and input/output devices would fail to allow large-scale production leading to economies of scale and, possibly, low unit (terminal) costs.

Another criticism is connected with the speed-philosophy which underlies the planned telecommunication networks.

As the transmission speed (bits per second) which the telecommunications link allows is a key question for on-line-systems applications, shortcomings here may be disastrous. It is mostly decided within the telecommunication network which remote operating applications are possible and which not (or only with tremendous additional costs).

Unfortunately, a doctrinal status has developed as to the speeds necessary in a public switched network for the next 10-15 years. There are no sophisticated applications described to be considered for planning but the respective telecommunications authorities believe and use as planning parameters "that 85% to 97% of all data transmission applications will only need low or medium transmission speeds, i.e. 2,400 bits and below"[16] by 1985.

Correspondingly, a British consultant's report on which the Post Office is said to rely heavily for its present planning has estimated that by 1983, 99.9%[17] of all data terminals likely to be installed will operate at slow and medium transmission speeds, slow and medium being defined as up to 10 k bits/s.[18]

Thus, with a switched network designed for these limited possibilities, by 1985 most of the applications using visual displays as for computer aided instruction, consumer information systems would be excluded again. Only "less sophisticated" terminals like printers would be possible and it is no exaggeration to conclude that most common carriers regard the data network requirements as merely an upgrading of their present telex or other switched telegraph networks.

This seems to be all the more regrettable as in most European countries all means of telecommunications such as radio, T.V. and telephone activities are more or less all under the same jurisdiction of government and thus could be much better and more "easily" integrated than, for example, in the U.S., where these high capacity techniques of telecommunications are in private hands and some are even not subject to regulation through the FCC (Cable T.V.). Consequently, the conditions for integrated use of these facilities meet many more "natural" constraints in the U.S. than in Europe or Japan.

I mentioned earlier that these systems might help to ease the time and paper-consuming activities in the service sector. Thus in order to realise the trend towards paperless or less paper "instant" societies, the terminals then in use cannot be teleprinter devices but will be ultrafast facsimile and visual display terminals (such as Bitzer's plasma device) allowing instant interaction, such as dynamic line drawing for CAI, for example. To do so, however, and abstracted from resolution techniques, which still have to be developed, hundreds of thousands of bits have to be transported economically over the network.

This refers particularly to the local loop-network. Consequently, the FITCE-Experts[19] recommend that a switched telecommunication network with capacities of some million bits per second should be envisaged.

In this connection there is another excellent case study of the interdependence of common carrier services and the system designs of related industries: the videophone. As it is now, with its very small screen (5"x5"), it has merely intangible benefits and so remains a gadget. But its screen is small because telecommunication facilities offered by common carriers do not allow a TV sized or any larger screen.

On the terminal side of the computer-utility concept, the comparative results of alternative network capabilities could thus be described in this tabular fashion.

State of the art of the telecommunications network	Terminal characteristics
Low speed	- sophisticated terminals with memory - prohibitively expensive - complicated to handle
High speed	- TV screened interactive real time applications - cheap terminals - widespread use (no concentration of power) - easy operation, etc.

The Interdependence of Network Performance and Terminal Characteristics.

Although the forward planning of common carriers is difficult to debate, it seems to be no exaggeration to conclude that their existing networks and their known plans are simply geared to transmit data rather than being application oriented. In other words common carriers do not look, for example, into the transmission speeds necessary for the introduction of electronic, "instant" mail or newspaper delivery services, or pictorial information services. Neither do they examine the question of what bit-streams the network should provide in order to allow remote computer aided education and the re-shuffling of the various software packages, vital for this individual service.

The same criticism of thinking in technical bit streams only without any reference to applications accounts also for those new specialised common carriers, whic have made famous market studies, showing how the "transmission pie" will be divided into the various economic sectors such as; retailing, manufacturing, securities and finance, or in terms of volume of calls these sectors will place (traffice volume).

Again, although these forecasts might reveal interesting parameters, the basic network-planning parameters such as the various applications of computers, their telecommunication requirements, the type of data - whether digitally coded or in analogue form - to be transmitted, are not taken into account.

To give another example of the vital important of the latter question the following considerations must be taken into account. In the so called "instant world" of the post-industrial societies, where instant information delivery services are common, as are cashless and paperless services which use a pass-word or a credit-card or similar devices for cash transaction, identification and security purposes. I personally would rather see an identification medium which is carried on the person and thus cannot be lost and thus cannot be falsified. For example, voice or fingerprints. However, given the existing resolution and redundancy technology in connection with a reasonable "response time" for transmission of such analogue represented data, broadband communications facilities are needed. Up until now one has been missing such parameters in the planning efforts of common carriers.

Criticisms of the Pricing Philosophy

Although the pros and cons of the forward planning, in technical terms, of the common carriers is debatable, it is quite astonishing that the old telephone-communication-pricing philosophy will be applied secretly to the "new" services, i.e., depending on the distance of the user, time of the "communication" and time of day when the "communication" is placed.

This is the more astonishing as users are looking for a flat rate, or a tariff based on the amount of "information" transmitted and both seem to be technologically possible.

A National Policy Programme

Consequently, if most of the promises discussed earlier are to be translated into action and the most serious dangers avoided, something more imaginative than:

- total laissez-faire and shareholders interests of common carriers;
- the PTT concept, which leaves most decisions to "the consideration" of the administration;

will be needed.

To leave the optimal assessment to the existing carriers and administrations within the present environment is a course which is not welcomed because of the built-in stifling effect with regard to innovation and the lack of imagination and courage to invest sufficiently in this vital future resource.

However the creation of national systems of the scale and the potential described earlier appears economically and politically impossible without a concerted approach on the broadest possible base.

This approach should cover the following segments:

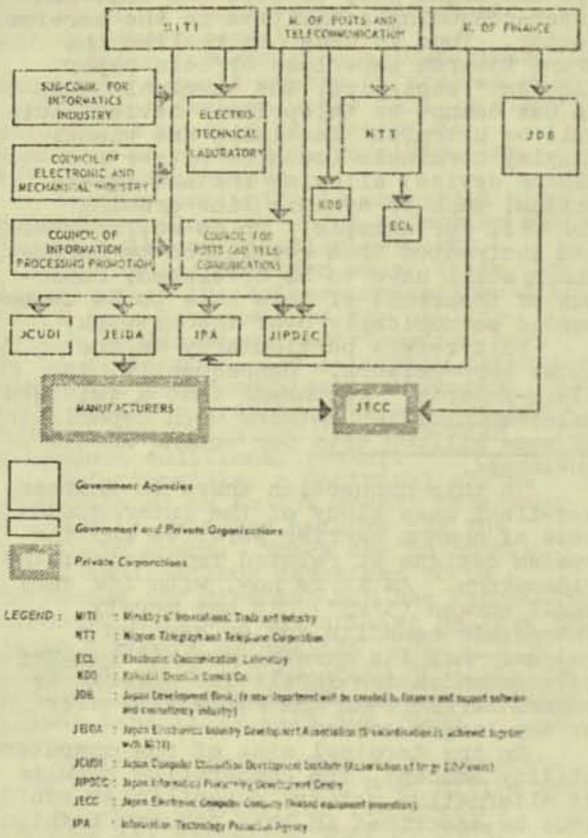
- computer manufacturers (including the wide spectrum of terminals and related component manufacturers);
- common carriers and broadcasters;
- software houses;
- present and potential users.

Consequently, there is need for a "Major Programme Approach" which is trying, like any other national undertaking, such as the Moon Apollo Project, or supersonic aircraft, to identify major objectives of social and economic needs likely to be fulfilled by the computer telecommunication technology, and to imbed it into the national science and technological policy as well as economic policy. Some of the objectives, i.e. applications, as pointed out earlier, could be:

- computer aided instruction,
- the paperless, instant tertiary sector,
- innovations within critical public domains, which actually cause tremendous deficits (postal services for example) and thus absorb urgently needed public financial resources.

At the moment, I am only aware of one nation where on a high policy level such a programme seems to find an echo - Japan. A look at the attached Japanese programme seems to be worthwhile although the social means and ends seem to be omitted, whereas the key industry considerations are dominant.

With respect to the institutions that could coordinate the activities and inventives involved in such a national undertaking to assure the integrated and balanced development, in technical, economic and social terms of computer/telecommunication systems, a new approach seems to be necessary. This even more so as the traditional national monopoly philosophy of the telecommunication industry proves more and more having been a matter of dogma and presumption rather than subject to economic reasoning. Under this umbrella common carriers have become, through horizontal and vertical integration, a tremendous macro-economic power which now seems not to fulfill its services in the public interest.



Institutionalisation of the Industrial Promotion Programme of Information Technologies in Japan.

Some of the symptoms characteristic of the unreceptive environment in the field of telecommunications merit considerable examination: whereas the computer industry has experienced rapid technological change and considerable cost decreases within the last 10-15 years, the fact still remains that the telecommunication network we now have was basically designed and built some 40 years ago.

Depreciation schedules for the computer and satellite industry are some 4-6 years. Typical values with the traditional telecommunication carriers are 25-35 years for local switching gear and 20-25 years for toll service gear.

As a consequence, we might have to wait some generations before 1972 technology is introduced as a means of telecommunications services on a large scale.

Another example of the fact that market forces hardly work in the telecommunication industry may be illustrated by the fact that another Atlantic undersea cable between Europe and the USA is planned for operation by 1976, although it is sufficiently proven that the 1972-satellite generation (Intelsat IV) is economically superior to cable and although the satellits capacity is only partially used.

As a consequence, such a major Programme Approach in terms of applications could hardly succeed unless the parallelling of the telecommunications industry structure, both in terms of the market entry issue of new telecommunications common carriers and the relationship between common carriers and the related equipment suppliers, is carefully examined.

There have been strong pleas for a "new institutional environment" which could be characterised respectively by the strict separation of planning and policy making from operational functions of telecommunication and computing services. Physically the network would integrate the various sub-networks such as telephone, telex, facsimile data and video, cable T.V. and radio broadcasting perfectly, and thus, in technical terms, accomplish perfect systems integrity. The ownership of "parts" of such an integrated telecommunication network however, could be shared by many common carriers. Thus, the whole issue of economies through horizontal and vertical integration is brought into question.

As such a far-sighted set of telecommunication policies would considerably influence the design of computer, terminals and related devices and solve most of the standardisation problems, the market entry of new common carriers would no longer be a technical issue but rather of economic and political nature. The market entry issue could then be discussed according to stated criteria, which could principally be operational (users) needs and economic advantages, such as cheaper and qualitatively better services.

References

1. Parkhill, D.F., "The Challenge of the Computer Utility", 1969, p.3
2. Merill, T., "Business Week", Dec. 6, 1969 and Nov. 6, 1971.
3. "Le Marché de l'Information", L'Expansion, Paris, No. 21 July-August, 1969, p.96-105.
4. Private information from Mr. Shimo, Ministry of International Trade and Industry, Japan, Feb, 1972.
5. Hoskyns Group Ltd., "United Kingdom Computer Industry Trends, 1970-80", October 1969, 10.(55)
6. The IED/EIA Response to the Federal Communications Commission Docket 18397, Part V.
7. Irwin, M., "Computers and Communications - Towards a Computer Utility",1968 p. 172.
8. "The President's Task Force on Communications Policy, A Survey of Telecommunication Technology", Staff Paper IV, PB 184 412, p. 44.
9. Private information from the Fernmeldetechnische Zentralamt der Deutschen Bundespost, Darmstadt, February 1972.
10. Telecommission Study 6(a), "Reports on the Seminar on the Wired City", Dept. of Communications, Canada, 1970.
11. Ministry for International Trade and Industry, EPD, February, 1972.
12. FCC Docket No. 18397, "Notice of Proposed Rule Making and Notice of Inquiry", adopted 12 Dec, 1968, para.8-9.
13. FCC Docket No. 18509, "Final Report and Order" adopted 28 Jan, 1970, para.47.
14. The figures for France, Germany, Japan, U.K. and U.S. represent data terminals, i.e., a connection or interface with common carrier transmission services irrespective of the actual number of connections. Correspondingly the quantification in terms of modems (modulator-demodulator) or of so called data-sets identifies data transmission users on the analogue network but not number of terminals connected to one data set or modem.
14a. Figures for U.K. refer to the years 1973, 1978 and 1983.
14b. Datran Report.
15. Annual Report, American Telephone and Telegraph Company, 1970 and 1971.
16. FITCE Report for the Commission of the European Communities, op.cit. p. 44.
17. Calculated from: Data Transmission, Post Office Telecommunications, U.K.
18. COPEP, Comité VIème Plan, 1970, p. 3: French Administration believes that 99% of the data terminals operate at no higher than 10k bits/s speeds.
19. FITCE - Federation of Academic Telecommunication Engineers of the European Community, Brussels, Commission of the European Communities, Nov. 1969, p. 35

SUMMARY OF THE EXISTING DATA COMMUNICATIONS SERVICES IN WESTERN EUROPE AND TENTATIVE FORECAST OF NEW SERVICES FOR THE NEXT DECADE

August Ohlmer
Bundesministerium Fuer Das Post Und Fernmeldewesen, Federal Republic of Germany

Abstract

The data services offered in Western Europe are already quite numerous. The development in the number of data sets is still in the steep initial phase. Changes in the growth rates can, however, be ascertained. The development since 1967 is shown with the aid of diagrams, first in detail for the Federal Republic of Germany and then as a summary comparison for several European countries. The future development up to 1982 is estimated. The telex and telephone services are used for comparison of the development phases. The picture is completed by a structural analysis of the data sets, based on the figures for 1971. New data services are anticipated as a result of the CCITT studies on separate new data communications networks. These are explained using the basic initial results of the GM NRD; concerned are Draft Recommendations X 1 and X 2. On the other hand further extensive investigations are still necessary. These are explained with the aid of examples from the NRD programme for the period 1972-1976. This whole contribution must only be considered as an interim report; sometime during 1973 the CEPT market study on data transmission will be available.

Introduction

One section of my subject is concerned with new services. The term "new" is relative and must therefore be more closely defined. Thus I should like first to give a description - albeit brief - of the development up to the present time. Of necessity I must confine my observations to Western Europe. My source of information is primarily documents of the CEPT or of the individual CEPT countries. Unfortunately, however, one decisively important source of basic information is not yet available, that is, the CEPT Market Study on data transmissions for the next 15 years for 15 countries. This will not be completed until some time in 1973. Technical development, too, is very much in a state of flux. A new technique may not be identical with new services, but it does present possibilities for these. For these reasons it would, theoretically, be more suitable to postpone today's report on the situation in Europe until 1973/74 or, and this is a more realistic proposition, to repeat it then on a better basis. I ask, therefore, for your understanding if my report today seems to be of a rather interim nature. Enough on the delimitation of the subject.

Development up to the present

You will remember that data processing first began in the USA in the second half of the fifties. Soon also the European telecommunication administrations were faced with this problem, using their existing telecommunication networks; they conducted technical investigations and developments and attended to the drawing up of regulations both for use and organization as well as to the establishment of standards. In the subsequent phase new transmission possibilities and networks came into being, e.g. 200 Baud data network, broad band transmission, use of parallel modems, etc. The importance of international cooperation and standardization for the technical systems and services was particularly emphasized by the increasingly close associa-

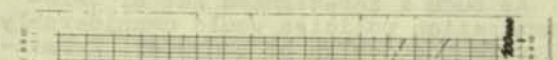


Fig. 1 Number of data sets in the Fed.Rep. of Germany

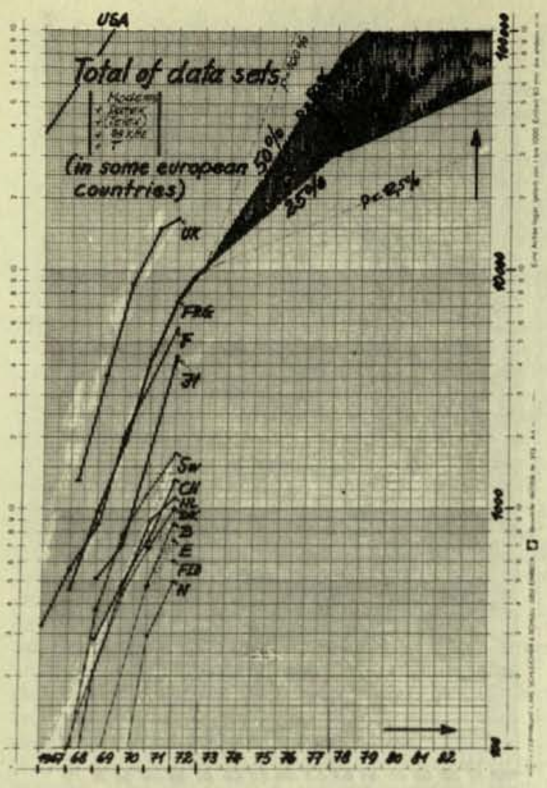


Fig. 2 Total of data sets

tion of the individual European countries. The data services have, therefore, been established for some time in Europe too; the services offered are already numerous but the development, in figures, of the data sets is still more or less in the initial phase. I should like to show you some "home made" diagrams to illustrate this and indicate the order of magnitude.
Figure I - number of data sets - illustrates the situation in the Federal Republic of Germany since 1967. The ordinate is divided logarithmically. The growth rate (p.c.) are indicated as a bundle of rays for easier identification.

1.1. The curve (a) consists of the total of private data sets on fixed circuits (private lease) for the following services: 200 bit/s (proportion approx. 35% in June, 1972); 600 - 1200 bit/s (proportion approx. 27% in June, 1972); up to 2400 bit/s (proportion approx. 20% in June, 1972); over 2400 bits (proportion approx. 18% in June, 1972). There is a steady, steep increase with a growth rate of approximately 100%. For the sake of comparison, the relevant figures for Great Britain are 8,418 data sets at the end of 1971. In France there are 2,956, in Italy 2,588 data sets.

1.2. The curve (b) "modems" comprises the total of post-owned data sets connected to the public telephone network for the following services in June 1972: 200 bit/s (proportion approx. 36%); 600 - 1200 bits (proportion approx. 36%); parallel modems (external station) (proportion approx. 25%); parallel modems (central station) (proportion approx. 3%). Both curves can be compared. The slow change from the steep increase phase to a moderate transitional phase (25 - 50%) is implied, but it is not yet statistically proved. For comparison I shall give the relevant figures for the end of 1971 for Great Britian = 7,971; France = 2,104; Italy = 1,263.

1.3 The development of the Datex service (200 Baud) runs almost parallel to that of the first two services mentioned; however, the change in the growth rate is more regular; for the next 5 years I anticipate that the curve will in all probability go through the central area, i.e. between p = 25 to 50%, or 12.5 to 25% for the period from 1978 to 1982. I see this development as a model case for

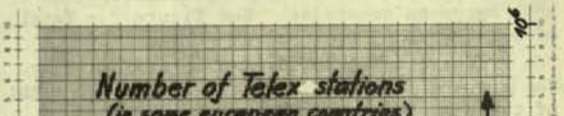


Fig. 3 Number of Telex stations

comparable "new data networks" of the future. At the end of 1971 there were 147 Datex stations in France. To complete the picture 3 further data services should be mentioned:

1.4. Telex traffic with computers, i.e. Telex stations carrying out data operations (50 Baud) This service is still small and it is difficult to give statistics. Further figures for the end of 1971 are only available for France (49).

1.5. Provision of broad band circuits (48 kHz). The development of the figures regarding data sets for this service is still in a state of flux. Roughly the same figures are available for Great Britian for 1971 (49). This is followed by Switzerland (10), Sweden (4) and Spain (2).

1.6. Use of telegraph circuits in traffic with computers (leased circuits up to 200 Baud; curve "T"). The number of sets is relatively high, but the slow development tendency reduces their importance and future share in the entire data transmission. The situation is similar in Italy (478), France (394), Switzerland (164) and the Netherlands (76). After this initial survey with the aid of a detailed description of the situation in the Federal Republic of Germany I move on to a summary comparison of some European countries. See Figure 2 - Total of data sets.

The ordinate is again divided logarithmically. The growth lines are indicated to the top right of the example of the Federal Republic of Germany; the development area which I foresee for the next 5 years lies between 25 and 50%. The respective total of all data sets of a country since the beginning of 1968 is indicated. We have no comparable figures available for before this time. At the top is Great Britain (UK), followed at some distance by Federal Republic of Germany (FRG), France (F) and Italy (It). The growth rate of the development hitherto is approximately 100%; i.e. these countries are still going through the initial phase. However, a transition can already be ascertained in the case of UK and FRG to values of p = 25 to 50% for the next 5 years etc. Further down the page come the following countries, close on each other's heels: Sweden (Ss), Switzerland (CH), Netherlands (NL), Denmark (DK), Belgium (B), Spain (E), Finland (Fld) and Norway (N). The sequence does not always correspond to the size and importance of telecommunications in these countries. There, too, a typical initial phase is ascertainable. As an overall impression it can be said that the development tendency in Western Europe is largely uniform and that the initial phase should soon be over. To enable a comparison with the development in the USA you will find in the upper left-hand corner a curve marked "USA" for the years 1967 to 1969. The real difference in time (delay) between the individual European countries and the USA could probably only be ascertained with the aid of corrected figures; the correction factor is obtained, for example, from the comparison of the gross national product of the USA with the gross national product of each European country. In the case of the Federal Republic of Germany

$$f = \text{approx. } \frac{991 \times 10^9 \ \$}{186 \times 10^9 \ \$} = \text{approx. } 5.3$$

(valid for 1970)

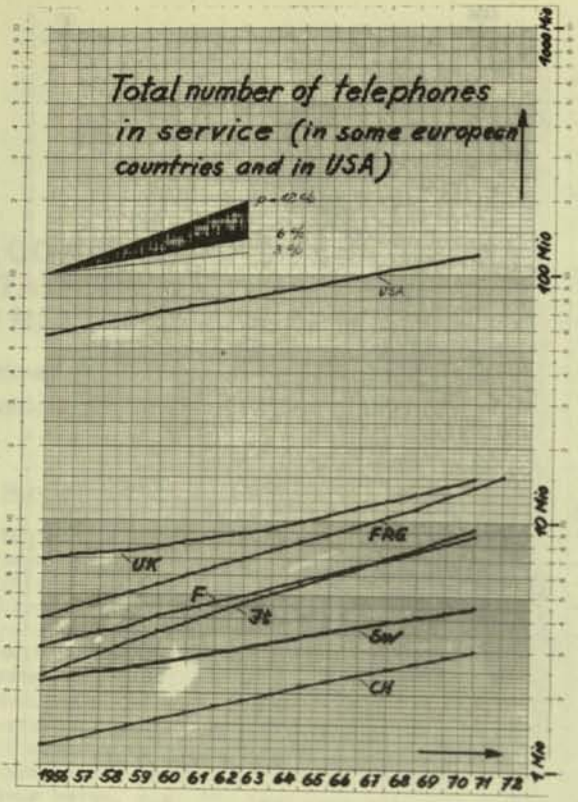


Fig. 4 Total number of telephones in service

The graphical evaluation gives a result of roughly 5 years time difference. I have tried, with the aid of Figure 2, to warn against an over-evaluation of the future development in Europe. To strengthen my point I should like to draw your attention to the telex service which is very highly developed in Western Europe and which is unequivocally to be considered a purely commercial service. See Figure 3 - Number of telex stations. The figure covers 6 countries which have at present the greatest amount of data traffic and, for comparison purposes, the development in Japan. Italy developed its telex service relatively late. One can still ascertain the end of the steep initial phase (end of 1963) and the transition from 1964 to 1969 to a

steady development. A balanced, steady development is indicated by the development over many years in the Federal Republic of Germany (curve FRG). In my opinion the other countries, such as Great Britain (UK), France (F), Switzerland (CH), Sweden (Sw) and Italy (It) are also moving gradually in this direction Even an overseas country like Japan will probably not be spared this unavoidable development To sum up with regard to Figure 3 I should like to say: 'The telex service in Western Europe shows a way in which the development tendencies of the previous data services can be realistically estimated. Now, just a short look at the telephone service of the 6 aforementioned countries plus the USA. See Figure 4 - Total number of telephones in service.
Already there is a steady development everywhere The sequence of the European countries corresponds practically to that in the data service, as indicated in Figure 2. You will have noted that I have made no mention of the international data service. For this we should wait for the results of the market study. However, I should like to show you one diagram giving information on the entire outgoing telephone and telex traffic of the Federal Republic of Germany.
See Figure 5 - Development of the telex service as compared with the telephone service.
This shows a random sample for the month of July In the top band you will recognize how both services run almost parallel for many years. The relationship telephone: telex fluctuates between 2 : 1 to 2.5 : 1. I would remind you of what I said earlier about the telex service as a guide line. Perhaps there is a possibility here that some day an informative correlation will also be computed for the international data traffic. The lower curves illustrate individual traffic relations.
Figure 6 - Structure of transmission speeds - brings us back to the present state of affairs in the data services in European countries. It follows on from Figure 2 - Total of data sets - and presents a percentage distribution of transmission speeds for 6 important countries; all figures apply to the end of 1971 and were calculated from the number of data sets. In the formation of the groups sets attached to fixed circuits and sets attached to public networks were combined. The result is the following distribution:: up to 200 Baud - from 37.4 to 69.2% - on average 54.9%; 600/1200 to 2400 bit/s - from 30.6 to 62.3% - on average 42.8%; 4800 to 9600 bit/s - from 0 to 6.3% - on average 2%; broadband (48 kHz) - from 0 to 0.7% - on average 0.3%. On the extreme right of the diagram you will see figures which are taken from papers of the "Zurich Seminar" of March, 1972 and which are intended as an estimation for 1980. These figures seem quite plausible in an optical comparison with the results for 1971. On the other hand, certain schools of thought anticipate a considerably higher proportion for "low-speed" in the future. This concludes my section on "Development up to the present". I should now like to proceed to the second part of the subject.

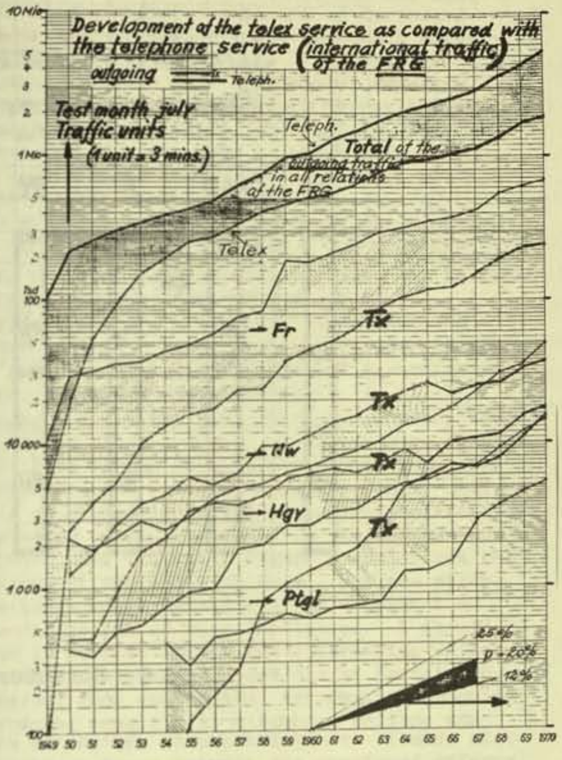


Fig. 5 Development of the telex service as compared with the telephone service (outgoing international traffic of the Federal Rep. of Germany)

New Services

Here we are concerned with a non-technical description of new services which are envisaged for the next decade. This task is in conformity with the deliberations of the CCIT with regard to the need for new data services and with regard to the possibility of fulfilling the requirements by means of digital switching systems. The CCITT Mixed Working Party "New Data Networks" (GM NRD) has undertaken relevant studies of a separate data network and, at the end of this study period, submitted the first results and proposals for the approval of the Vth Plenary Assembly (December 1972 in Geneva). The essential requirement was to reach agreement on the basic services and facilities to be provided in the future, independently of the switching or transmission technology that is used nationally. This independence is important as there are two main proposals for new networks on hand in Europe (Federal Republic of Germany and Great Britain). The CCITT documents which are of interest here are documents NRD-99 and NRD-101 of May, 1972.

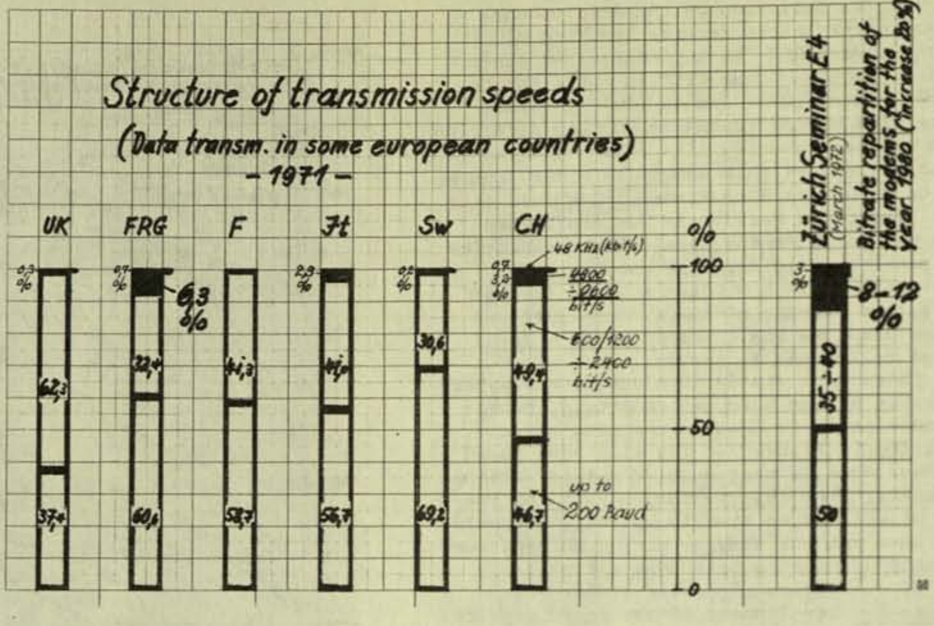


Fig. 6 Structure of transmission speeds

User classes of service and data signalling rates
for public data networks

Class	User data signalling rate and transmission mode	Address selection and service signals
1	200 bit/s, 11*) units/ char. start/stop	200 bit/s, Alphabet No. 5
2	50-200 bit/s, 7.5-12 units/ char. start/stop	200 bit/s, Alphabet No. 5
3	600 bit/s, synchronous	600 bit/s, Alphabet No. 5
4	2400 bit/s, synchronous	2400 bit/s, Alphabet No. 5
5	9600 bit/s, synchronous	9600 bit/s, Alphabet No. 5
6	48000 bit/s, synchronous	48000 bit/s, Alphabet No. 5

*) Usage in accordance with Recommendation V.4.

Fig. 7 User classes of service and data signalling rates for public data networks

Doc. 99 contains the two important Draft Recommendations; X 1 - User classes of service and data signaling rates for public networks and X 2 - Recommended user facilities available in public data networks. Doc. 101 contains the study questions for the period 1972 - 1975.

Figure 7 shows the main contents of Draft Recommendation X 1 in the form of a table. I shall quote the basic objectives from the Draft Recommendation: "The establishment in various countries of public networks for data transmission creates a need to standardize user data signaling rates, transmission modes, address selection and service signals; such standardization facilitates international interworking, bearing in mind that in the data mode the network should be code insensitive. Recommendations in the V Series already standardize data signaling rates for synchronous data transmission in the General Telephone Network and modulation rates for modems. These rates are, however, not necessarily most suitable for networks devoted entirely to data transmission and this leads to the requirement for an additional standard. Bearing in mind the desirability of providing sufficient data signaling rates to meet users' needs, the requirement to optimize terminal, transmission and switching costs to provide an overall economic service to the user, the interaction between users' requirements and tariff structure, the C.C.I.T.T declares the following view: Users' data transmission requirements via public data networks may best be served by defined user classes of service" etc. Only 6 user classes of service are still in existence. Only classes 1 and 2 are planned for "start-stop". Classes 3 to 6 are intended for "sychronous". The phase 50 bit/s originally also envisages was dropped after long discussion as being untypical for data networks. It was agreed that this phase in data networks should not be envisaged in addition to an already existing, separate telex network. In those networks in which the telex and the data services are combined the speed of 50 bit/s is a component of the integrated network in any case. As an example of this I would mention the new EDS system of the German Administration. In the long term the public data network planned by Great Britain will also carry telex services. Class 2 will provide, for example, for operation at the following speeds and code structures: 75 bauds (7.5 unit); 100 bauds (7.5 and 10 units); 110 bands (11 unit); 134.5 bauds (9 unit). The long-range development direction for start/stop character-oriented further study. The proposals of some administrations and organizations to expand class 2 (50 to 200 bit/s) to 300 bit/s did not meet data transmission is expected to be toward synchronous operation. The user data signaling rates shown in classes 3 - 6 are the maximum possible user information transfer rates; users may also operate at particular compatible lower rates, for example, half the rates quoted. The ways in which this is to be accomplished are to be the subject of with general approval. The same applied to the proposals for the additional introduction of 4,800 bit/s, synchronous. But I should like to point out that these points can be taken up again in the next CCITT study period. I refer here to Doc. NRD-101, Question 1/NRD - point A (urgent study point) - Standardization of user data signaling rates and transmission modes in public data networks.

Figure 8 gives the main part of the Draft Recommendation X 2. A number of performance characteristics of new data networks have been compiled here and classified as necessary, desirable or optional. All these characteristics are envisaged or representable in the German EDS network, for example. Figure 8 summarizes the outcome of the study to date. But the study continues under Question 1/NRD - point B - Standarization of user facilities. I should like to mention just a few interesting points from this question: speed and/or format conversion, code conversion, information (enquiry) services, repeat call attempts, polling and data collection at users request. I must also mention at least briefly the continued efforts to attain application flexibility. These, too, serve only the one aim, that is, to compute all the system parameters of a data network. I include in this especially Question 1/NRD - point D - Call set-up and clear-down time. There exist already certain detailed ideas on this subject but these require further study before they can be made into a Draft Recommendation. Questions 1/NRD - point F - Grade of service and 1/NRD - point G - Transmission quality, belong, among others, to this section. You will find all these questions laid out in detail in CCITT Doc. NRD-101. Three facts emerge from this rather compressed and incomplete report. They are: (1) The data services are developing very rapidly in Europe too, (2) Europe is working intensively and systematically on the development of new data networks and services together with other interested countries, and (3) Much has already been achieved, but there is much more still to be done during the next decade.

User facilities in public data networks

Title of facility	Availability
Symmetrical duplex	P
Bit sequence independence	N (User classes of service Nos. 1-2) P (User classes of service Nos. 3-6)
Manual and automatic calling and answering	P
Direct call	E
Abbreviated address	E
Closed user group	E
Closed user group with outgoing access	O
Remote terminal identification	O
Multi-address	O
Delayed delivery	O (User class of service No. 1) N (User classes of service Nos. 2-6)

Note : P = A facility supplied as part of the user class of service.

E = A facility recommended to be available to the user in public data networks

O = A facility which may be provided in public data networks

N = Not required

Fig. 8 User facilities in public data networks

PUBLIC TELEPHONE NETWORK AND COMPUTER-COMMUNICATION

Ken'ichiro Hirota
Nippon Telegraph & Telephone Public Corporation
Tokyo, Japan

Summary

The transmission media for computer-communication in Japan are at present mainly leased circuits. The use of public telephone network for computer-communication, however, will soon be completely open to general computer users with the introduction of a new message rate system.

In this connection, it is expected that to a considerable extent, Japanese telephone network will meet most of the present data transmission demand, at least from the viewpoint of error-rate, since the majority of exchange equipment in Japan is crossbar-type and the effect of exchange noise on data transmission will be practically negligible.

An important role of the public telephone network in the future of computer-communication is considered as general telephone subscribers can be provided with some computer-communication services. As an example, NTT provides its push-button telephone subscribers in Tokyo and Osaka with a calculation service called "DIALS."

On the other hand, a basic limitation to data transmission via the public telephone network is the speed restriction due to the voice-band, and to meet the rising demand for high-speed data and other wide-band communications, NTT plans to introduce new facilities for wide-band switched services by adding electronic switching equipment to the existing telephone network.

Since the public telephone network, even if associated with the wide-band facilities, will still be on an analogue and circuit-switching bases. NTT therefore plans to extend the scope of digital services on leased circuits. Also, basic research is under way at NTT's laboratories for a switched digital network with a store-and-forward facility.

Data Transmission Facilities for Computer-Communication in Japan

Public telecommunication facilities for domestic use in Japan are exclusively provided by NTT (Nippon Telegraph & Telephone Public Corporation).

The data transmission media for computer-communication systems in Japan are, in principle, dedicated circuits leased from NTT. As of the end of March 1972, approximately 13,500 such circuits were in use for 295 computer-communication systems operated by a variety of governmental agencies or private enterprises. Data transmission speeds on these circuits are 50, 100, 200, 1,200 and 2,400 bit/sec. Detailed figures thereof are shown in Annex 1. It has been a general tendency this past few years that the yearly increase in the total number of circuits is 40 to 60% and that the basic speed is moving from 50 bit/sec to 1,200 bit/sec.

Besides the above-mentioned systems, NTT provides its customers, as of May 1972, with thirteen computer-communication systems serving a total of 950 terminals, mainly for nationwide and/or public use, and with two DIALS (Calculation by Telephone) systems for approximately a hundred-thousand push-button telephone subscribers in Tokyo and Osaka. Of these systems operated by NTT, seven including two DIALS, utilize the public telephone network which primarily serves twenty million telephone lines.

The use of the public telephone network for computer-communication, which has been legally limited to those systems operated by NTT, will soon be admitted to general computer users in accordance with the recent revision of the Public Telecommunication Law. The revision includes the introduction of a three-minute message rate system into every local telephone area. The said release of the public telephone network for computer use will be effective immediately after the introduction of the new telephone rate, which will be made area by area between November 1972 and August 1973. The new telephone rate, although primarily for other reasons than computer-communication, is also considered instrumental in moderating possible congestion in the telephone network due to the very long duration of some computer-communication calls.

It has been legally defined that, before connection of a given computer or terminal with the telephone network be allowed, their interface characters should be recognized by NTT to conform to certain technical standard for avoiding any undesirable effect on telephone traffic. Also, the traffic conditions of the telephone exchanges concerned will be checked in advance by NTT.

This may in some cases result in the installation of additional exchange equipment prior to the connection of a computer.

As for the high-speed data transmission, 48 kilobit/sec data-sets based on CCITT Recommendation V35 are in service on non-switched leased lines. It is planned that this type of service may be available on switched circuits also, for users in limited areas following the introduction of DEX (NTT type electronic exchange equipment).

Besides the data transmission facilities NTT, in cooperation with Japanese manufacturers, is developing a very large-scale computer equipment called "DIPS" (Dendenkosha Information Processing System), with the aim of making it the standard computer equipment for a variety of computer-communication services to be operated by NTT. The first commercial use of DIPS will be in Tokyo next year for the DEMOS-E service, i.e., a conversational time-sharing service for approximately five hundred user terminals.

Japanese Telephone Network

Telephone network facilities are not primarily aimed at the transmission of digital signals. In fact, according to the experience of most countries where computers are already connected with the telephone network, it seems to be the general view that the network facilities are insufficient for data transmission. The main reasons therefor are considered to be the speed limitation within the voice-band, possible errors due to exchange noises and group-delay distortion, and postdialling delay. Of these the noise encountered in step-by-step exchanges may be the most significant.

The above-mentioned shortcomings also exist basically in the Japanese public telephone network and it is considered difficult to guarantee completely the quality of data transmission over the entire telephone network, even in the case of low-speed data transmission within the voice-band. Nevertheless, as for the Japanese network, there are also certain other aspects perhaps a little different from those in other countries, which might have a less harmful influence on the data transmission quality due to NTT's experience in operating its systems. These points are described below.

(1) In the Japanese telephone network, practically all the trunk switching equipment and two-thirds of local switching equipment are of the crossbar type, the remainder being step-by-step. Moreover, in the coming five years, the volume of exchange equipment will be doubled with the addition of solely crossbar or DEX equipment and with the removal of some old step-by-step equipment.

(2) Consequently, it would in most cases be feasible even at present that any computer or terminal could be switched to each other solely through crossbar equipment by exceptional use of remote exchange lines.

(3) It is generally known that the exchange noise in crossbar exchanges has practically no serious effect upon the quality of data transmission. This was also proved to be true from NTT's experience of operation in the past two years of the computer-communication systems, DEMOS (Scientific & Engineering Calculation) and DRESS (Sales & Inventory Management), which utilize the telephone network. Furthermore, measurements are being made of the bit-error-rate of 1,200 bit/sec data transmission through actual crossbar-switched connections. The results so far showed that the average bit-error-rates would be lower than 10^{-6} for local connections and 10^{-5} for five-carrier-link connections respectively.

(4) The postdialling delay varies depending on the type and number of switching points. For a connection through four crossbar exchanges, which is considered standard in Japan, the delay is about five seconds and this figure will become less with the introduction of DEX equipment.

(5) All the line signalling systems in the Japanese telephone network are of the outband type and the voice-band is fully available for data transmission.

In consideration of the above, it is the author's view that the Japanese telephone network could meet the requirements for computer-communication to fairly a wide extent, as long as the demand for high-speed data transmission beyond the voice-band is not high, and that it would be premature to start construction of a separate data network at this moment.

Computer-Communication Service for Telephone Subscribers

Although the public telephone network could not be made completely suitable for data transmission, it will not only continue to be a powerful means of transmission but also play an important role in future computer-communication. In other words, with the rapid progress in this field of communication, the need principally would arise for the advantages of computer-communication, which are at present available to specific enterprises, should be shared perhaps by many people. From this point of view, the advent of computer-communication services for general telephone subscribers is considered very desirable.

It may be generally said that a computer-communication service for telephone subscribers should satisfy at least the following conditions:

a. Available at any time without special subscription procedures
b. Simple manipulation without any special skill or training

c. Low cost
d. Universally welcomed by the majority of subscribers
e. With the similar grade of availability to that in the telephone service

As an example of such types of service, NTT provides all push-button telephone subscribers in Tokyo and Osaka with a calculation service called DIALS, a simple description of which is given in the Annex 2.

In designing the DIALS system, the following procedures were followed in the telephone network to meet the above-mentioned requirements:
1) The service charge is determined proportionate to the holding time of connection with the computer centre and as recorded on the subscriber meter exactly as in multi-metering trunk calls.
2) For the above charging method, the access code to the centre is chosen from the spare codes in the national numbering plan.
3) Consequently, the routing of DIALS calls becomes similar to that of trunk calls and limitation of service areas and low-loss transmission for DIALS calls were achieved.

Furthermore, the following considerations were given to the computer system side:
1) In principle, the terminal equipment is an ordinary push-button telephone and the output voice-response is given in the form of mechanically compiled word and/or phrase groupings. A simple keyboard or a printer can be available as required by the subscriber.
2) The computer system is on a duplex basis with an automatic switch-over facility from normal to stand-by. As a result, the service availability so far has been 99.99%.
3) To provide as many subscribers as possible with the service, the system is designed so as to serve 500 simultaneous calls at maximum and the internal processing time for a given calculation is limited to within 200 milliseconds. Also, care is taken in designing the supervisory programmes to minimize the differences in the response times for a variety of calculations.

Aside from DIALS, various computer-communication services such as information inquiry or retrieval is likely to appear some time in the future. The system configuration for these future services may be basically similar to that of DIALS as set forth above.

Wide-Band Circuit Switching Service

Over the public telephone network, the upper limit of data transmission speed is in general 2,400 bit/sec, and is possibly around 4,800 bit/sec even under special arrangement. From this point of view, some other means than the telephone network must be provided for the computer-communication systems with higher data transmission such as 48 kilobit/sec or 240 kilobit/sec, which are expected to grow in demand.

High-speed communications in Japan, including high-speed data and high-speed facsimile transmission utilize, for the present, non-switched leased line facilities. It is, however, considered highly desirable that these high-speed communications be served over a switched network, since these carry high-density information and generally consume very little time.

On the other hand, from the common carrier standpoint, the present volume of high-speed communication traffic may not be considered sufficient to economically justify the immediate initiation of a separate network.

From the above points of view, NTT plans to upgrade the telephone network by introducing new facilities. It is projected that all principal cities throughout Japan will be equipped with DEXs (NTT type electronic exchange equipment), which will exchange not only ordinary telephone circuits but also, where necessary, groups of wide-band circuits.

A subscriber who has a terminal to DEX would be able to set up a connection over a series of wide-band circuits in addition to normal telephone circuit connection with another subscriber in the same category, to transmit high-speed data, etc. To obtain wide-band circuit connection, the subscriber, before dialling a number, would have to dial a two or three digit prefix code corresponding to the desired frequency band. For instance, the prefix "✱✱" for the videophone service and "XY✱" (XY is not fixed yet) for the 48 kilobit/sec data transmission would have to be dialled respectively, prior to dialling the desired number.

The above-mentioned facilities will be put under commercial test next year for a limited number of subscribers in Tokyo and Osaka. Also DEX, which should play the most important role in the network, will be introduced by 1977 into at least one hundred cities.

Basic Researches for a Switched Digital Network

It is considered for the time-being that the public telephone network may be suitable for low-speed data transmission to a considerable extent and that the wide-band switched service described in the preceding section may be a fair solution for high-speed data transmission. The concept of the services, however, is still on an analogue and a circuit-switching bases and the store-and-forward facilities for, e.g., the code or speed conversion, packet-switching, etc. are not yet included in the services.

Considering these, NTT will continuously make effort to improve data transmission services. As a first step, NTT is planning for the near future, to extend the scope of digital services for leased circuit users, both in speed and in distance, since the present digital services are limited to 50 baud telex network and 100 bit/sec telegraph-type leased circuit services.

On the other hand, basic research toward a data switching network in the future is being conducted at NTT's Electrical Communication Laboratories. The research includes a variety of digital transmission techniques, time-division electronic switching and packet switching. However, the ultimate network configuration together with the period for initiation of such a data network is still nebulous, since it depends on the volume of demand and the progress of the technologies concerned.

In fact, the volume of data traffic at present may be considered too small to establish a separate network. As an example, the number of long-distance leased circuits used for data transmission in Japan is around 1% to the total number of trunk telephone circuits. Considering, however, the expected rapid growth of computer-communication and video-communication, it would not be too far in the future that a separate end-to-end digital network could be economically justified.

ANNEX 1

LEASED CIRCUITS FOR COMPUTER-COMMUNICATION IN JAPAN

1. As of the end of March 1972, 13,512 circuits leased from NTT were in use for 295 computer-communication systems. Detailed figures therefor are shown in the following table:

Data Signalling Speed (bit/sec)		Number of Circuits	%
Telegraph-Type	50	6,318	47
	100	126	1
Telephone-Type	200	3,623	27
	1,200	3,248	24
	2,400	197	1
Total		13,512	100

Notes: 1) The above figures are solely for on-line systems use.
2) Telegraph-type 100 bit/sec service just opened in September 1971.

2. Besides, the computer-communication systems operated by NTT, utilized 171 (200 bit/sec) and 327 (1,200 bit/sec) circuits, as of the end of March 1972. These circuits are not included in the above table. Also, 449 terminals (200 bit/sec) were connected with the telephone network for DEMOS and DRESS systems operated by NTT.

3. In addition to the present leased circuit services, the introduction of telegraph-type 200 and 1,200 bit/sec services and telephone-type 4,800 and 9,600 bit/sec services is under consideration.

ANNEX 2

A CALCULATION SERVICE FOR TELEPHONE SUBSCRIBERS -DIALS

In September 1970, NTT inaugurated DIALS (Dendenkosha Immediate Arithmetic and Library Calculation System) in Tokyo - a calculation service for telephone subscribers.

DIALS service is available daily from 8 a.m. to 9 p.m. for all telephone subscribers with a push-button telephone, which is equipped with ten numeral buttons and two additional function buttons, * and #.

Such a subscriber can have access to the DIALS centre by sending the seven-digit special code "0100111", and then, by pushing buttons according to a calculation formula followed by a command, "**#", he will receive the answer to the formula by voice-response, mechanically compiled from 90 word and/or phrase groupings. A subscriber wishing a hard-copy answer will be provided with a handy printer.

Characters and symbols necessary for calculation are composed of not more than three numeral and/or function button strokes. A subscriber wishing a single-stroke operation for inputting a character or a symbol, may be provided with a thirty-four-button handy keyboard. Another advantage of the keyboard will be that it may enable subscribers with ordinary rotary-dial telephone to utilize DIALS service by switching the telephone line to the keyboard after having set up a dialling connection to the DIALS centre.

DIALS users can probably make more complicated calculations than those possible by conventional desk-top calculators. In other words, a telephone subscriber may make use of approximately a hundred library programmes stored in the DIALS computer system. Also, he may temporarily store a calculation formula defined by himself and repeatedly substitute numerals for its variables during connection to the DIALS centre. In DIALS service, however, some limits have been set to the calculations, e.g., the order is not more than sixth in solving an algebraic equation, since the internal processing time is limited to within 200 milliseconds.

The charging rate for calculation is seven yen per twenty-one seconds duration of a connection, independent of the calculation category, and charges are recorded on the telephone subscriber's meter on a multi-metering basis.

The DIALS centre in Tokyo is now serving approximately seventy-thousand push-button subscribers and the average volume of usage is about twenty-thousand interactions per working day, the holding time per call averaging a hundred seconds.

DIALS service has also been in operation in Osaka, since March 1971, and will be introduced in Nagoya in December of this year and in other cities seriatim.

COMPUTERS AND A TELEPHONE COMMUNICATION SYSTEM OF THE FUTURE

CHAIRMAN:

Laurin G. Fischer
Systems Division, Computer Sciences Corporation
Falls Church, Virginia, U.S.A.

PROGRAM COORDINATOR:

Dr. Wayne B. Swift
Policyholder Service Corporation
Falls Church, Virginia, U.S.A.

PAPERS:

J. Crompton, Plessey Company, Ltd., Liverpool, England, *"Structure and Internal Communications of a Telephone Control System"*

D. C. Cosserat, Plessey Company, Ltd., Liverpool England, *"A Capability Oriented Multi-Processor System for Real-Time Applications"*

K. J. Hamer-Hodges, Plessey Company, Ltd., Liverpool, England, *"Fault Resistance and Recovery within System 250"*

Dr. C. S. Repton, Plessey Company, Ltd., Liverpool, England, *"Reliability Assurance for System 250: A Reliable Real-Time Control System"*

DISCUSSANTS:

Mark Davies, Bell-Northern Research, Ottawa, Ontario, Canada

Paul E. Muench, American Telephone and Telegraph Company, New York City, New York, U.S.A.

THEME: The problems of developing a dial telephone system with programmable digital control equipment are analyzed in the context of the U.K.: an approach taken at Plessey to solve these problems is outlined to the extent of illuminating clearly the nature of the choices available and a rationale for the choice of one particular approach.

STRUCTURE AND INTERNAL COMMUNICATIONS OF A TELEPHONE CONTROL SYSTEM

J. Crompton
Plessey Co. Ltd.
Liverpool, England

Summary

Current plans for the introduction of computer controlled telephone exchanges to Great Britain envisage the formation of a new telephone network which will interwork with the current network and ultimately replace it. The basic module of the new network is a Switching Unit, which is controlled by a Processing Utility. Switching Units are composed of a number of subsystems, and each subsystem is subject to standard definitions both for its interfaces and for the function it performs. The hardware/software ratio of each subsystem is at the discretion of the manufacturer, but subsystems can be regarded conceptually as having a hardware component and a software control component. The action of the subsystem control programs is coordinated by a further control program, and a great amount of interaction is necessary between these programs during the setting up of a telephone call. The software mechanisms necessary for internal message handling and process creation must be chosen with great care bearing in mind the various trade-offs possible, processor utilization, and the definitions of the subsystem standard interfaces.

Introduction

Development of the British Telephone Network is guided largely by the Advisory Group on System Definition (AGSD) - a body consisting of representatives from both the Administration (British Post Office) and from the various manufacturers of telephone equipment. Any future computer controlled telephone exchange which is to be used in the United Kingdom will be subject to constraints laid down by AGSD.

The concept currently proposed by AGSD is to form an "overlay" network of Stored Program Control (SPC) exchanges. By this is meant a system which could start off life in a very small way - possibly a single exchange - interworking with the existing telephone network, and which could then grow in discrete stages. This would form a new, small network of SPC exchanges, which interfaced with the old network at selected points. As the new overlay network grows, it will slowly replace parts of the old, until eventually the entire system will consist only of SPC exchanges.

System Structure

Switching Units

The basic module of this new network is known as a Switching Unit. Switching units are of several different types, and each type can have many different designs and constituent elements. Basically, the function of each switching unit is to provide facilities whereby various telephone circuits can be monitored and interconnected under the instructions of a centralized control. This control may be located with the switching unit, but equally may be remote and operate via a data link. The centralized control is known as a Processing Utility. The two most common types of switching unit are:-

1. Subscriber Switching Unit
The Subscriber Switching Unit interfaces directly with the telephone user, by means of wires from the subscriber's premises. Fig. 1 indicates schematically a subscriber switching unit, which, with its interface to the existing network, could provide the start of the new network.

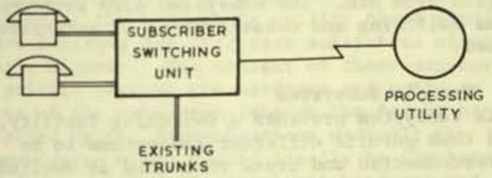


Fig. 1

TYPICAL INTERCONNECTION OF SWITCHING UNITS

2. Main Switching Unit
The Main Switching Unit is normally connected to a number of subscriber switching units, and also may be connected to other main switching units. This is indicated schematically in Fig. 2.

Although figures 1 and 2 show the switching units as connected by a single data link to a single processing utility, more complex arrangements will apply in practice, for security reasons.

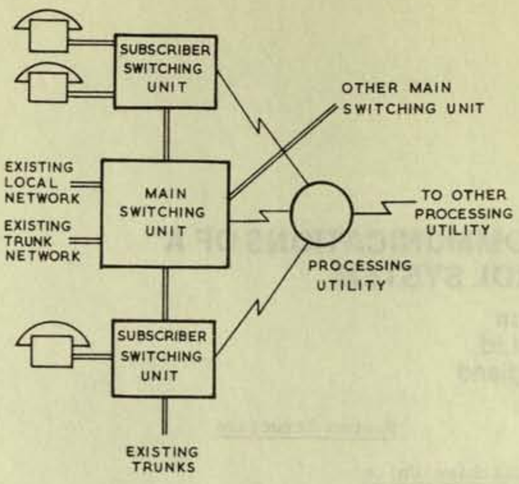


Fig. 2

MAIN SWITCHING UNIT

Subsystems

Each of these switching units is composed of a number of distinct elements, known as subsystems. Subsystem units are so chosen to provide interfaces which can be rigidly defined, and remain constant between equipment manufacturers, enabling equipments of various designs to interwork satisfactorily. Each subsystem performs a distinct function within its interface boundaries; it is the declared intention of AGSD to define these functions and interfaces. Some typical functions which can readily form subsystems, however, are:-

1. Subscribers Subsystem
This subsystem provides the complete interface between a particular group of subscribers and the rest of the network. It provides all system communication with the subscriber - for example it will provide dial tone and busy tone to the subscriber, and will accept dialled or keyed digits from him. The subsystem also performs some switching and concentration of subscribers lines.

2. Transit Subsystem
This subsystem provides a switching facility, and thus permits different subsystems to be interconnected and cross connected as desired.

3. Interface Subsystem
This subsystem is used to connect the new network with the old. It must provide all types of signalling and all facilities in use on the particular existing junctions with which it is connected.

4. Manual Board Subsystem
This subsystem provides the second "human" interface into the system (the first being the subscriber). The subsystem must provide all facilities necessary for operational staff to provide assistance to subscribers, monitor and test lines etc.

5. Miscellaneous Terminations Subsystem
This subsystem contains the various devices required by the administration - such as time announcement machines, message recorders, facilities for interconnecting multi-subscriber calls.

The configuration of Fig. 2 is redrawn in Fig. 3 to show some subsystems which could typically be involved, and the ways in which they could interface with each other.

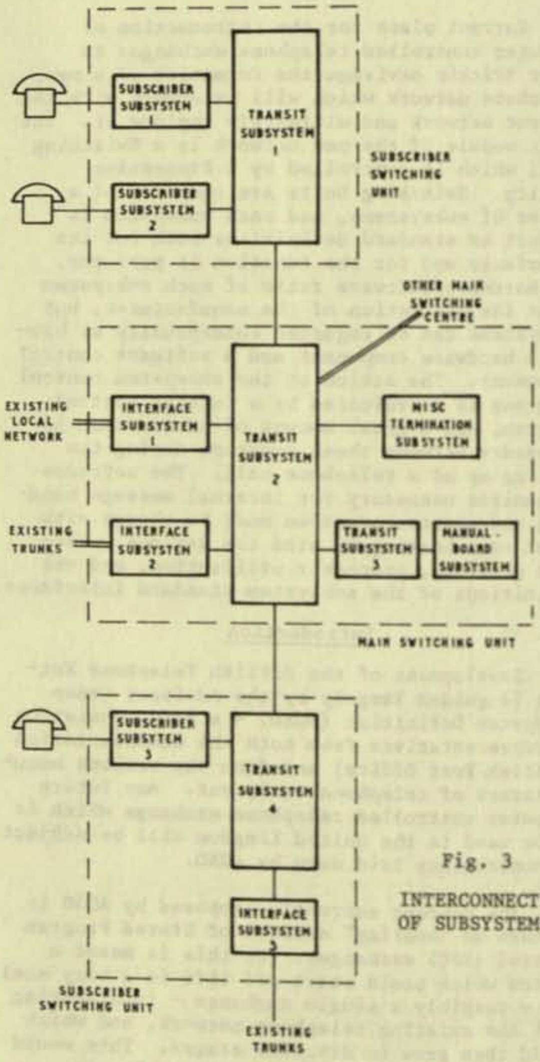


Fig. 3

INTERCONNECTION OF SUBSYSTEMS

Subsystem Structure

Subsystems are chosen to perform particular functions within defined interfaces; the manner in which the functions are performed will depend upon the particular method of implementation chosen by the particular manufacturer. This detailed implementation can vary greatly - not only between manufacturers, but within manufacturers as technologies advance. In particular, the amount of work performed by the hardware and the amount performed by the software can vary. For example, consider the hardware/software trade-offs which are possible in the design of the switchblock part of a subsystem:-

The basic requirement is to connect one particular input from a group of inputs, to a particular output. Fig. 4 shows a group of 12 inputs and 6 outputs, and a possible method of performing the connections by two stages of switching - each point marked X represents a switch or "crosspoint". It can be seen that by judicious operation of two crosspoints, any input can be connected by one of several paths to any output - provided that the paths are not already in use for another connection. Several methods of arranging this connection are possible: for example:-

1. Use of "intelligent" hardware, which would accept the identities of the two terminals to be connected, effect the connection if possible, and then return a "success" or "fail" message. This solution requires a minimum of software.

2. Use of simple hardware, which would merely activate or deactivate any nominated crosspoint, as instructed. This solution requires all the work to be done by the software - even to the extent of keeping a "map" of the crosspoints, in which busy ones are marked, and from which an available path can be selected, details of which are sent to the hardware.

3. Use of hardware falling between these extremes for example hardware which would activate and de-activate nominated crosspoints, check and report upon the success of the operation, and also provide facilities for the software to interrogate the state of nominated crosspoints. This solution leaves the "intelligence" with the software, but provides security for the current details of crosspoint settings.

4. Use of solution 2. or 3. above, but placing the necessary software in a local mini or micro processor, which acts upon instructions received from the processing utility.

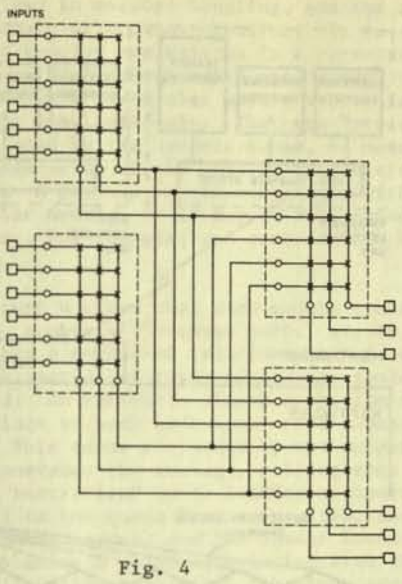


Fig. 4

TYPICAL 2-STAGE SWITCHING

Since it is possible for any subsystem to contain software, it is logical to consider each subsystem as consisting of two interdependent parts - the hardware, and the software within the processing utility. Any program structure, therefore, will conceptually contain a number of distinct subsystem control programs, but in the limit of complex hardware or local mini-processor implementation, the control programs will be simple message handlers.

These subsystem control programs must be able to transmit and receive messages to and from their hardware counterparts. The physical means of this message transmission may include a data link, and most probably will include methods of multiplexing and de-multiplexing along some message transmission medium - for example the normal I/O handling software of the virtual machine in the processing utility could interleave messages for different subsystems along a single highway. The content of these messages is private between the hardware and software parts of the subsystem (though the format may be affected by the communications medium); the transmission means should ideally be transparent. The interposition of additional hardware and software between the subsystem hardware and its control program in order to resolve message addressing and transmission problems in no way violates the concept of the subsystem with defined interfaces; it merely provides a transparent interconnection. Figure 5 shows schematically the type of arrangement that could exist for the system depicted in Figure 3. Each hardware subsystem has its software counterpart in the processing utility. The points marked X indicate interfaces which are likely to be defined as AGSD standard interfaces - these are interfaces at the software end of the subsystems; other interfaces subject to definition lie at the hardware end at the human interfaces.

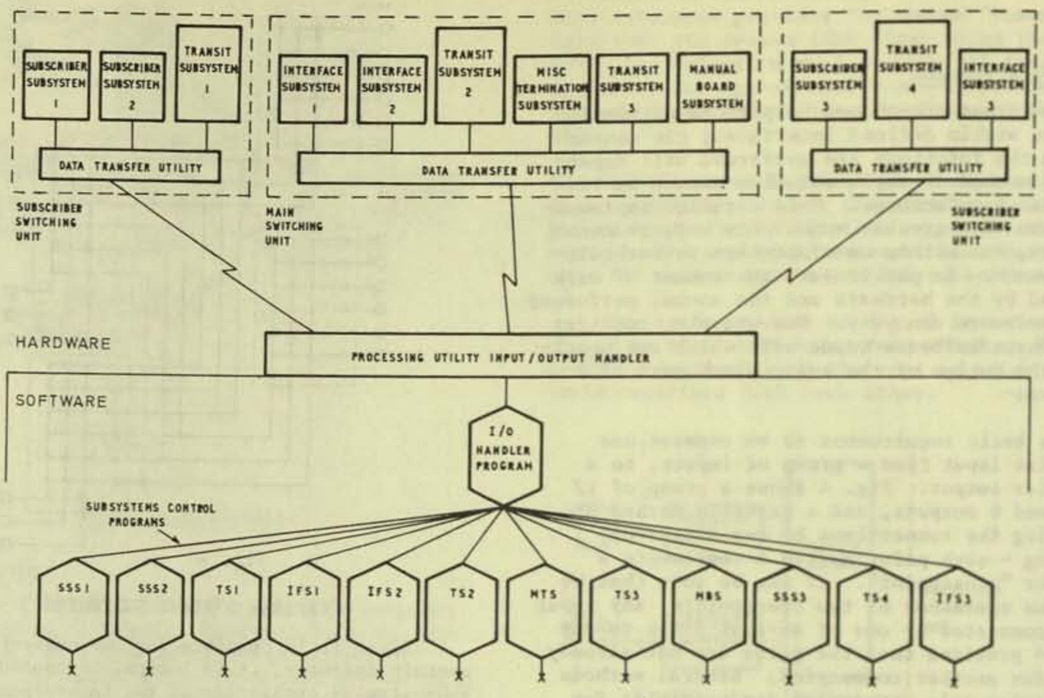


Fig. 5 TYPICAL INTERCONNECTION OF HARDWARE AND SOFTWARE SUBSYSTEMS

In practice the arrangement will probably differ from that shown in Figure 5. Frequently a single manufacturer will be responsible for a complete switching unit - if not for several co-located switching units. Two or more similar hardware subsystems could be controlled by a single control program, operating upon several data bases. Also, economics could dictate that certain pieces of equipment be shared by several subsystems. The hardware part of each subsystem consists of several devices, which are each treated as peripheral units. One of these equipments for example, which is often expensive, is called a Marker. The marker is the peripheral which controls the operation of crosspoints in the switchblock, as explained previously. A marker may well have the capacity to control more than one switchblock, so its costs could be shared among subsystems located together. Figure 6 indicates a possible method of re-structuring the software configuration of Figure 5. The handling programs for the individual peripherals (such as the marker) are shown, and these programs communicate with the subsystem control program proper, which must co-ordinate the operation of the peripherals in its subsystem hardware. Only one subscriber subsystem control program is shown - this will handle all three subsystems from three separate data bases; similarly for the other subsystems.

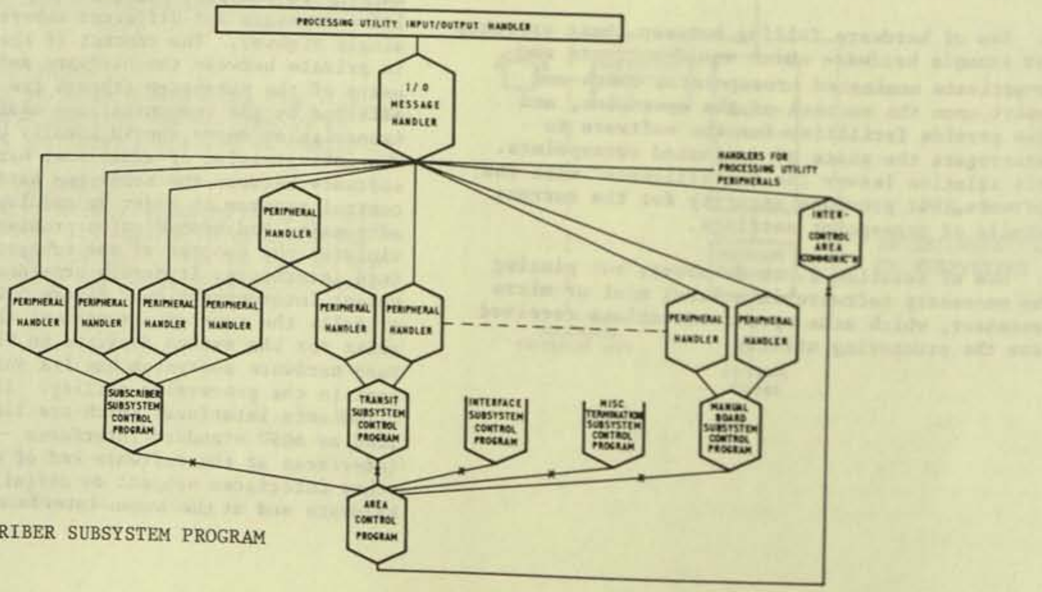


Fig. 6

TYPICAL SUBSCRIBER SUBSYSTEM PROGRAM

A software structure is now starting to emerge. A number of software "modules" have been identified, and some have software interfaces which are the subject of future definitions. The word module is used in the sense of a self-contained piece of software, which could be written and tested in isolation. As yet, no means of co-ordinating the operation of these modules has been mentioned; it is here that the concept of a Control Area is introduced. A Control Area consists of a group of switching units which are controlled by the same processing utility, and within which it is possible to select overall the path that will be used by a particular call, before that path is set up. In Figure 6, a software module called Area Control has been introduced, and this co-ordinates the operation of the individual subsystem control programs. The area control program can be regarded as holding the intelligence for the call, and the subsystems execute specific commands given by area control.

Internal Communication

Consider the type of interaction that will be necessary between the software modules of Figure 6 - for example when a subscriber in subscriber subsystem 1 wishes to make a call to a subscriber in subscriber subsystem 3 (Figure 3). The first indication that a call is to be made is given by the subscriber lifting off his handset; this event is detected by a peripheral called a Subscribers Line Circuit within the subscribers subsystem. The event will pass via the I/O message handler, peripheral handler and subsystem control program to the area control program which must examine its records to determine the type of service permitted to the particular subscriber. If dialling out is permitted, the subscriber subsystem will be instructed to connect the appropriate type of digit receiver, in anticipation of keyed or dialled digits, to connect any supervisory circuits that may be required, and to connect dial tone to the line. As digits are received, they pass via the chain to the area control program until eventually sufficient digits have been received to determine the destination of the call. After checking the availability and status of the called subscriber, instructions must be provided to the appropriate subsystems at the appropriate times to provide ringing current to the called subscriber, ring tone to the calling subscriber, to remove these conditions, to set up a path via the two transit subsystems, to check upon the continuity of the path - and eventually to clear down all connections. Should the destination lie in a different control area, messages must be sent either directly to the destination control area or to some intermediate (transit) control area which must itself activate appropriate subsystems.

It can be seen then, that within the software structure which has now emerged, there is a great requirement for the passing of messages between individual modules. Careful study must be given during system design to the mechanisms that will be involved in message handling, and the closely allied topic of process creation. It is assumed that all modules are written in a re-entrant manner, so that conceptually one or many processes using a particular module may be in existence simultaneously. The term "process" is here used in the dynamic sense, to mean the serial execution of the code in a module or program. A process may be associated with a particular module, or it may be associated with a message, for example, and cross module boundaries.

Figure 6 shows that each module contains a discrete number of "message ports" or, in other words, has a number of interfaces across each of which particular types of message could be expected. An extremely simple mechanism could be to place at each such input port a message queue. This queue is loaded by the output port, which generates the message, calling upon a common, centralized queue loading mechanism. The call to the queue loader specifies the name of the wanted queue, and the loader locates the queue by using a close association with the space allocation mechanism. (Absolute addresses could not be used, because in a system of this nature with a requirement of many years mean time between failures, it is necessary to move programs and data around, when system components either fail or are released for scheduled maintenance). Once one of these modules is scheduled, it runs until all its queues have been emptied, at which point it terminates. The time scheduling algorithm can be constructed to any arbitrary degree of complexity. This system has several attractions - the time and space overheads involved are quite small, and there is no danger of messages getting out of sequence and "jumping their queues". Also no contentions arise for file access; since only one process exists on any particular module at any instant access to that module's in-core data bases need not be subject to any lock and key control.

This system, however, could become quite inefficient as traffic grew. The number of processes is equal to the number of software modules - but it is quite probable that the processing utility consists of a number of processors working in parallel in a load-sharing manner. The precise number of processors is of no interest to the applications programs, provided that collectively the processors provide sufficient processing power; the supervision and co-ordination of the individual processors can be regarded as a function of the "virtual machine". The net result, however, is that several processes may be able to run concurrently. The system of one process per module forces all telephone calls to queue for sequential service by the area control program, whereas logically there is no reason why separate calls should not be processed in parallel by use of several processes on several processors - thus removing what could become a serious bottleneck on processor time utilization as traffic increased.

Such considerations lead to the proposal of creating a separate process for each telephone call. As this process completes execution of one module, it transfers control to the next module required by the particular call. Data associated with a particular call is carried in the "process base", or workspace associated with the process, and this reduces the overhead of message passing between processes. A number of difficulties are found with this approach, however. All messages entering the system require a certain amount of processing before thay can be associated with a particular call, and it is only when a message has travelled a certain distance that it can be picked up by its parent process - and this distance will depend upon the point at which the parent process last suspended. Parallel processing of certain activities associated with a call is not possible - for example, in the subscriber call described previously, the parallel actions by the two subscriber subsystems of setting up ring tone and ring current would need to be carried out sequentially (although of course each subsystem could be simultaneously active upon different calls). Even more serious difficulties are encountered when a call requires to be handled by a second (or third) control area - it is not feasible for the process to cross control area boundaries. This approach also entails a process crossing a subsystem "standard interface" which ideally should be defined in terms of messages only.

Yet another possibility is to use a process per module per telephone call. This approach requires a message passing mechanism which can deal with a high message rate, and which can associate messages wuth processes. The combination suggests a centralized system which uses semaphores for communication, and which is intimately associated with the time scheduler. The system must also allow information to be passed with each semaphore. Such a general mechanism is currently being implemented by one telephone manufacturer; it is conceptually simple, and flexible in application, permitting easy system expansion.

The mechanism readily provides association of messages with processes, and provides for reactivation of suspended processes. Great care must be exercised in its use, however, because of the space and time overheads inherent in such an approach. Even though it provides a useful mechanism for the process per module per telephone call. Some problems still remain with this approach - such as the file locking problem mentioned previously. For example, in a subsystem which contains a switchblock, and which maintains a store image of that switchblock, each call which uses the image to select a path will require unique access to the image for duration of the path choice algorithm; some method of constraint is necessary if a process per module per telephone call approach is used.

Obviously, none of these systems represents black or white; all are shades of grey. An attractive compromise is to treat different types of module in different ways. The I/O message handler is really a function of the processing utility, but can be considered here as being in two parts; input, and output. The input part is a single process, which is scheduled at regular intervals, and once scheduled runs cyclically until all incoming messages have been handled. As each message arrives, the addresses and other information from the Data Transfer Utility will identify the device from which the message originated, and the message is then passed to the appropriate peripheral handler (by the semaphore mechanism), where a process is activated for the particular message. The output part is also a single process, which is activated whenever a message is sent to it via a semaphore, from any peripheral handler program; having dealt with the message, the process suspends itself awaiting the next message.

The peripheral handler processes, in the case of input, will "funnel" down to a single process, which runs the subsystem control program. This funnelling can be achieved by private queues, in the case of unshared peripherals, or by use of the semaphore mechanism for peripheral handlers shared between subsystems. The single subsystem process can now service its messages in cyclic fashion, and has no contention problems for its files. In the case of subsystem control programs controlling several different subsystems, each on its own database, there is one process per database.

At control area level, yet another arrangement applies. Communication with the subsystem control programs is handled by the semaphore mechanism, thus maintaining a message interface. Each instruction given to the subsystem is accompanied by a "tag" which uniquely identifies the particular call (for example a call number) and this tag is later returned by the subsystem when reporting upon the action performed. Within the area control program, a process is initiated which handles the originating part of all calls. When the initial message comes from the subscribers subsystem indicating that the handset has been lifted, this Originating Call Process allocates the unique tag to the call and handles the early parts of call set up. This single process limits itself to handling a fixed number of calls - the number is dependent upon the structure of the processing utility; once the number is exceeded, a further parallel process is created to handle subsequent calls (thus ensuring equitable use of processing resources). Once sufficient is known about the destination of the call, a second process is created. This process may be in the same control area or a different control area, depending upon the destination of the call - creation of the second process, however, ensures that all calls can be treated in standard manner, whether they be inter or intra control area. The second process may be a Terminating Call Process, but the originating and terminating call processes can be separated by one or more Transit Call Processes, if the call needs to pass through several control areas. Each such process runs in a cyclic manner, and will itself create further processes as the traffic load increases. All these processes existing in the area control program communicate with the subsystems by the semaphore mechanism.

Conclusions

This approach is by no means the only possible, but it does illustrate the type of mechanisms which are necessary for the organizing of intercommunication between various parts residing inside an SPC computer system. Most of the work currently being performed in this area is subject to change, since the definition of the interfaces is not yet available. As detailed implementation of SPC progresses, however, manufacturers are solving problems in increasing detail, thus permitting interface and functional definitions to be arrived at which are both efficient and enduring.

Acknowledgement

The author wishes to thank his many colleagues upon whose work this paper is based and the Management and Directors of The Plessey Company for permission to make this information available.

A CAPABILITY ORIENTED MULTI-PROCESSOR SYSTEM FOR REAL-TIME APPLICATIONS

D.C. Cosserat
Plessey Co. Ltd.
Liverpool, England

Summary

The system under consideration is a multi-processor, multi-storage module configuration adapted to the processing and fault security requirements of such real-time applications as telephone switching, message switching, and radar systems control. Each processor accesses store independently and asynchronously and each region of store to which it has immediate access is bounded by an addressing structure known as a Capability. The capability has a dual purpose. It acts as a protection mechanism against hardware and software failure; and it defines a logical unit of contiguous storage space (a "segment") out of which all operating system and application data structures are built. A segment may contain either data or capabilities permitting a list structure of interconnected segments to be established.

Each processor executes instructions contained in data segments, achieving linkage from segment to segment, and in so doing manipulates the data structure appropriately. One particularly significant feature of the system is that it is possible for a running program to make copies of capabilities which it can store arbitrarily into the data structure. The operating system reduces essentially to a series of 'protected subroutines', each subroutine possessing just the capabilities required to gain appropriate access to the data structure. There exists therefore a graded approach to storage protection and a complete lack of the visual division into 'special' and 'normal' modes of machine hardware operation.

Introduction

Computer systems are characterised by their ability to provide 'general purpose' solutions to specific logical problems. In the telecommunications field, in particular, the computer may be used as a centralised control mechanism which replaces the logical functions often formerly provided by distributed hardware devices. Thus, for example, the centralised computer system can be used to control telephone switching hardware which itself contains little or no sequential logic either from a mechanical or electronic point of view. Similarly a computer may be used to provide automatic routing of messages in a message switching network; automatic information retrieval, computation, and display in an air traffic control environment; centralised control of industrial processes; network control of distributed systems such as electricity and gas grids; area control of road traffic schemes; etc.

Three important factors relevant to these systems are security, growth and obsolescence. Real-time systems whose operation affects a large number of human beings must be capable of withstanding long periods between system failures. In some cases this requirement arises from an economic or strategic need (in telephone switching systems, for example) and in others (such as air traffic control) human lives are directly involved. The second factor arises because telecommunications networks have traditionally been designed so that increases in size and 'traffic' carrying capacity can be accommodated over a period of years. Thirdly the nature of telecommunications networks and, in particular, the amount of capital investment required implies that systems installed today should not become rapidly obsolescent.

Traditionally, the kind of computers that have been applied to these real-time control tasks have emerged from two quite different stables; on the one hand, system designers have made attempts to adapt computing equipment developed in the data processing environment to the requirements of real time control, and on the other hand engineers who have experience of existing electromechanical and electronic techniques have tried their hand at producing computer systems. This dichotomy of discipline has led in the past to a polarisation of ideas on how real-time centralised control systems should be built. As a broad generalisation it might be said that the computer engineers have failed to design systems which have the security and expandability features so characteristic of telecommunications systems, whilst the telecommunications engineers have failed to design systems which promote to the full the control flexibility afforded by software technology.

In order to illustrate the problems confronting the computer system designer in this field it is useful to select a particular case for analysis. The case chosen here is that of the telephone switching control problem because it represents a particularly comprehensive example of conflicting requirements. Designers who are interested in other real time application areas will, however, recognise many analogies with their own problems.

In the case of telephone switching control, it was thought for a long period of time that the major problem involved was the tricky technological one of designing a centralised control device to obey the necessary logical steps to control the switches involved in setting up a circuit from one telephone subscriber to another. This was obviously the immediate and central task and it was tackled in a variety of different ways. Some solutions involved hardware-wired logic as the means of centralised control, others involved the use of a computer-like device which fetched instructions from a read-only store, and yet others utilised a true computer configuration in which a processing unit fetched instructions from and modified data in, a read/write store. At the beginning, it did not really seem to matter very much which particular system was chosen because the central problem of switch control was identical in all cases, and often the decision as to which system to adopt depended on the design experience and background of the individuals concerned.

On the basis of a computer's ability to set switches in a telephone network, it was also arguable that there were no obvious advantages in any of these approaches over the previous electromechanical systems. Certain peripheral advantages were said to accrue such as 'system flexibility', but what exactly did this mean, and how was such a nebulous term to be quantified? Nowadays, it is possible to enumerate a number of facilities provided by a computer controlled telephone switching system. For example,

automatic fault diagnosis
centralised maintenance
network monitoring
automatic accounting
integrated manual assistance facilities
special subscriber facilities

All these factors were, of course, recognised by the advocates of centralised control, but they were regarded as a bonus to the more immediate problem of the switching task itself. A closer examination of these and other similar facilities leads, however to a new concept: a centralised computer system for telephone switching control must be regarded as an administrative unit which interfaces primarily with the staff of the telephone administration. Except for the provision of special facilities, the interface to the telephone subscribers is of secondary significance since it necessarily remains very much the same as in all previous systems.

In a computer controlled telephone network all the above facilities can be provided in a centralised manner. Fault diagnosis and maintenance can be handled by a relatively small staff via interactive video-displays; network monitoring programs can be similarly controlled by a few staff at a centralised location. Automatic accounting software can remove the human data preparation link, passing metering information from individual calls into a form suitable for the direct printing of accounts. Manual board operators can communicate via similar video displays on which all information pertinent to the call is recorded. The operator has sufficient control to achieve the required objective without the necessity of any administrative overheads, such as the filling out of dockets: instead, the system records the call details and cost automatically and routes it directly through to the accounting software system. This approach implies a unification of system design and, where necessary, the derivation of simple and standard ergonomic interfaces with those staff who control it. The activities mentioned above are summarised in Figure 1.

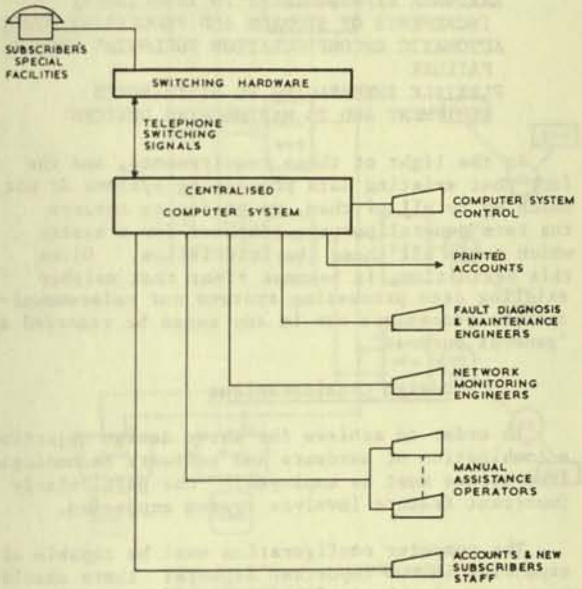


Fig. 1 ORGANISATION OF A CENTRAL PROCESSING SWITCHING UTILITY

Once the centralised control system is regarded as an aid to administration of the network, and once it is understood that it is here that the real economic advantages lie, the requirements of the centralised computing system necessary to support such activities become very different from those needed to handle the switching problem itself. In particular it is clear that wired logic or read-only program storage systems will not provide the necessary flexibility; and it is clear also that comprehensive software facilities are required to sensibly tackle the application requirements of what has become a real-time system with multi-access control.

The General Purpose Computer

It is clear from the foregoing telephone switching example that the problems of large and comprehensive central control systems are not amenable to solution either by dedicated telecommunications processors or by existing computers designed for the data processing environment.

Here is a list of some of the more obvious and important requirements:

- ABILITY TO RUN REAL-TIME PROGRAMS
- MULTI-PROGRAMMING FACILITIES
- MULTI-ACCESS FACILITIES FOR MAN/MACHINE CONTROL
- STRICT INFORMATION PROTECTION BETWEEN PROCESSES
- THE CONVERSE ABILITY TO SHARE INFORMATION BETWEEN PROCESSES WHERE REQUIRED
- HARDWARE EXPANDABILITY IN INDEPENDENT INCREMENTS OF STORAGE AND PROCESSING POWER
- AUTOMATIC RECONFIGURATION FOLLOWING SYSTEM FAILURE
- FLEXIBLE INTERFACING TO DISTRIBUTED EQUIPMENT AND TO MAN/MACHINE DEVICES

In the light of these requirements, and the fact that existing data processing systems do not match up to all of them, we prefer to reserve the term 'general purpose computer' for a system which meets all these characteristics. Given this definition, it becomes clear that neither existing data processing systems nor telecommunications processors can in any sense be regarded as 'general purpose'.

Design Considerations

In order to achieve the above design objectives, a combination of hardware and software technological innovations must be employed. One particularly important feature involves system expansion.

The computer configuration must be capable of expanding in two important aspects: there should be no practicable limits on the size of the fast store; and there should be as wide a range as possible of processing power. In each of these cases, the hardware should be expandable in reasonably small increments so as to permit a smooth rate of increase in capital investment in the system. It is particularly important that increases in storage and processing power can be achieved independently, since there is no obvious correlation between one and the other over a wide range of possible systems. Therefore a true multi-processor system which can contain a variable number of processing units and a variable number of storage modules is the ideal for the application.

The system must be resilient against both solid and transient hardware failures, and similarly against software bugs (which have many of the characteristics of transient hardware failures). This requirement means in practice that the system should be capable of automatic reconfiguration (i.e. switching out the failed hardware module) and recovery (i.e. the ability to return to the execution of a coherent program and data base).

The general purpose computer must also be capable of interfacing freely with a wide range of distributed telecommunications equipment, which may be remote from or local to the computer itself, and also must interface with man/machine devices such as video-displays and other computer peripheral devices.

Design Conflicts

The above remarks are addressed to some of the more obvious and important features of the general purpose computer. But some of the design requirements conflict and it is necessary to examine these conflicts in some detail.

The first design conflict arises from the requirement on the one hand to use the processors in a 'work sharing' mode to meet the requirements of a multi-programming, multi-processor system and on the other to respond quickly to interrupts generated by signals from the real-time system under control. Each processor must inherently be capable of obeying any program steps in the system (a functional approach involving the division of processors to specific tasks would conflict with the multi-processing requirement and with the need to expand the system with little software re-organisation).

The execution by a processor of a program is conventionally termed a process. In a multi-processor system there can clearly be as many processes in simultaneous execution as there are processors, but there may be an undefined number of additional processes which are blocked awaiting logical events or are freed but have no processor on which to run. When a process runs, the processor contains in its hardware registers information relevant to that process and when the process blocks, that information must be stored away. In a processor with several registers, the storing of their contents may involve many store accesses. An interrupt is caused by an event in the outside world which raises a signal into the computer system. This causes the processor to cease its present activity (i.e. to temporarily block the running process) and to execute an 'interrupt process' instead. The changeover from one process to the other involves the storing and loading of registers and hence there is a processing time 'overhead' on each interrupt.

In a single processor environment, this problem is often solved by the use of a second set of registers reserved for the interrupt process. In a 'work-sharing' multi-processor system this approach is not possible because the interrupted process is still logically free to run and may be picked up immediately by another processor. In this situation, the information concerning the process which is stored away in the first processor's register set is completely inaccessible to the second processor. Since the common medium of communication between processors is the store, it follows that the register information of an interrupted process must be written to store where it may be retrieved subsequently by another (or the same) processor. This register storing overhead is a theoretical limitation on a true multi-processor system and as such represents a design conflict between the attributes of such a system and the requirement to respond quickly to interrupts.

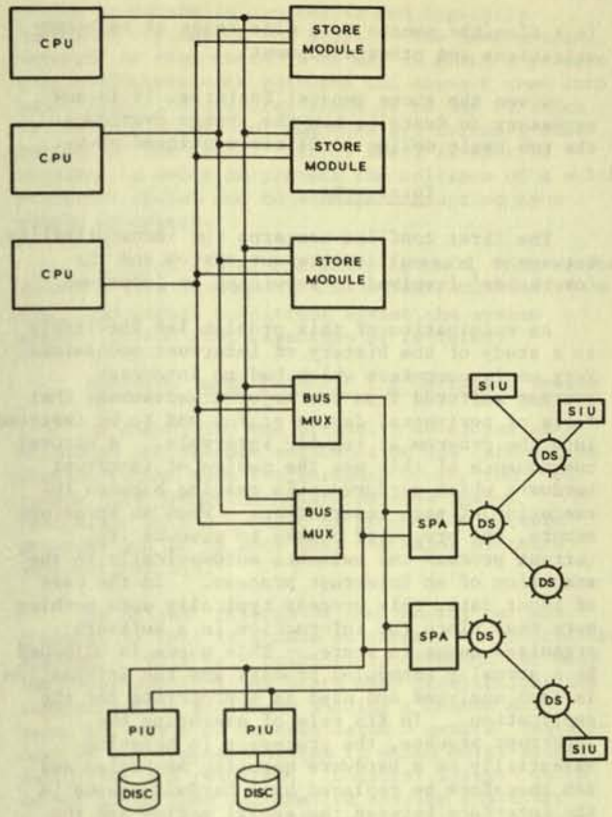
The second design conflict arises because of a potential fault security hazard in the universal sharing of store by all the processors in a multiprocessor system. In order to provide a system which is expandable in independent increments of storage and processing power it is necessary to abandon the usual concept of a computer as 'processing unit plus store'. The corollary of this divorce between processor and dedicated store is that a storage module belongs to no processor in particular and is equally accessible from all processors. Such a system organisation achieves the hardware modularity constraint at the expense of another; namely, resilience against store corruption which could lead to undetectable system failure. The concept of a multi-store, multi-processor system, which is so attractive from the point of view of modularity, is wide open to the possibility of storage corruption from any processor that fails transiently or permanently. Therefore there is a design conflict between 'equal availability of all storage locations' and 'potential damage to storage contents by a faulty processor'.

System 250 - A General Purpose Computer

An attempt has been made to embody the general design principles described above in the Plessey SYSTEM 250 central processing system. This system is designed for precisely the range of applications described and in particular for the control of administration and switching functions in a telephone switching environment. The design includes the following features which are compatible with and a consequence of the requirements of a general purpose computer.

Firstly, the hardware is designed as a multiprocessor, multi-storage module configuration as represented in Figure 2. Each processor may access any storage location in any store module over its own bus. Thus the modularity and incremental expansion requirements of the general purpose computer are satisfied.

Secondly, each processor is capable of detecting a range of hardware fault indications which will cause a fault interrupt to be automatically generated. The processor discontinues execution of the current process and switches to a fault interrupt process instead. The instructions obeyed by this process are, of course, fetched from store in the usual way but, should a subsequent fault interrupt be generated in that processor during the time that it is executing the fault interrupt process, the processor steps to the next storage module and recommences the process by fetching instructions from it. Thus, a failure in the storage module itself or corruption of its contents does not cause a permanent failure of the processor which received the original fault interrupt. The essential hardware mechanisms are therefore provided as a basis for an automatic reconfiguration software system.



CPU - CENTRAL PROCESSOR UNIT
MUX - MULTIPLEXER
SPA - SERIAL/PARALLEL ADAPTOR
DS - DATA SWITCH
PIU - PARALLEL INTERFACE UNIT
SIU - SERIAL INTERFACE UNIT

FIG. 2 TYPICAL HARDWARE CONFIGURATION

Thirdly, a flexible interfacing medium enables the system to be adapted to a wide range of peripheral equipment. The interconnection network is in the form of a bit-serial medium which transfers 'messages' between computer system and the periphery in both directions. A standard serial interface makes it possible to build a structure of 'primary' and 'secondary' electronic switches to suit a particular configuration and to interface simply to man/machine interactive devices. This satisfies the requirement that a general purpose computer should be connectable in a flexible manner to a wide range of telecommunications and other equipment.

Given the above general features, it is now necessary to describe how the system overcomes the two basic design conflicts mentioned above.

Interrupts

The first conflict concerns the incompatibility between a true multi-processor system and the 'overheads' involved in servicing an interrupt.

An examination of this problem led inevitably to a study of the history of interrupt mechanisms. Very early computers which had no interrupt systems suffered from the major disadvantage that tests of peripheral device status had to be inserted into the program at regular intervals. A natural consequence of this was the design of interrupt hardware which performs this testing between the execution of each instruction. When an interrupt occurs, the processor ceases to execute its current process and switches automatically to the execution of an interrupt process. In the case of input data, this process typically does nothing more than place the information in a software organised queue in store. This queue is unloaded by a normally scheduled process and the information is then analysed and used as appropriate for the application. In its role of executing the interrupt process, the processor is behaving essentially as a hardware queueing mechanism and can therefore be replaced by a hardware queue in the interface between the serial medium and the processing system. In SYSTEM 250 the unit known as the Serial/Parallel adaptor (see Figure 2) performs this function and, typically, can queue up to sixteen messages to and sixteen messages from the serial medium. A normally scheduled process unloads the messages from the hardware queue directly.

Another input/output requirement involves the transfer of data from magnetic backing storage devices such as drums and discs. Since it is usually uneconomic in conventional systems to withstand an interrupt for each word or character transferred, the standard approach is to use channelling hardware which moves data directly between 'burst mode' devices and store. During the transfer the processor is free to perform its usual functions and only receives an interrupt from the channelling hardware when the data transfer is complete. The usual characteristic of channelling hardware is that it is both elaborate and expensive; and it is, perhaps, unfortunate that there appears to be a tendency amongst computer designers in the direction of more elaboration and more expense. Some channellers are completely hard-wired, others obey special instructions fetched from store and begin to look very much like special purpose computers. SYSTEM 250 has taken this trend to its logical conclusion and utilises the standard processor module as a channeller. This approach has two very important advantages. Firstly, in a secure system there is no additional 'sparing' problem. Whereas it would have been necessary to provide a second channeller as a fault security backup, the additional processor now required for channelling work can share the existing spare processor(s) required to maintain processing security. Thus, in the majority of system configurations, the cost comparison is between one processing module and two channeling modules. Secondly, there is now no requirement for interrupt generation at the end of a data transfer because the processor itself can continue to process what would have been the interrupt response routine.

The two features of SYSTEM 250 described above, namely hardware message queueing and the use of the processors as channellers, has abolished the need for external interrupts and has therefore resolved the conflict between interrupts and the efficient operation of a true multiprocessor system. Additionally it has provided a cheap and conceptually elegant form of input/output control.

It should be observed, in conclusion, that there are still three mechanisms in a SYSTEM 250 processing module which can force a change of process: the first is due to a program trap condition; the second occurs when the processor's interval timer clock value reaches zero; and the third is due to the occurrence of an internally detected fault condition. Although these conditions may colloquially be referred to as 'interrupts', the common characteristic is that in no case is the condition externally imposed. The abolition of inter-processor and device-processor interrupt lines has a significant effect on the security of the hardware and makes it simpler to isolate processors and peripheral units following hardware failure.

It can be seen from the above discussion that the requirements of the general purpose computer are highly interactive. Both security requirements and the need for interrupt free operation of the multi-processor system affect the input/output economics in an unexpected way. By turning these conflicting constraints to advantage rather than by adopting some conventional compromise solution, it has proved possible to realise a simpler and more economically attractive solution to the problem.

Storage Protection - The Capability

The second design conflict which must be solved in the quest for the general purpose computer concerns the potential for storage corruption in a true multi-processor system. Two separate fears may be expressed on this subject. Firstly, there is the fear that processors which have access to the whole of the storage system may corrupt the program and read-only data held there. This will almost certainly result in a catastrophic failure of the system with instructions and data constants coverted to random values. This problem has led some designers to criticise the nature of an alterable store for critical real-time applications and to suggest that the older schemes of wired logic processors or of computers with their own dedicated storage modules are more adapted to the requirements.

As will be shown below, the nature of the problem is not so much the volatility of the storage medium itself as a lack of discipline on the part of the processors in their attempts to access it. It is this latter aspect to which attention has therefore been turned in an attempt to preserve the general purpose features of a freely alterable storage system.

The second fear is that, even in a system where read-only and read-write information is strictly segregated, there is still the possibility that faulty processors will obey random instruction sequences, attempt to obey read-only data as instructions, and alter read-write data values to which the currently obeyed program has no logical access. In short, even in a partitioned system of this type there is still much scope for corruption of store and therefore of system failure.

Solutions to this information protection problem typically involve the use of base-limit protection registers which partition the store into a number of contiguous regions or segments. Further protection measures may be applied to restrict access, such as the 'rings of protection' scheme suggested by Graham (reference 1). What is required, then, is a mechanism which permits the programmer precisely to define those data structures which will be made accessible to a running process and, by default, those which will not. There must be no system feature which prevents information sharing where this is logically required, and conversely, no system feature which permits information sharing where this is not logically necessary.

The solution which has been chosen in SYSTEM 250 involves the provision of hardware protection features which permit a given running process to access only those regions of store that the programmer originally intended. This is achieved by means of a universal segment identifier known as a capability. A capability is an invarient address which defines (a) the absolute location of a segment of storage, (b) the length of the segment, and (c) the kind of access permitted (read-only, execute only, read-write, etc.). What distinguishes a capability from a traditional base-limit protection address is that it can be freely copied by the running process itself (i.e. it can be loaded into a machine register and can be stored into a storage location), but that its contents can in no way be altered. The concept of a capability originated in the work of Dennis and Van Horn (reference 2), and was proposed in the present freely copiable form by Fabry (reference 3). The use of the capability mechanisms in SYSTEM 250 has already been described in detail elsewhere (reference 4) and no further elaboration will be attempted here.

The essential feature of a capability is its ability to permit the currently running process access to carefully controlled and logically necessary regions of the store. The hardware is arranged so that there is no way in which a process can manufacture data patterns and convert them into capabilities; therefore, there is no way in which it can gain access to, and possibly corrupt, other regions of the store. This, then, is what is required in order to prevent the collapse of a multi-processor system due to storage corruption by a single processor.

The corollary of the above is that, when faults do occur in a processor, the strict control of base, limit and access conditions assist the system greatly in the fast detection of failures.

Software Implications - The Operating System

The capability was primarily developed as a mechanism for storage segmentation and information sharing rather than for hardware protection. Of course, its protection features were always recognised in the context of protection between programs and it is here that the major software implications lie.

One of the criteria of the general purpose computer is that it should be capable of information sharing. This is a critical requirement for many real-time applications where many transactions are represented by processes sharing a common data base, but may also be considered a general requirement of any computer system in which multi-access facilities are required. Computer systems which do not allow good information sharing characteristics must resort to software control of shared storage and sometimes to the provision of separate copies of program for each process which requires to obey it. We may restate the requirements as follows: a multi-processor system should be able freely to execute code re-entrantly and should be able to access shared information when, and only when, this is a requirement of the program logic. The capability mechanism gives us exactly this property. Information sharing is permitted when required, and entirely denied when access is not logically necessary.

The protection afforded by the capability mechanism is extended in SYSTEM 250 to the interfaces between subroutine linkages. A program can only perform a subroutine call if it possesses the necessary capability for the subroutine. The access condition set into that capability permits 'enter' access only: that is, the capability can only be used to perform a subroutine call and not to gain access to the called subroutine's capabilities and hence to its data structure. Therefore the called subroutine's data structure is completely inaccessible to the calling routine. Similarly, once a routine has performed a subroutine call, the capabilities owned by that routine are denied to the subroutine and this satisfies the converse condition, that the calling routine's data structure is completely inaccessible to the called routine. Information interchange between two such routines is therefore strictly limited to that which the programmer intended: information may be passed as parameters in the form of data and/or capabilities in the machine registers; or information may be made permanently accessible to both calling and called routines, by placing in each routine's data structure a capability pointing to the shared information.

Given the inter-routine capability protection mechanism, it is now possible to construct all programs in a subroutine hierarchy irrespective of whether these programs are conventionally regarded as part of the application software or part of the Operating System software. This fact has had a dramatic effect on the design of the Operating System for SYSTEM 250 because it permits us no longer to regard it as a monolithic software package protected from application software corruption by means of a single impenetratable barrier. Rather, each logical function in the Operating System is treated as a distinct protected subroutine so that the storage protection philosophy within the Operating System structure relies on the same capability mechanisms as those utilised by the application programs. The result of such an organisation is that the system is not split into separate application and Operating System monoliths separated by a 'special supervisor mode' of hardware operation and the distinction between an Operating System and an application subroutine becomes one of administrative significance only.

List Structured Addressing

It has been stated previously that what distinguishes a capability from a conventional base-limit protection mechanism is the ability of the running process to perform load and store operations on capabilities by means of hardware instructions embedded in the program. This contrasts strongly with systems in which the reloading of base-limit registers is undertaken indirectly by software in 'supervisor mode'.

The free copiability property of capabilities enables the programmer to use them as invarient addresses in an arbitrary list structure and, indeed, an unlimited number of copies of a given capability can be generated. This distinguishes the capability mechanism from other invarient address schemes, such as the Burroughs descriptor (reference 5) which essentially restricts the data structure to a tree-like representation.

The arbitrary information sharing properties of the capability are exploited in the SYSTEM 250 Operating System to provide, in a simple manner, multi-programming and multi-access facilities. Firstly, the ability to arbitrarily share code segments means that all Operating System routines can be obeyed re-entrantly by many processes. Secondly, it is possible to strictly protect the information belonging to one multi-access user from that belonging to another. And, thirdly, it is possible for multi-access usersto share information in a controlled manner through a system of directories. The directory structure is similar in concept to that provided by the MULTICS Operating System (reference 6) but it differs in the following important respect: whereas, the directory structure in the MULTICS system is organised as a tree, the directory structure in SYSTEM 250 can be organised as any arbitrary list. Thus, the inter-connection of directories exactly mirrors the hardware level at which the capability mechanism permits an arbitrary interconnection of segments. This feature can be exploited to give precise information sharing properties to a system comprising groups of users of various classes. Our telephone switching example illustrated some of the many man/machine interaction requirements involving the sharing of some information. However, many of the classes of user are performing quite specific and separate tasks which do not require a great deal of administrative interaction. This is reflected in the organisation of directories to which these users are given access: it is the responsibility of the administration to organise the directories into a suitable list structure.

Conclusions

In conclusion, therefore, it has been shown that the requirements of computer systems in telecommunications applications are far removed from the facilities conventionally provided by either telecommunications processors or data processing machines. The facilities of a 'general purpose computer' suitable for these applications have been derived, the main features being incremental expandability of storage and of processing power, automatic reconfiguration of the system following hardware or software failures, and the simple interconnection to distributed telecommunications equipment and to man/machine interface devices.

It has been argued that to satisfy the above features, a computer system should be organised as a multi-processor with each processor equally capable of sharing the work available. This requirement in turn leads to two design conflicts which have been resolved in the design of the SYSTEM 250 computer system by, firstly, the abolition of external interrupts and, secondly, the use of a universal segment identifier known as a capability.

It has further been illustrated that the design solutions to these two conflicts have been turned to our own advantage because the problems involved have forced us to think out from first principles the necessary and sufficient features of a true 'general purpose' computer system. In particular we have been able to avoid an expensive and self-defeating approach to the production of channelling hardware, by recognising that the trend in this area towards increasing complexity implies a trend towards the use of standard processing equipment; we have been able to capitalise on the protection features of the capability mechanism by the design of a modular Operating System organised as a series of protected subroutines; and we have used the concept of free copiability of capabilities to reflect into the user terminal level of the system the idea of an arbitrarily interconnected structure of directories.

In particular, the capability mechanism, which is such a central feature of the SYSTEM 250 hardware architecture, enables us to claim three quite distinct achievements: the protection of information in a multi-processor system against hardware failure, the modularisation of Operating System and application software into a protected subroutine hierarchy, and the efficient and arbitrarily constrained sharing of data structures between competing processes. This leads us to believe that this concept represents a significant and essential advance in both hardware and software technology and that SYSTEM 250 provides both the sufficient and the necessary features of a 'general purpose computer'.

References

1. Graham, R.M. "Protection in an Information Utility", Comm. ACM, 11, 5 (May 1968) pp. 365-369.
2. Dennis, J.B, and E.C. Van Horn "Programming Semantics for Multiprogrammed Computations", Comm. ACM, 9, 3 (March 1966), pp. 143-155.
3. Fabry, R.S. "List Structured Addressing". PhD. dissertation - University of Chicago, Illinois, (June 1970).
4. England, D.M. "Operating System of System 250". Proceedings of the International Switching Conference, Boston, Mas. (June 1972).
5. Burroughs Corporation "The Descriptor" - A Definition of the B5000 Information Processing System - Detroit, 1961.
6. Bensoussan, A, Clingen, C.T, Daley, R.C. "The Multics Virtual Memory". Proc. Second ACM Symp. on Operating Systems Principles, Princeton, N.J. (Oct. 1969).

Acknowledgement

I would like to acknowledge the contribution of my colleagues who were involved in the design of the System 250 and, in particular, the valuable contribution of M. O'Halloran who understood before most of us what Operating Systems were all about. I would also like to thank the Directors of the Plessey Company for permission to publish this paper.

FAULT RESISTANCE AND RECOVERY WITHIN SYSTEM 250

K. J. Hamer-Hodges
Plessey Co. Ltd.
Liverpool, England

Summary

This paper describes some of the aspects of the Plessey SYSTEM 250 real-time processing system, and is an accompanying paper to those presented by my colleagues from Plessey U.K.

The requirements of a Real Time processor system suitable for the control of a communications application are evaluated. The ability of SYSTEM 250 to fulfil these requirements and the hardware architecture which provides the characteristics so urgently required by the communications industry is described.

A general description of the hardware of the processor is included and the use made of capabilities in ensuring the detection and isolation of fault occurences within the working system is described. Particular attention is drawn to the fault recovery sequence and the diagnostic facilities which enable the working system to live through fault conditions and offer the grade of service required by the application.

General Introduction

SYSTEM 250 was designed at the outset to meet the exacting control requirement of telephone or data message switching systems. It should be appreciated that this application demands an exceedingly high standard of performance in almost all of the areas considered important in Real Time applications. Convential computer systems are inadequate when examined against the essential requirements already established by conventional switch equipments. The characteristics of a computer system which will satisfy the stringent requirements of exchange control are summarised under the following headings:

Continuity of service
Ease of expansion
Ability to Evolve
System Partitioning and Security
Flexibility
High Power/Cost Ratio

Continuity of Service

The British Post Office has devised a sliding scale defining the allowed minimum reliability of telephone exchange control equipment. The scale ranges over steps from failures of the control equipment of less than 15 seconds which can be tolerated up to 50 times per year, to failures of more than ten minutes which should not occur more than once in 50 years. These reliability figures must be maintained despite:

(a) The existence of undetected software errors within the system.

(b) Occasional on-line expansion or modification of both the hardware and the software components.

(c) The need for long periods of unattended operation.

Ease of Expansion

A further requirement is that each individual system should be economically viable from the date of first installation. They must offer a growth potential such that the system is capable of ON-LINE expansion of any facility (e.g. Storage, Processing Power, or Input-Output Capability) by a factor of three during the expected life of 25 years. These extensions should not require alterations or re-compilation of the existing programs or cause any loss of service.

Ability to Evolve

A computer system which is expected to be operational for more than two decades can only remain economic if its architecture permits the inclusion of advances in hardware technology. The software architecture must also provide the flexibility necessary to absorb the undoubted changes which will be required to provide the, as yet, unforeseen facilities to be offered in the future.

System Partitioning and Security

The system hardware and software must be partititioned in a secure manner such that information transfers can be monitored, and faults or errors detected quickly and contained. The aim is to prevent corruption of and/or unauthorised access to system resources, in particular storage media, with minimal overheads in power, cost and complexity.

Flexibility

The control system is required to be flexible in both the hardware and software architecture such that a wide range of configurations with differing requirements can be controlled by differing configurations which minimise the cost of each system. In particular the system must be capable of efficiently controlling large numbers of low activity peripheral devices.

Introduction to System 250

SYSTEM 250 is a modular multi-processor system. The central system modules are Stores, Processors, and Multiplexors. Standard and non-standard Peripheral devices of all types can be attached as will be described subsequently. Twenty four bit word lengths are used for all memory addressing, instruction formats, and data storage. Thus the total memory capacity is in theory in excess of 16 million words. The instruction repertoire has been simplified to twenty seven basic operations, with inter-register, store and register or literal options available when meaningful.

Peripherals Devices are addressed via Control and Data registers which appear to the Processor to be exactly similar to the normal random access storage connected to the Processors, and it has, therefore, been possible to eliminate all specific peripheral handling instructions. Instead the normal Load Register and Store Register instructions are used, with addresses which specify the appropriate register within the desired peripheral device.

System Architecture

Interconnection of the Processors to attendant storage and peripherals is achieved over a 60 signal bus system, each Processor having an individual bus. Interface Units are used to attach stores and peripherals to these buses. See Fig. 1.

The function of the Interface Unit is to recognise requests for access to the module , resolve contention between individual requests from separate processors, and to allocate each request a cycle of access to the module. A system with up to eight processors is currently possible, with each of the Store Modules equipped with an 8 port Interface Unit. Peripheral Devices, however, are equipped with only 2 port Interface Units. When, therefore, there are 3 or more processors, peripherals connect to a Peripheral Bus system, driven by Multiplexors which can be equipped with 8 ports. Thus the more expensive 8 port Interface Units are not required throughout the Peripheral area. Two Multiplexors are required for security, and if either one should fail all traffic is passed through the alternative unit. See Fig. 2.

Up to 40 Modules can be attached to each Bus, over distances of 100 metres.

Only high activity, or fast speed devices, need be connected directly to the Bus system, e.g. Backing store devices. Low activity or slow speed devices such as user terminals or the application terminals of a real time system are connected to a serial data collection and distribution system known as the Serial Medium. See Fig. 3.

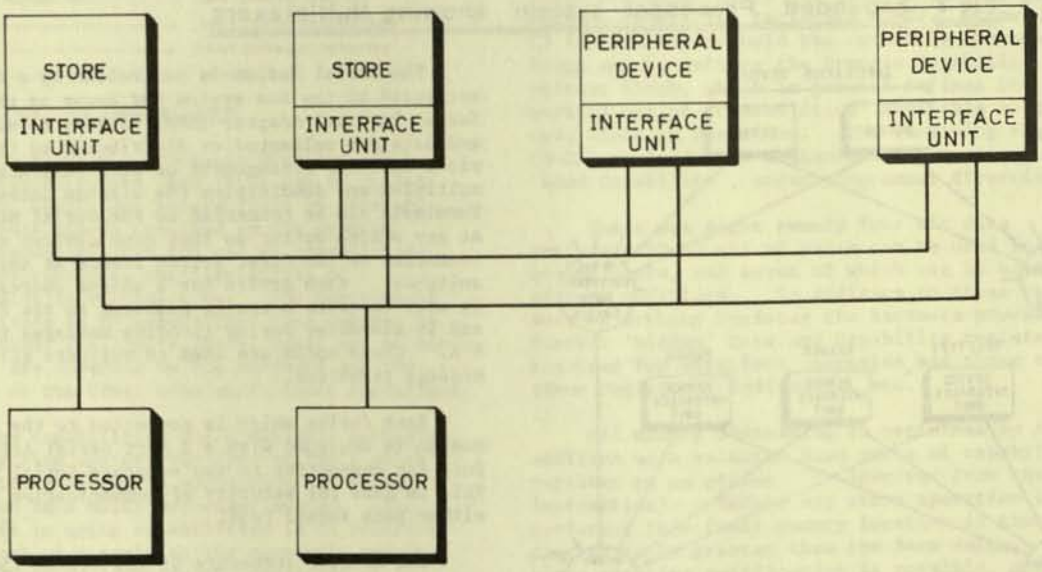


Fig 1 Typical two Processor system

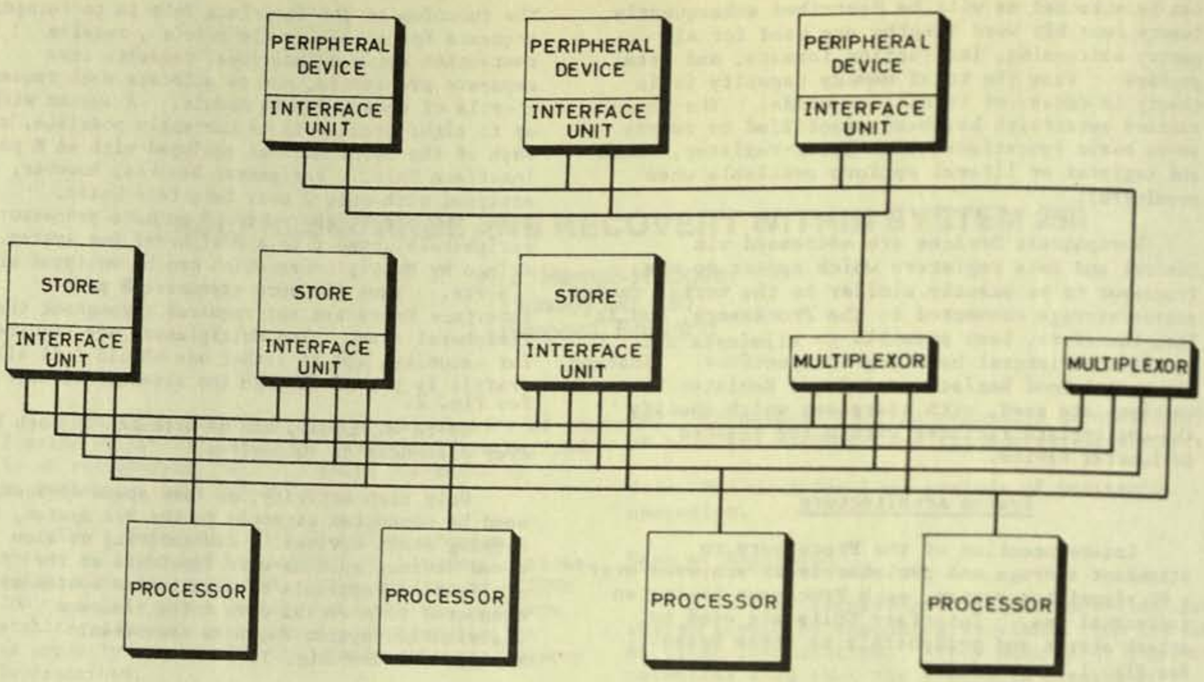


Fig 2 Expanded Processor system showing Multiplexors

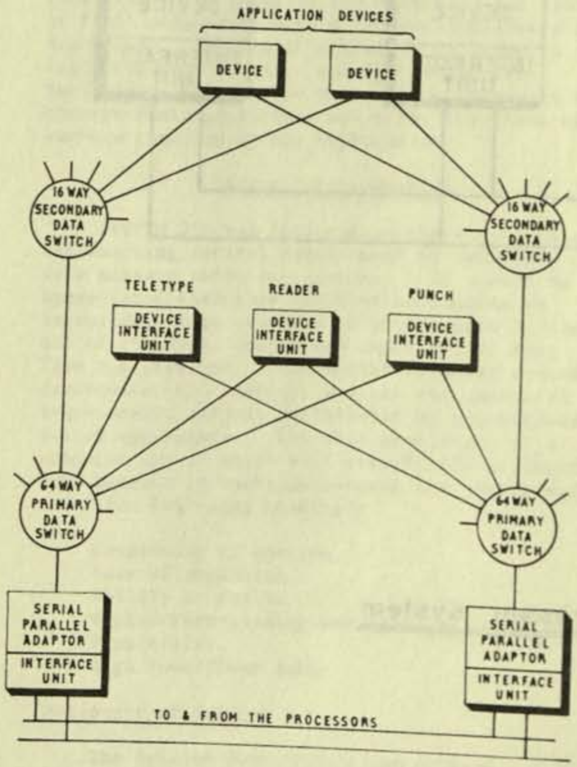


Fig 3 The Serial Medium

The Serial Medium is controlled by a device connected to the Bus system and known as the Serial Parallel Adapter (SPA). Packets of address and data are collected or distributed by the SPA, via a cascaded arrangement of Data Switches which multiplex and demultiplex the message paths. Terminals can be connected to the Serial Medium at any switch outlet so that some devices may be connected at the first switch others at subsequent switches. Each device has a unique address which is used to route outgoing messages to the device, and is assembled during incoming messages to the SPA. Check codes are used to validate all message transfers.

Each device which is connected to the Serial Medium is equipped with a 2 port Serial Access Unit for connection to two separate Serial Mediums. This is done for security of communication if either path should fail.

The modular structure of the SYSTEM 250 has been arranged so that individual system parameters can be matched in the most economic way possible, Stores for example can be built up in units of 8, 16 or 32K in slow, medium or fast access times ranging from 1μs to 300ns to match the data storage requirements. Numbers of processors can likewise be matched to the work requirements and the security requirements. The number of peripheral terminals can similarly be equated to the requirements of each installation. Further the System can be expanded in small steps by the addition only of the required module.

Capabilities

Each Processor has access to all modules connected to the system. Consequently each Processor represents a security hazard if either a hardware fault or a software error should corrupt a location by accident. The concept of Capabilities has therefore been implemented in the Processors to protect against corruption of invalid areas of storage, including Peripheral devices. Reference 1 discusses the necessity of capabilities and provides more detailed references. Capabilities are descriptors which identify the separate 'logical' entities within the system and the users access rights to the logical block. The Operating System loads these logical blocks into physical address space and allocates the Base and Limit address values accordingly via a map known as the System Capability Table.

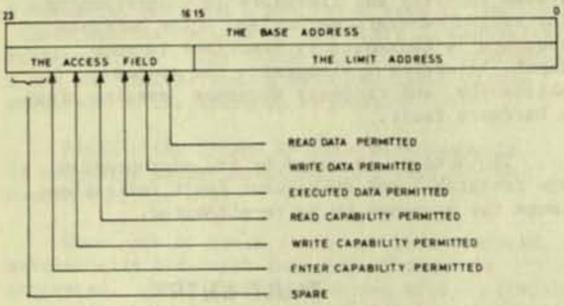


Fig 4 The Capability format

The Hardware of the Processor provides eight Capability Registers into which can be loaded the Base, Limit, and Access field of separate addressable blocks required by the program. It is emphasised that the use of capabilities in no way restricts the flexibility of programming at writing time. The function of capabilities is to ensure that once defined by the program, the limits (Base, Limit and Access Rights) are observed by the hardware and the process at run time, even under fault conditions.

Thus capabilities are a valuable mechanism in protecting against the type of fault which causes the progressive corruption and final breakdown in a multi-processor system. The basic aim in using capabilities is to restrict the effect of a fault to the currently running process, and to identify the existence of a fault immediately it occurs.

Processor Architecture

In order to understand the principles of system operation it is necessary to describe the architecture of the Processor . Reference 2 describes many of the hardware aspects of the processor which are not described in detail here.

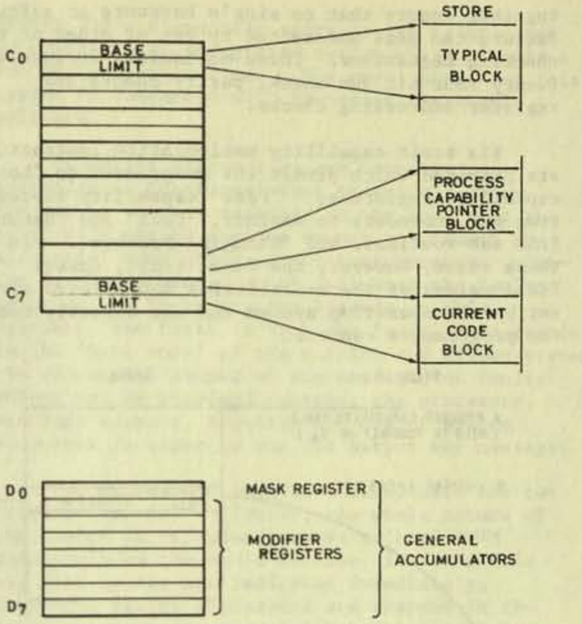


Fig 5 The working Register set

There are eight fifty bit (48 + 2 parity) general purpose capability registers C0-C7. Conventionall C7 is required to hold the currently executed code block and C6 defines the Process Capability pointer block, which in general defines the working set of 'capabilities' available to the code block in execution. The remaining registers C0-C5 are loaded by a standard instruction, 'Load Capability', under programmer direction.

There are eight twenty four bit data registers D0-D7 all of which can be used as accumulators, and seven of which can be used as address modifiers. In addition to these two sets of working register the hardware provides further 'hidden' Data and Capability registers required for efficient operation, and these cover timer registers, indicators, etc.

All memory addressing is performed by the addition of a selected Base value of capability register to an offset (derived from the instruction). Before any store operation is performed this final memory location is then checked to be greater than the Base value, since negative modification is possible, and less than the Limit value. Similarly, the micro-program action, Read or Write, is checked to be permitted by the Access field of the selected capability before the Store operation is allowed to complete. Fault interrupts are generated if any violation of the capability is attempted.

Clearly the system places great reliance upon the validity of the capability registers and the data held by them. Therefore a considerable number of checks are involved when loading and using capability registers which together ensure that no single hardware or software failure can pass undetected by one or other of the checking mechanisms. These mechanisms include a twenty four bit sum check, parity checks and register addressing checks.

Six basic capability manipulation instructions are provided which permit the programmer to 'Load' capability registers, 'Pass' capability blocks from one procedure to another, 'Call' and 'Return' from sub-routines, and 'Changing Process'. In all these cases, however, the Base, Limit, Adcess Field values of the capabilities manipulated are set by the operating system and not directly under the programmers control.

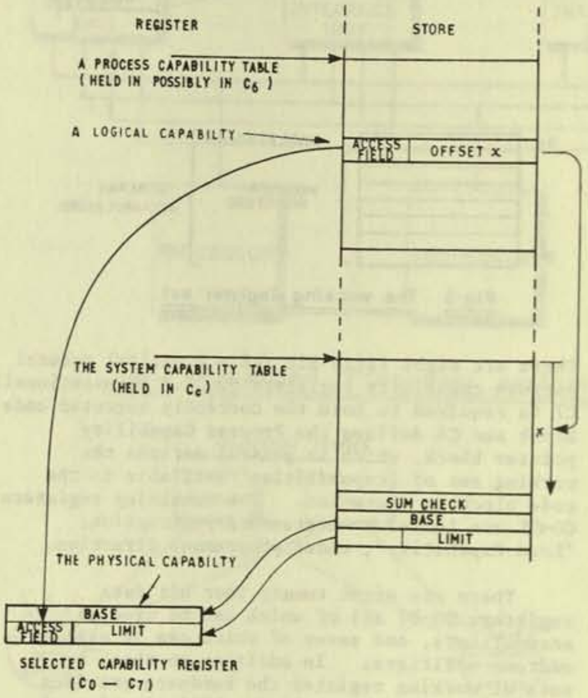


Fig.6 The Load capability sequence

The selection of the Base, Limit and Access Field is arranged via the System Capability Table held in one of the 'hidden' capability registers. Within this 'map' is described all the currently available blocks referenced in main store. Each user of the system, has a set of 'capabilities'. Each capability specifies an Access Field and the offset of one of the 3 word packets held in the 'map'. The Capability manipulation instructions reference the available 'capabilities', this in turn enables the hardware to select and load the assigned Base, Limit and Access Fields into the Capability registers of the Processor. Thus the logical capability is converted into a physical address at run time.

It must be stressed that although the programmer is at liberty to load into a register the assigned Base and Limit values of any of his available capabilities he cannot alter either the values of his own set of logical capabilities, or the corresponding physical Base and Limit values. This is effected by disallowing the WRITE DATA facility on a cap ability block.

Therefore at all times, the range of Memory locations which can be accessed is limited to the available 'capabilities' and the corresponding Base Limit values held in the System Capability Table.

Fault Detection and Recovery

In order to protect the working system from progressive collapse due to the migration of faults through the system, the Processor performs a Fault Interrupt immediately the fault condition is recognised, and before any actual capability violation can occur. The Fault Interrupt sequence is critical in order to preserve the system security and therefore in understanding the system recovery mechanism. The hardware sequence is consequently described in some depth, Reference 3 elaborates the system philosophy, and recovery sequence ensuing after a hardware fault.

The actions executed by the microprogram are repeatable and subsequent fault indications cause the sequence to be re-attempted.

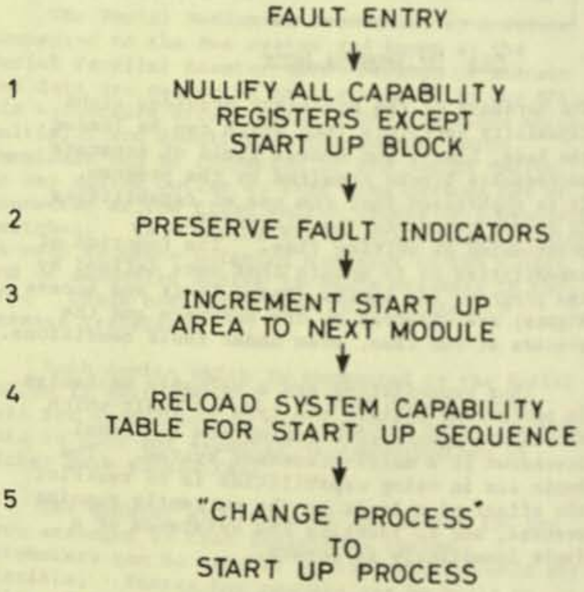


Fig. 7 The Fault Interrupt Sequence

Firstly the currently loaded Base and Limit values are corrupted to give invalid parity detection. This ensures that even given incorrect sequencing through the Fault Interrupt microprogram any attempts to access the memory locations of the previously running Process are prevented.

Secondly the Fault Indicator register is stored into a hidden register to preserve the fault indication. The Fault Indicator register is then cleared.

Thirdly, the Start Up Capability Register C(S) is incremented by 2^{16} so that it now references a different memory module. This ensures that during multiple fault conditions the Processor attempts to 'Start Up' from each of the available store modules in turn until it succeeds.

Fourthly, the Capability Register referencing the System Capability Table C(C) is reloaded with a Sumcheck, Base and Limit value held as the first three entries defined by the Start Up Capability register C(S). The block thus loaded references a new and limited set of Base and Limit values available to the Start Up Process.

Finally the Change Process microsequence is attempted using the Capability held as the fourth entry in C(S).

When each of these steps has been executed successfully the Fault Interrupt Process is activated. This Process will run with a limited set of memory locations available thus preventing interference with other Fault Free Processors.

The pre-requisite of the Fault Interrupt sequence is that at least one valid copy of the Start Up Block and the associated Program and Data block exists in any one of the equipped Store modules. Similarly if the Processor has a hardware failure which prevents the successful activation or completion of the Start up Process the hardware is condemned to an eternal cycling of the Fault sequence in an endless attempt to recover.

Note that the system recovery sequence which follows a fault detection can be made as rigorous as the application requires, Reference 3 discusses this in more detail.

On-Line Diagnostic Facilities

In order to achieve high reliability at reasonable cost the Mean Time to Repair faulty modules must be reduced to a minimum. In broad terms this has two effects. Firstly, the possibility of a second failure within the critical part of the system during the 'down time' of the first module is minimised, thus improving the system reliability, or alternatively, for a fixed reliability the number of redundant modules of any one type is minimised thus reducing system cost.

SYSTEM 250's diagnostic software and maintenance procedure is an integrated system which minimises the system repair time. The novel aspect of this system is concerned with Processor diagnostic software.

Processor diagnostics are normally an extension of functional test programs. They are run on suspect machines in the hope that the fault will not be serious enough to prevent the successful completion of the test program. Output is then produced which indicates the faulty component. There are two hazards in this approach, the first is that the fault could reside in the 'hard core' of the machine and either prevent the successful output of any message, or faulty output may be obtained, second, the processor, although suspect, requires the use of system resources in order to run and output any message.

For System 250 this is unacceptable for two crucial reasons. Firstly, the whole nature of the design is oriented towards a 'hard core' which includes the whole machine, it is in this way that faults are indicated immediately. Secondly, faulty processors are trapped in the fault recovery sequence deliberately so that they cannot make use of systems resources.

However, as a consequence of System 250's multiprocessor philosophy, it has been arranged that the diagnostic routines run on a working processor which then interrogates the suspect machine.

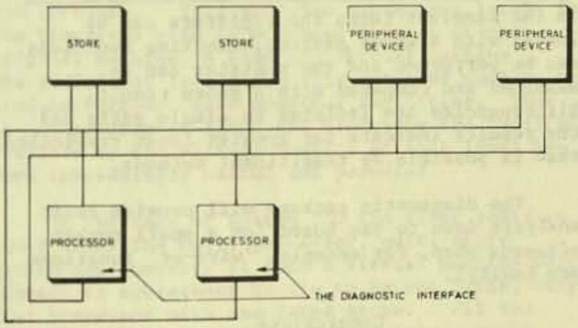


Fig 8 The connections of the Processor Diagnostic Interface

Each Processor module has an optional 'Diagnostic Interface'. This interface is exactly the same in operation as the Store and Peripheral Interfaces connected to the Store Bus. Each Processor can therefore be connected to the Test Interface of one of the other machines in the System, either directly or via a Multiplexor. The internal logic of each processor is therefore addressed as memory locations. The appropriate 'Capabilities' must be loaded into the hardware registers of the interrogating processor in order to address the suspect machine.

A set of Commands are provided as part of the Diagnostic Interface which facilitate the operation of certain essential functions.

STOP MAIN PROCESSOR CLOCK

START MAIN PROCESSOR CLOCK

PERFORM SINGLE SLOT WORKING

PERFORM SINGLE INSTRUCTION WORKING

REPEAT A PARTICULAR INSTRUCTION

STOP AT A PARTICULAR INSTRUCTION ADDRESS

STOP AFTER "n" SLOTS

STOP AT A PARTICULAR MICROPROGRAM SLOT

STOP AT A FAULT CONDITION

MONITOR MICROPROGRAM CONTROL SIGNALS

FORCE MICROPROGRAM CONTROL SIGNALS

MONITOR INTERNAL REGISTERS

FORCE INTERNAL REGISTERS

Fig.9 The Diagnostic Interface Commands

In the simplest terms the registers can be loaded with a known pattern, clocking functions can be performed and the register can be examined and compared with a known result. Discrepancies are isolated to single paths and the results indicate far greater fault resolution than is possible by traditional methods.

The diagnostic package will provide fault analysis down to one board (or a small number of boards when, for example, 'wire-or' functions are faulty).

Conclusions

Each characteristic of SYSTEM 250 was conceived to satisfy one or more of the design requirements detailed at the start of this paper.

A SYSTEM 250 CHARACTERISTIC	THE REQUIREMENT
CAPABILITY PROTECTION MODULARITY REDUNDANCY MULTIPROCESSOR TRAFFIC SHARING FAULT DETECTION AND RECOVERY ON-LINE DIAGNOSTICS	CONTINUITY OF SERVICE
STANDARD HARDWARE INTERFACES STANDARD SOFTWARE INTERFACES MODULARITY MULTIPROCESSOR DATA COLLECTION AND DISTRIBUTION	EASE OF EXPANSION,ABILITY TO EVOLVE, FLEXABILITY & POWER/COST RATIO
CAPABILITY PROTECTION STANDARD HARDWARE INTERFACES STANDARD SOFTWARE INTERFACES	SYSTEM PARTITIONING & SECURITY

Fig.10 The System Characteristics

While not exhaustive, it is hoped that this paper, in conjunction with the others presented by my colleagues, has indicated the principles of operation of SYSTEM 250, its architecture, and its power.

References

1. D.C. Cosserat - 'A Capability Oriented Multi-Processor System for Real-Time Application' presented at this Conference.

2. D. Halton - 'Hardware of SYSTEM 250 for Communication Control' Proceedings of the International Switching Conference, Boston, Mass. June 1972.

3. C.S. Repton - 'Reliability Assurance for SYSTEM 250 a reliable, Real-Time Control System' presented at this Conference.

Acknowledgement

I would like to thank the many colleagues on whose work this paper is based and the Directors of the Plessey Company for permission to publish it.

RELIABILITY ASSURANCE FOR SYSTEM 250
A RELIABLE, REAL-TIME CONTROL SYSTEM

C. S. Repton
Plessey Co. Ltd.
Liverpool, England

Summary

System 250 is a multi-processor system designed for real-time communication applications where very reliable operation is required. The initial application of this system (control of a telephone exchange) is required to achieve a mean time between failure of 50 years, where a failure is defined as a system outage lasting over ten minutes.

The paper describes in a general way the problems involved in providing this degree of reliability, and some solutions which can be adopted. The approach which is being used in the design of System 250 is described.

Particular emphasis is placed on the initial stages of recovery which ensure that a fault-free system configuration is set up and that a basic minimum set of programs are correctly loaded and working, allowing the system to bootstrap its way back into full operation. The hardware and software mechanisms used to achieve this basic level of recovery are described in some detail, and the methods used to secure these mechanisms themselves against the effect of fault conditions are also considered.

Introduction

The application of computer systems to real-time control situations is rapidly expanding. Many of these applications, such as air traffic control and communication systems, are essentially continuous activities which demand very reliable control systems. This means that the design of highly reliable computer systems is becoming increasingly important. This paper describes the methods used to secure a real-time, multi-processor system (System 250) against failure and discusses some of the problems involved in providing reliable system operation.

System 250 has been designed for communication applications, such as control of telephone switching, where continuous, reliable operation is required. A typical requirement of this type of application is a mean time between system failure of 50 years, where a system failure is defined as an outage lasting over ten minutes.

Previous papers[1,2] have outlined the overall configuration of System 250. The main features are that the system uses a group of functionally identical processor units connected to a group of identical store units. This type of configuration can be made to perform like one large, very powerful computer, and yet its power can be economically increased in small steps simply by adding more processor or storage units. Since all units are functionally identical any store module can replace any other store module, and similarly any processor unit can replace any other processor. This means that equipment failures can be catered for fairly simply. In the event of a unit failure the faulty unit is isolated and the functions of that unit are then reallocated to other modules in the system which have some spare capacity.

The software which is used to control this hardware configuration can usefully be considered as a number of distinct layers or levels[3]. As each new layer of software is added to the system it is used to extend, or present in a more convenient form, the facilities which are available. In effect the first layer takes the bare facilities provided by the machine instruction set and adds to them by providing further facilities within the software. This provides subsequent levels in the hierarchy with an enhanced version of the original machine, a kind of 'virtual machine'. The additional levels use this extended machine to produce further, more powerful facilities. Thus as one progresses along the hierarchy the facilities provided by the virtual machine at each level become increasingly useful and powerful.

In the case of System 250 the first software level takes the multi-processor, multi-store system and converts it into a virtual machine which appears to subsequent levels to be one large, very fast processor with one large store. All the problems associated with multiple processor operation are handled at this initial level, and subsequent layers need not consider the multi-processor nature of the system. The next level in the hierarchy provides convenient input/output facilities and controls the backing store devices such as discs, so that lower layers see a very much larger store system than that provided by the main store alone. The next level provides operator communication and facilities such as program assembly, editing, job control etc. Finally on the last level come the application programs which actually perform the real-time operations (Fig. 1).

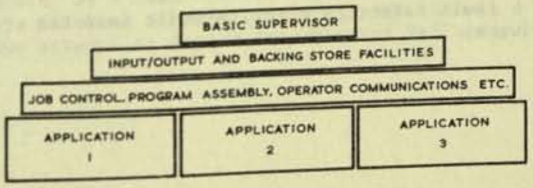


Fig. 1
Software Structure of System 250

There may be several sets of application programs in a system such as this. For example, one central control system may control several remote telephone exchanges. Other functions may be required which are related to, but not part of, the main real-time activity. For example, a maintenance sub-system to allow on-line testing may be added or a program development sub-system to allow new programs to be developed and debugged before being introduced into the real-time system.

Recovery Mechanisms

Based on this broad description of System 250 let us now consider the type of facilities and mechanisms which will have to be built into the system to allow it to recover automatically from fault situations.

Obviously the system will have to cope with failures within individual processor units and store modules, so that we require some means of detecting that a fault has occurred and locating the fault to a particular module. The faulty module can then be isolated so that it cannot interfere with the rest of the system. Finally any data which may have been lost or corrupted by the fault must be restored so that normal operation can continue. Typically this will involve reloading lost programs and data in the event of a store fault, and abandoning or reconstituting suspect data after a processor failure.

The system will also have to deal with software faults. On the basis of past experience it seems inevitable that even after thorough testing and commissioning all but the smallest system will still contain design errors in the software. This means that the system will occasionally behave unpredictably when certain, rather rare, combinations of data or timing circumstances occur. All that is required in this case is to reset any data which has been affected by the failure and restart processing using fresh data. This type of data recovery mechanism is similar to that required to deal with the after-effect of processor failures, as described above.

Thus, in general, each recovery action includes three distinct phases:- The first is the detection that an error has occurred. The second is an attempt to locate the fault to a particular hardware unit. This may not succeed, either because insufficient information is available, or because the fault is caused by a software problem. Finally the third phase will involve some form of data recovery or restart procedure which will allow the system to resume normal processing.

Within System 250 the mechanisms used at each stage of recovery are as follows:-

The error detection mechanisms which are used are:-

(1) fault detection circuits built into the hardware.

(2) software consistency checks and time-outs to monitor overall system performance.

(3) test routines run in background mode.

The methods used to locate the fault to a particular unit are:

(1) Persistent fault conditions reported by check circuits.

(2) If the error detection mechanism implicates a particular unit or units (for example hardware check circuits or test routines) a fault count associated with the unit or units can be incremented in order to detect persistently failing devices.

(3) A localised test procedure can be used to test units which are suspect as the result of an error indication from a hardware check circuit or failed test routine.

(4) The testing sequence can be extended to cover all units within the control system.

(5) As a last resort units can be switched out of system on a trial basis in an attempt to find a viable system configuration.

There is obviously a very wide range of data recovery and restart procedures which can be adopted. We have found it useful to adopt three stages of recovery action which provide progressively more extensive restart facilities. These are:-

(1) Process Restart Each process, or transaction, in the system has a defined recovery action which can be activated if that process meets any form of error condition. The recovery action involved will vary depending on the nature of the transaction, and these can range from regenerating data areas, and restarting the failed process in the case of a vital system function, such as a disc handler, to simply ending the failed process and printing diagnostics.

(2) Area Restart Each functional area within the system has a defined recovery action which will allow read/write data to be regenerated from duplicate files held on disc by that area. This may allow complete data regeneration, but more usually, some transactions will be abandoned and only the most important functions will be made restartable by storing redundant information on disc. This type of restart is commonly referred to as a 'warm start'.

(3) Area Reload Each functional area also has a defined recovery action which will allow processing to be restarted from read-only information in duplicate, sum-checked files held on write-protected areas of the backing store. This form of recovery obviously involves abandoning all current transactions, reinitialising the system and then resuming processing new

transactions. This type of restart is commonly known as a 'cold start'.

Recovery Procedures

We have now considered the basic elements which are available for use in constructing the required recovery procedures. Before moving on to discuss the form taken by these recovery procedures it is worth making the following observations:-

(1) the hardware test and data recovery procedures involved can themselves disrupt system operation, for example it is difficult to perform a complete test on every hardware unit in the system without causing some disturbance to normal on-line processing, and the various data recovery procedures often abandon perfectly valid transactions rather than attempt a complex validity checking operation.

(2) the error indications do not always pinpoint the source of the fault or the identity of the corrupt data. Processors may trigger hardware check circuits as the result of attempting to process invalid data corrupted by faults elsewhere in the system, and it is impossible to predict just how much data may have been disturbed by any given software fault.

This means that it is very difficult to adjust the recovery action so that the fault is corrected and yet the disturbance to system operation is minimised.

In the circumstances the best strategy is to combine the various fault location and data recovery/restart procedures into a sequence of recovery actions. Initially the action which causes least disruption to system operation is used. If this fails to clear the fault, as indicated by further error reports, then increasingly powerful (and hence more disruptive) recovery actions are used until the fault is cleared, as indicated by the absence of further error indications.

The sequence of actions which has been adopted in System 250 is shown in Fig. 2. Error indications which do imply the location of a fault (hardware check circuits and failed test routines) cause a fault count to be incremented for the unit, or units involved. If one unit is consistently implicated then the fault count indicates this. A local testing procedure for the suspect units is also activated. If either of these mechanisms detect a consistent fault the system is reconfigured to isolate the faulty unit. In the case of a hardware check circuit indication it is also necessary to restart the process which was running at the time of failure as the data associated with this transaction is now suspect. Repetitive errors detected by hardware check circuits within a short time interval suggest that the fault may be due to a software problem within the failing area rather than a hardware fault. Therefore in this case the recovery action is extended to cover the failing area rather than just the process involved.

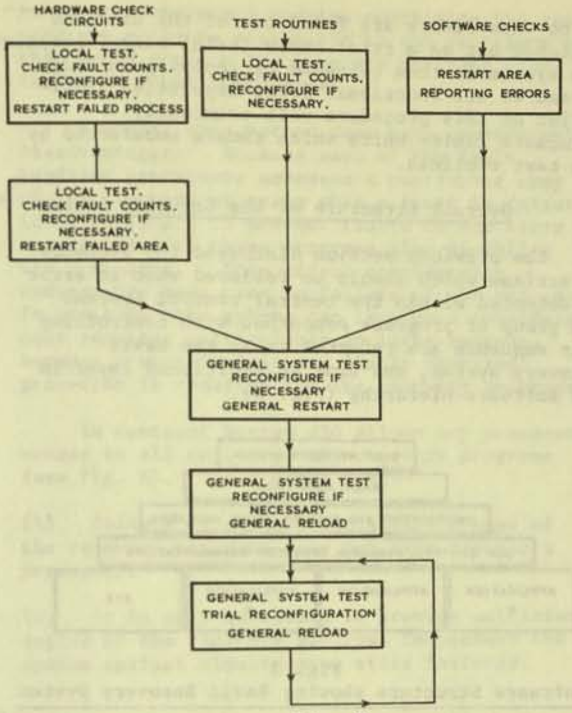


Fig. 2

Sequence of Recovery actions used by System 250

Faults detected by software checks cause a restart of the functional area detecting the fault. If the error is due to a software problem within that area this should clear the fault.

Further repeated error indications of any kind cause a general system test to be performed which thoroughly tests all control system elements. Any faulty units are isolated and the system is restarted by means of an Area Restart applied to all functional areas. This procedure will eliminate any data corruption in main store and will recover the vast majority of all solid hardware faults.

If further error indications are generated then another general system test is initiated in the hope of detecting possible intermittent hardware faults. Any faulty units are isolated and the system is reloaded by means of an Area Reload applied to all functional areas. This will reload the system using duplicate read-only files from backing store. This eliminates any possibility that further system failures can be caused by corrupted data generated by an earlier fault.

After this stage the only faults which can remain undetected are intermittent hardware faults or solid faults not detected by the test routines. Therefore, as a last resort, subsequent fault reports initiate another general system test in a further attempt to detect intermittent failures.

If no new faults are found one of the units is switched out on a trial basis (trial reconfiguration). The system is then reloaded by applying an Area Reload to all functional areas Repetitive application of this procedure will eventually eliminate faulty units which remain undetected by the test routines.

Overall Structure of the Security System

The previous section discussed the sequence of actions which should be followed when an error is detected within the central control system. The group of programs concerned with controlling this sequence are referred to as the basic recovery system, and form an additional layer in the software hierarchy (Fig. 3).

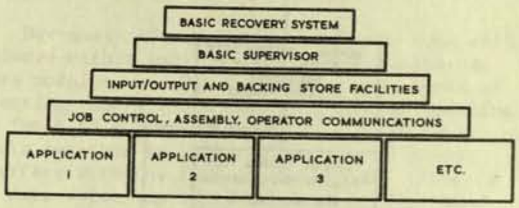


Fig. 3
Software Structure showing Basic Recovery System

When discussing the functions provided by the various levels in the hierarchy it was shown how the basic supervisor, which contains the scheduling and store allocations routines, effectively concealed the multi-processor, multi-store nature of the system from the lower levels. Programs involved in lower levels could be written on the assumption that they would run on one large processor with one large store. The basic recovery system performs a similar function in that processor and store failures are dealt with at this level, and lower levels in the hierarchy do not need to be concerned with the possibility of hardware failures. They can be written on the assumption that they are always held in a fault-free store module, and are obeyed by a fault-free processor. Thus although several copies of the basic recovery procedures must be available to protect this level against store failures, programs on lower levels do not need to be duplicated. If a store module fails, the programs held in that module will be reloaded into a new module by the basic recovery system. Therefore, placing the basic recovery system at the highest level in the hierarchy reduces to a minimum the amount of program which must be replicated. It also simplifies the system since lower levels do not need to consider the possibility of hardware faults.

The software checks required to provide an error detection mechanism should be distributed throughout the system so that each level contains its own independent set of checks. Similarly it is convenient to provide data recovery and restart procedures on a per level basis. This means that each level becomes an independent functional area, with its own set of software checks, and its own restart procedures.

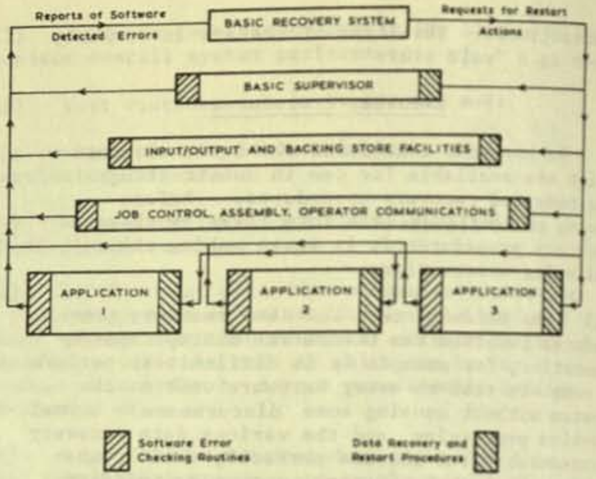


Fig. 4
Software structure showing Basic Recovery System & Communication Paths to the Rest of the System

The software checks report any errors to the basic recovery system which can then initiate the appropriate recovery action, which may involve invoking restart procedures provided by the lower levels. This modified hierarchical diagram is shown in Fig. 4.

This type of system structure means that as one progresses down the hierarchy not only do the number of facilities available increase, but it is also possible to make wider assumptions about the state of health of the system. Below the basic recovery system programs may be written on the assumption that all hardware faults have been eliminated from the system. The only responsibility that these lower levels have with respect to system reliability is to maintain an overall measure of performance through the software checks on that level, to report consistent faults to the basic recovery system on the assumption that the degradation is due to some form of system fault, and to provide the standard recovery procedures. Below the level of the basic supervisor it may also be assumed that reliable store allocation, and scheduling facilities are available, since it is the responsibility of the software checks and restart procedures within the basic supervisor to ensure this. Below the input/output level it may also be assumed that reliable system peripherals are available, and, for example, an application program written to test a particular piece of application hardware can ignore possible side effects due to faults on the input/output channels. It is the responsibility of the input/output routines within the operating system to eliminate these faults. This expanding level of confidence continues right down to the application/operating system interface where it may be assumed that processors, stores, input/output channels and system peripherals are working correctly and that the full range of operating system facilities is available. Of course, it is the responsibility of the application programs to cover the effects of faults in any special peripherals controlled wholly by that application.

Thus the overall reliability of the system is based on a hierarchy of guarantees. At the top of the hierarchy the basic recovery system provides fault-free stores and processors. Working from this base the other levels can then guarantee fault-free input/output devices and operating facilities to the application programs. By using this wider base the application programs can now secure their own specialised peripherals against failure. In many ways this hierarchy of guarantees parallels the functional build-up of the system, which is based on using the facilities provided by higher levels to make extensive or sophisticated facilities available to lower levels.

Securing the Security System

In the scheme outlined above everything depends on the ability of the basic recovery system to guarantee fault-free processors and stores to the lower levels. One of the main problems involved in producing a workable security system is to ensure that the basic recovery system itself is not disabled by fault conditions. Obviously several copies of these recovery programs must be provided in seperate store modules to protect them against store failure, and some form of protection must be provided to prevent these multiple copies being overwritten by a faulty processor. The recovery programs must also be accessible to several processor modules, to cover processor failures.

These requirements could be most easily met by nominating some, or all, of the processors as 'fault handling' units and providing each with a private store module containing a copy of the recovery programs, Fig. 5. In the event of a store or processor failure one, at most, of the store/processor pairs would be disabled and unable to take effective action. The other processors would then be able to clear the fault and recovery system operation.

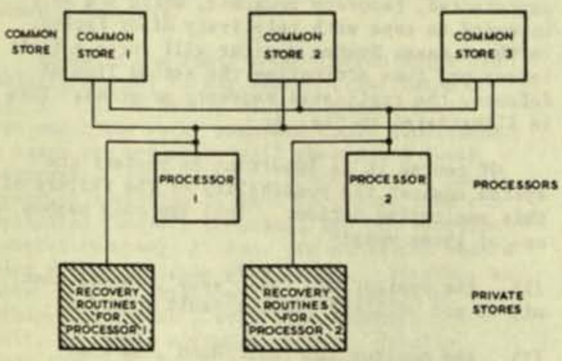


Fig. 5
System Diagram illustrating 'Store per fault-handling processor' approach

As each of the store modules containing the recovery programs would be accessible to one processor only, this would protect the recovery programs from faults in other processors.

However, this method does have considerable disadvantages. Because each of the fault handling processors accesses a particular copy of the recovery programs when a fault is detected it is difficult to prevent faults in the store associated with these programs also disabling the processor. This effect considerably reduces the mean time to failure of the processors. In addition this scheme can involve a considerable cost penalty, particularly in large systems, because a store module per 'fault handling' processor is required for the recovery programs.

In contrast System 250 allows any processor access to all copies of the recovery programs (see Fig. 6). This means that:-

(1) failure of a store containing a copy of the recovery programs does not also disable a processor.

(2) it is only necessary to provide sufficient copies of the recovery programs to protect the system against simultaneous store failures.

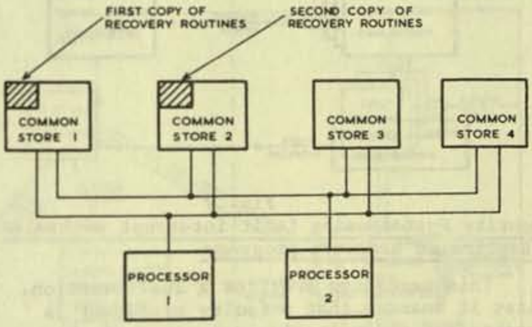


Fig. 6
System diagram illustrating System 250 Approach to Fault Handling

This arrangement is made possible by two features of the processor hardware:-

(1) the capability mechanism, which was described in a previous paper[1], provides a very secure store protection facility, and protects the recovery programs against the possibility of being over-written in the event of a hardware or software fault.

(2) the fault interrupt mechanism, also described in a previous paper[1], which together with the test program in the first section of the recovery programs, is used to control access to the recovery programs.

To illustrate this scheme assume, for the moment, that the only form of entry to the recovery programs is via a fault interrupt. This may be an involuntary interrupt resulting from an attempt to perform some illegal operation, or it may be a deliberate attempt to invoke the recovery mechanism because some error condition has been detected by the software. On taking a fault interrupt the PP250 hardware first disables all the current capabilities held in the machine, thus preventing further access to store. It then attempts to reload a new set of capabilities from a pre-designated location in store. If this is completed successfully the resulting capabilities are used to access the first part of the recovery program. This is a test program which is arranged as a maze. The only possible exit from this maze is via a further capability which is created bit by bit as the machine proceeds through a series of tests. These tests are designed to completely check the hardware and the 'read only' blocks (programs and data) associated with the recovery program. If an error is detected at any stage then another fault interrupt is forced. This causes the processor to reattempt the capability load from the next available store module (see Fig. 7).

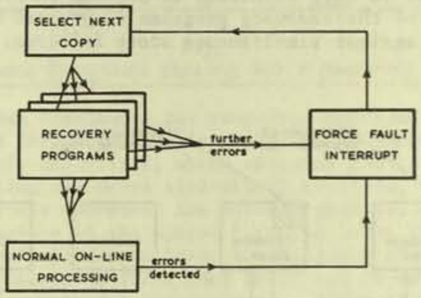


Fig. 7
Security System using fault interrupt mechanism & Replicated Recovery programs

This mechanism provides a dual function. First it ensures that a faulty processor is constrained to endlessly cycle through the storage system in an attempt to find a test program which it can obey successfully. The only capabilities available to the faulty machine at this time are associated with the test program, so that it is unable to interfere in any way with the operation of the on-line system. Secondly it allows fault-free processors to search through the storage system to find an uncorrupted version of the recovery programs.

The mechanism described above, although considerably better than the 'copy per processor' method, does have some disadvantages. The first is that before any recovery action can be taken the processor involved must obey a lengthy (100-200 msec) test program. The second is that all the recovery programs must be replicated. Both of these precautions are unnecessary in some fault situations where the fault is unlikely to disable the on-line system in any way, and the recovery action is fairly simple. For example, software faults which corrupt data within application programs are unlikely to affect the normal running of other programs. Once the fault condition is detected it is only necessary to activate the data recovery/restart routines for the particular application to recover system operation.

This rather minor kind of fault can be dealt with quite adequately by programs which exist in the on-line system and run in the normal way. However these programs do need some form of protection so that if they themselves are disabled by the fault, or are unable to cope with the fault situation in some other way, then the more powerful, replicated programs can be activated. Thus some form of monitor mechanism which can detect the failure of these unreplicated programs is required, as shown in Fig. 8.

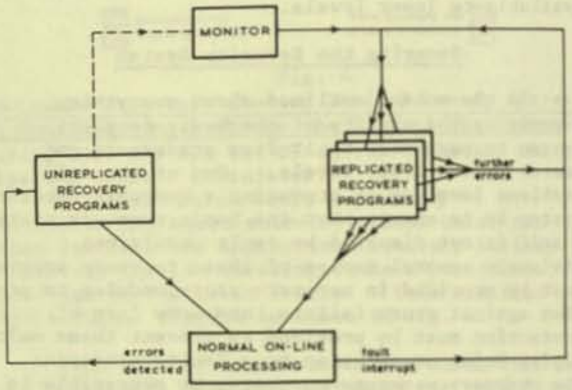


Fig. 8
Security System using a combination of replicated & unreplicated recovery programs

In the proposed implementation of System 250 the monitoring mechanism is made an implicit part of the unreplicated recovery programs. These recovery programs are activated by a process called the 'System Monitor'. This process runs at regular intervals and scans the system fault indicators. If any fault condition is detected then the appropriate recovery routine is activated. If persistent or multiple error conditions are detected then this implies that the fault is beyond the scope of the simple, unprotected, recovery programs, which are only intended to cope with relatively minor faults. In these cases System Monitor will force a fault interrupt, thus activating the second line of defence, the replicated recovery programs. This is illustrated in Fig. 9.

Of course it is important to protect the system against the possibility of the failure of this monitoring action. This can only happen in one of three ways:-

(1) the monitor can fail 'sane', detect that all is not well and force a fault interrupt.

(2) the monitor can fail 'dead', so that either it does not run at all, or does not perform any meaningful action when it does run.

(3) the monitor can fail 'crazy' so that it apparently runs correctly at regular intervals and yet does not respond to fault conditions.

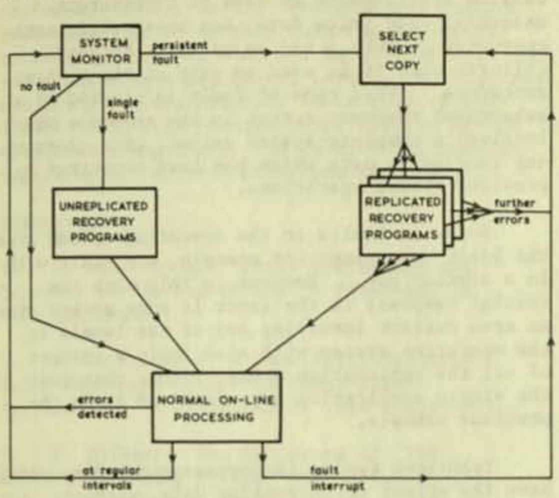


Fig. 9 .
Basic Structure of System 250 Security System

If the monitor fails 'sane' then the replicated recovery system is activated explicity by the monitor, and it can take effective action to recovery system operation. If the monitor fails 'dead' then an independent time-out mechanism is used to force a fault interrupt. This is equivalent to a periodic 'OK' signal which is used to reset a time-out, thus indicating that the system is operating correctly.

There remains the possibility that the monotor can fail 'crazy'. The probability of this happening can be reduced to any arbitrary level by incorporating sufficient self-checks into System Monitor, and ensuring that sufficient overlapping, independent software checks exist in the on-line system.

In general it is relatively easy to ensure that if the system fails then eventually, somewhere, one of the processors will generate a fault interrupt, thus activiating the replicated recovery programs.

There is one final modification which can usefully be made to the system illustrated in Fig. 9. It is fairly easy to ensure that even under the worst possible fault conditions at least one processor will generate a fault interrupt at some stage. Therefore the fault interrupt mechanism is used to ensure that the replicated recovery programs, and the associated powerful recovery actions, are activated when a major system collapse does occur. However, an isolated fault interrupt is symptomatic of nothing worse than a transient hardware fault, or simple software error. Ideally these should be dealt with by the unprotected programs, using recovery actions which cause minimum disruption to system operation. Only repetitive or multiple fault interrupts should drive the system into the rather more drastic recovery measures adopted within the replicated recovery programs.

This feature can be incorporated fairly easily. After a processor has successfully completed the test program which forms the first part of the replicated programs, it places a message in a location which is scanned at regular intervals by System Monitor. When this message is detected, the other error indications are checked together with a fault count for the processor which generated the message. Provided that this is an isolated occurrence the monitor process passes capabilities to the faulted proessor which allows it to rejoin the on-line system. If this particular processor has suffered a succession of fault interrupts it is assumed that it either has an intermittent fault, or a solid fault which is not detected by the test program. In either case it is not passed the capabilities which allow it to rejoin the on-line system but is forced back into the test program.

If System Monitor does not respond to the message then the assumption is made that either System Monitor has failed or that multiple error conditions have occurred. In this case the processor accesses the replicated recovery programs (Fig. 10).

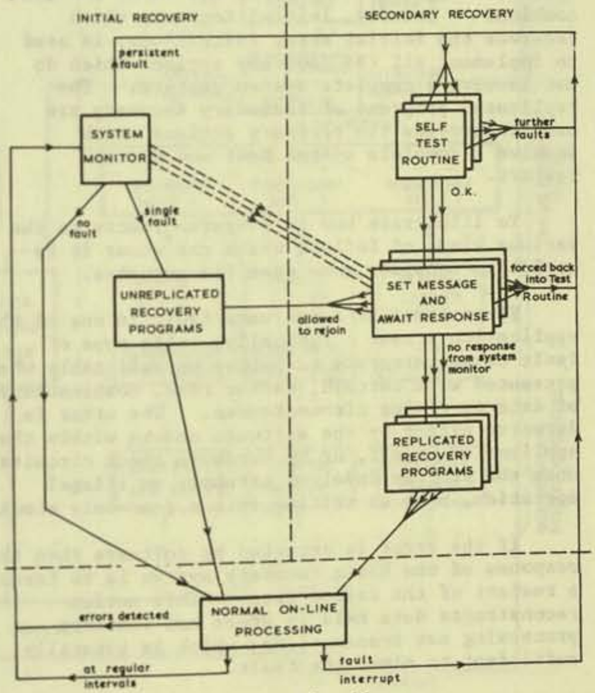


Fig. 10
Detailed Structure of System 250 Security System

Thus the general scheme is that in addition to the replicated programs which provide the basic level of recovery another group of programs is provided which run in the normal way as part of the basic supervisor. These programs form the first line of defence and provide a number of simple recovery actions which do not greatly disturb system operation. The general hierarchical structure shown in Fig. 3 is therefore extended by splitting the basic recovery system into two sections (Secondary Recovery and Initial Recovery). Only the essential kernal of this recovery system (Secondary Recovery) is replicated, and the rest (Initial Recovery) forms part of the basic supervisor (Fig. 11).

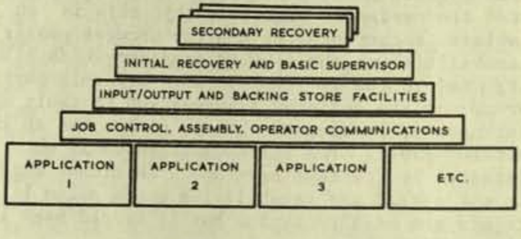


Fig. 11
General Operation of the Basic Recovery System

So far we have discussed the sequence of actions to be taken under fault conditions, and the overall structure of the basic recovery system. Fig. 12 illustrates how these two are combined . Briefly, Initial Recovery which receives the initial error indications, is used to implement all the recovery actions which do not involve a complete system restart. The replicated programs of Secondary Recovery are used to provide the recovery actions which involve a complete system test and general restart.

To illustrate how this system reacts to the various kinds of failure which can occur it is useful to consider some specific examples.

First consider a software fault in one of the application areas. Typically, this type of fault causes programs to behave unpredictably when presented with certain, rather rare, combinations of data or timing circumstances. The error is detected either by the software checks within the application itself, or by hardware check circuits when the program involved attempts an illegal operation, such as writing into a read-only block.

If the error is detected by software then the response of the basic recovery system is to force a restart of the failed area. This action reconstructs data held in store and restarts processing new transactions, which is generally sufficient to clear the fault.

If the error is detected by hardware then, after various hardware test procedures, the particular transaction involved is restarted. This may be sufficient to clear the fault, but if it is not then subsequent faults will force an area restart.

In very rare circumstances the area restart may fail to clear the fault. This can only happen if the duplicate information held on backing store, which is used to reconstruct essential read write data, has been consistently corrupted in such a way as to cause further failures when it is used as part of the restart procedure. This type of fault is cleared by a subsequent recovery action in the sequence which involves a complete system reload, thus clearing any read/write data which has been generated by previous system operations.

Software faults in the operating system area, the basic supervisor for example, are dealt with in a similar way. However in this case the initial response to the error is more severe since an area restart involving any of the levels in the operating system will also imply a restart of all the application areas, rather than just the single application area involved as in the previous example.

Transient faults in processors or store modules have the effect or corrupting data, without permanently disabling a hardware unit, so that the immediate after-effects are indistinguishable from software faults. Thus the remarks made above also apply to this type of failure mode.

Consider the possibility of a processor fault. Recent trials on the system indicate that faulty processors usually take a fault interrupt very quickly after the incidence of a fault, within one or two milliseconds. Also recent tests have shown that the 'fault capture' level of the test program, which is obeyed by a processor after taking a fault interrupt, is very good, better than 99.5%. Thus the vast majority of processor faults will very quickly cause the faulty machine to take a fault interrupt. It is then trapped in the maze of the test program, which isolates it from the rest of the on-line system.

In general store faults will have an obvious and immediate effect on the system. Usually all the processors receive a parity fault indication very soon after the fault has occurred. This effectively disables the on-line system so that recovery is achieved through Secondary Recovery via a general test of the system and a complete warm start.

Hardware faults which are not located by the test routines, either because they are intermittent or beause the test routines are not comprehensive enough, are difficult to recover. They may be located by means of fault counts, or in the case of intermittant faults by repetitive use of the hardware test programs. However if none of these mechanisms do locate the fault then the final, last ditch, action taken by the recovery system is to attempt to find a viable configuration by means of trial recongiguration. How quickly this is achieved depends on the nature of the fault. If the fault is seriously affecting system operation, so that its effects can be detected very earily, then a medium sized system can work through all possible combinations of the central control equipment in something like two minutes. If the fault only causes the failure of the occasional transaction then the system is performing useful work. Provided the reduced performance is acceptable then the automatic recovery mechanisms will not be activated at all, since the system is, to all intents and purposes, operating satisfactorily. This type of non-urgent fault will eventually be cleared by the maintenance engineers, who receive information regarding all error indications recorded.

References

1. K. Hamer-Hodges 'Fault Resistance and Recovery within System 250' - Presented at this conference.

2. D. C. Cosserat 'A Capability Oriented Multi-Processor System for Real-Time Applications' - Presented at this Conference.

3. E. Djikstra 'The Structure of 'THE' Multi-programming System" Com. A.C.M., Vol. 11, No. 5 May, 1968, pp. 341 - 346

Acknowledgement

I would like to thank the many colleagues on whose work this paper is based and the Directors of the Plessey Company for permission to publish it.

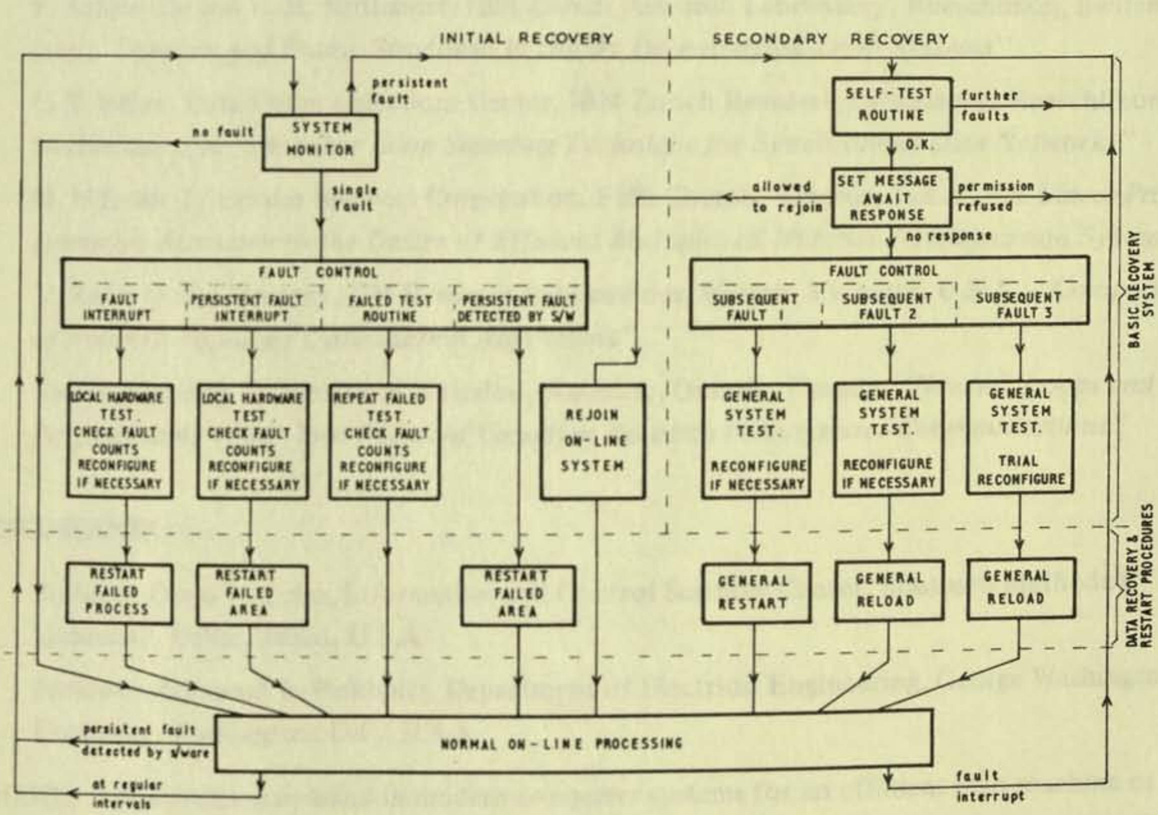


Fig. 12

Implementation of System 250 Security System

DATA NETWORK DESIGN PROBLEMS I

CHAIRMAN:
Dr. Paul Oliver, UNIVAC
Washington, D.C. U.S.A.

PROGRAM COORDINATOR:
Dr. Carl Hammer, Director,
Computer Sciences, UNIVAC,
Washington, D.C., U.S.A.

PAPERS:

P. Zafiropulo and E. H. Rothauser, IBM Zurich Research Laboratory, Rueschlikon, Switzerland, *"Signaling and Frame Structures in Highly Decentralized Loop Systems"*

C. S. Nokes, Data Communications Center, IBM Zurich Research Laboratory, Rueschlikon, Switzerland, *"A Subscriber Loop Signaling Technique for Synchronous Data Networks"*

M. Millman, Computer Sciences Corporation, Falls Church, Virginia, U.S.A., *"A Linear Programming Approach to the Design of Efficient Multiplexed Wideband Transmission Systems"*

V. Kevin Moore Whitney, GM Research Laboratories, Warren, Michigan, U.S.A., *"Comparison of Network Topology Optimization Algorithms"*

Eric G. Manning, University of Waterloo, Waterloo, Ontario, Canada, *"Newhall Loops and Programmable TDM: Two Facets of Canadian Research in Computer Communications"*

DISCUSSANTS:

Professor David L. Cohn, Information and Control Sciences Center, Southern Methodist University, Dallas, Texas, U.S.A.

Professor Raymond L. Pickholtz, Department of Electrical Engineering, George Washington University, Washington, D.C., U.S.A.

THEME: The increasing demand in modern computer systems for an efficient man-machine or machine-machine interface suggests that we are quickly approaching the point where most systems will be operating in a communications-oriented environment. This trend is further stimulated by financial and response-time considerations. Unfortunately, it is often the case that a computer network, rather than solving problems, simply shifts them from the data processing to the data communication parts of the system. The papers in this session address themselves to the techniques and considerations required in the design of cost-effective data communication networks. The design approaches and the types of networks to which these are applied differ considerably within the framework of the common goal; the improvement of cost-performance.

SIGNALING AND FRAME STRUCTURES IN HIGHLY DECENTRALIZED LOOP SYSTEMS

P. Zafiropulo, E. H. Rothauser
IBM Research Laboratory
Zurich, Switzerland

Summary

Today control in digital communication systems is practically always centralized. But in small systems the future growth of local control capability may start a trend towards greater decentralization. The paper considers such a future environment. The paper investigates the signaling procedures and frame structures necessary in highly decentralized communication systems. In order to keep the problems within manageable size we have limited ourselves to single loop systems.

1. Introduction

As terminals and their requirements become more sophisticated, additional control intelligence will get placed at the terminal. This tendency will be accelerated by the fact that the implementations of control capabilities, be it microprogrammed or stored-programmed, is becoming cheaper. Today control in digital communication systems is practically always centralized. But in small systems (up to several hundred terminals) this future growth of local control capability may start a trend towards greater decentralization.[1] This paper considers such a future environment which allows for sufficient control intelligence at the terminals for the realization of highly decentralized fully digital small communication systems.

The paper investigates the signaling procedures and frame structures necessary in highly decentralized communication systems. In order to keep the problems within manageable size, we have limited ourselves to single loop systems[2] and there again to two issues which we feel worthy of attention:

- signaling and frame structure for transmission over circuit-switched links for synchronous data such as speech,
- signaling and frame structure for the handling of data messages, using a packet-switched mode of communication.

The paper demonstrates the feasibility of signaling procedures for a communication system where the terminal controller builds-up, monitors and disconnects its own communication link, without the support of a central intelligent supervisor.

2. General Description of the Considered Class of Decentralized Loop Systems

2.1 Transmission between terminals

Figure 1 shows the basic system configuration. The terminal units attach to the bus line via connectors. The transmission is digital and the bus line exhibits unidirectional signal flow. The connectors read data from and insert data onto the bus line. The total data stream passes through every connector, therefore connectors retransmit data not destined for them without modifications and distortions. The bus line forms a closed loop, therefore the unidirectional transmission enables not only half-duplex, but also full-duplex message exchange. Terminals set up and monitor their own communication links.[3] To be more precise, each terminal is attached to a control unit which fulfills these duties. This terminal controller consists of a dedicated unit. When a terminal wishes to establish a communication link, the terminal controller looks for a free slot in the periodical data frame structure and occupies it. It then uses the occupied slot to establish contact with the receiving terminal.

The tasks of the loop controller consist of initiating and maintaining a data frame structure which can serve as a reference to separate the time-shared traffic between the attached terminal units. The loop controller is transparent to the bit stream. In a practical system additional features are necessary, e.g., connections to the outside world, to data banks and processing facilities, as well as the availability of certain services. We have assumed that these features can be handled by terminal-like units.

2.2 General frame structure

The heart of the system is the loop frame structure. A fixed number of initially generated bit-positions travel round the loop. All signals present on the loop at a given time instant are grouped to form a single frame. Beginning and end of the frame are separated by a gap called the interframe gap (Fig.2). The signal stream is unidirectional thus both frame and interframe gap rotate periodically round the loop. The time it takes for the frame to rotate once round the loop is the frame cycle time.

Terminals communicate by reading from and writing onto preassigned bit positions in the frame. To enable such a process each terminal controller is equipped with a bit-counter which is incremented every time a binary signal passes that terminal. These counters have a finite capacity and must therefore be periodically reset to zero. This last condition is accomplished by the interframe gap. When a terminal detects the interframe gap, a time-out feature resets that terminal counter to zero. The interframe gap guarantees that all terminal controller bit counters get periodically reset. This technique has the advantages that it ignores the state of a terminal counter at the reset instant thereby correcting erroneous counts before the arrival of the next frame.

The bit counters allow to superimpose various communication services on different sections of the frame. By introducing in each terminal-controller a gating system controlled by its counter, terminals can initiate read/write operations solely in prespecified frame sections. This method is used to separate the traffic from terminals which transmit information in a circuit-switching mode from traffic which utilizes a packet-switching mode.

3. A Control Algorithm for Circuit-Switched Connections

It seems good practice to assign a fixed segment of the available channel capacity throughout a connection to terminals which want to communicate with each other at some constant bit rate for a longer time.

The most prominent example is the telephone service in which 8 bit characters are transmitted at a rate of 8 kHz.

3.1 Slot structure and signaling

The frame segment used for circuit-switching is subdivided into M contiguous slots of equal size (Fig.3). Two terminals communicate with each other using one such slot, hence as many simultaneous full-duplex links are possible as there are slots in the segment. Let terminals A and B communicate with each other via the nth slot.

When the nth slot passes either terminal, a bit interleaved read/write operation takes place.

Terminal units set up their own communication links in the following way:

Event sequence at calling terminal:
1. Search for free slot.
2. If free slot detected occupy it and memorize slot position.
3. Insert destination address in occupied slot.
4. If called terminal responds with an off-hook message, begin communication.

Event sequence at called terminal:
1. If own terminal busy, ignore address in passing slots (ignore calling terminal).
2. If own terminal free search for slots containing own address (search for calling terminal).
3. If own address detected, memorize slot position. Respond by inserting in same slot ring back signal.
4. If own terminal goes off hook insert off-hook message and begin communication.

To perform these routines, those for disconnection, as well as to enable extra signaling for special features, each slot must be able to carry the following state information: free slot, occupied slot, address information, ring-back signal, transmission mode, on hook (terminate), and various control information for special features.

Thus each slot must be able to transport seven different control states, the data samples, address codes and special feature codes. The chosen coding method within the slots will now be described. The scheme considered is such that never more than one bit delay occurs at each terminal connector. This is an important condition, because the signal delay around the loop should not become too large. Furthermore it reduces the number of circuit elements in the critical loop path to a minimum, thereby increasing transmission reliability.

Each slot is subdivided into two compartments (Fig.3). Compartment-1 is permanently reserved for control purposes. Compartment-2 is primarily used for the transmission of either speech samples or address information. During the transmission build-up process the entire destination address must be inserted into one slot. Should a destination address require say x more bits to be transmitted than can be inserted into a single compartment-2 then up to "2 to the power x" terminals will respond to the transmission of the first address section. Hence the maximum number of attachable synchronous terminals on the loop must not exceed the number of addresses provided by compartment-2. In a band limited situation compartment-1 should contain the least possible number of bits. Endeavoring to achieve this goal has led us to develop a coding method which is highly immune to transmission errors and therefore interesting on its own merit.

Compartment-1 is chosen to contain two bits. Transmission of the seven different states becomes only possible if the information is distributed over more than one frame cycle. This fact led us to design a continuous mode of signaling rather than a spurt mode. High immunity against nonsystematic transmission errors is thereby achieved. A repetitive pattern will always recover from a non-systematic error so that the involved terminals revert to the desired state. Such immunity considerations are more crucial in decentralized systems than in centralized ones. In the latter systems the central control unit can contain sophisticated error recovery routines whereas it becomes unreasonable, in decentralized systems, that each terminal unit contain such routines. The penalty for having only two bits in compartment-1 is a certain reduction in the speed of establishing communication links.

3.2 The proposed control strategy

The two bit positions in compartment-1 of each slot serve permanently for control purposes. The first bit allows differentiation between free and busy slots. A free slot is not monitored by any terminal. It is characterized by the first bit being either permanently 1 or 0. Active terminals modify the first-bit value of slots they are using. Hence a busy slot is characterized by a first-bit value which changes. It will be shown that the first-bit value of a busy slot changes at least once every three consecutive frame rotations. Therefore, a terminal searches for a free slot by monitoring consecutive slots starting with slot one. The first slot, in which the terminal observes no first-bit modification during three consecutive frame cycles is free. The terminal occupies the slot by immediately modifying the first-bit value. This immediate action guarantees that never two or more terminals, searching for a free slot, will occupy the same slot. An error modifying a busy slot's first-bit at a crucial time could cause a terminal to consider that slot to be free. If, however, a terminal observes a slot during more than 3 consecutive frame cycles before occupying it, then the system can tolerate erroneous first-bit modifications without an occupancy error occurring. The ensuing penalty is a larger call set-up time.

This busy/free technique has the advantage that non-systematic errors cannot lead to permanent status modifications. Hence no extra control is needed to observe whether occupied slots are actually being used. If the busy/free information were carried statically, i.e., by a constant bit pattern, single bit errors could modify erroneously a free slot into a busy one. An unlucky sequence of such errors would be able to cripple the communication capacity of the system.

A second purpose of this first bit is to describe whether the other terminal is on- or off-hook. The necessary algorithm is contained in the following laws:

LAW 1: If a terminal modifies the first bit of a slot during the nth frame passage it will retransmit that bit without modification during the (n+1)st frame passage.

LAW 2: If a terminal does not modify the first bit of a slot during the nth frame passage, it will modify that bit during the (n+1)st frame passage if and only if it is unchanged (i.e., if the received bit value is identical to the previously received bit value).

These laws apply only to active terminals which occupy a slot. They generate the first-bit pattern 1A of Fig.4 when only one terminal occupies a given slot. The patterns 1B and 1C are generated when two terminals, (X and Y) occupy the same slot. Then if pattern 1B is observed on the X-Y loop section, pattern 1C will appear in the Y-X loop section and vice versa.

When a called terminal goes off-hook, pattern 1A is superceded by the pattern combination 1B/1C. The sending terminal easily detects this change and is thereby alerted to the new state of the receiver. Conversely, when one of two terminals, involved in a communication, goes on-hook, the pattern combination automatically reverts to pattern 1A. This modification is again easily detected by the remaining active terminal and it thereby recognizes that its partner has gone on-hook.

The described method has the important advantage that if one terminal goes either off- or on-hook the partner will eventually detect this even though transmission errors occur. Other advantages of the proposed scheme are:

. no master-slave relationship ever exists between two terminals communicating with each other: i.e., both terminals use identical first-bit modification routines.

. If in a communication link one terminal goes on-hook, then the other automatically keeps occupying the slot and can therefore directly initiate a new call.

Now the role of the second bit in compartment-1 will be discussed. When a slot has to carry an address in compartment-2 then the code in compartment-1 must be such that terminals immediately know this upon reading the second bit in compartment-1. Assuming that address information is depicted by the "1" value of the second bit, then all other signal and control states must be coded with a "0" valued second bit. How this is done will be shown by describing the coding procedure which initiates a speech link.

A terminal wanting to set up a communication link occupies a free slot by generating pattern 1A. When this calling terminal has the entire receiver address assembled and coded in binary form, it sets the second-bit value to "1" and inserts the address in compartment-2. Idle terminals observe the second-bit value of each slot. If they detect a "1" they read the content of compartment-2. When the address is not their own, they undertake nothing and start observing the next slot. When the address is their own they continue to observe only that slot and stop monitoring the others.

Let us assume that the called terminal has been idle and it detects its address during passage of the nth frame cycle. It can only recognize the address after reading the last bit of compartment-2, by then compartment-1 has already been retransmitted , hence the terminal undertakes nothing. During the next frame-cycle (n+1) the called terminal will, after reading the second bit, immediately set that bit to "0".In this way the called terminal generates the ring-back acknowledgement. If the called terminal detects that during the (n+1)st cycle the second bit arrives again with value "1" and compartment-2 again contains the own address, the called terminal will, during the (n+2)nd frame cycle, continue to set the second bit to "0". On the other hand, if neither compartment-2 contains the address nor the second-bit value is "1" then the terminal will revert to the idle state. As long as the called terminal remains on-hook and modifies the second-bit value to "0", it will ring. And as long as the calling terminal detects a "0" valued second bit it will generate the ring-back signal. On the other hand, if the calling terminal detects "1" valued second bits it will generate the busy signal. If now the called terminal goes off-hook it will begin to modify the first bit and thereby modify the bit pattern to either 1B or 1C depending on the off-hook instant. Upon detecting this new pattern the calling terminal will know its partner has gone off-hook and will interpret the contents of compartment-2 as text (speech) samples, it will at the same time refrain from presetting the second bit in compartment-1 to "1" and will start to insert its own samples into compartment-2.

The reader can check for himself that the setting-up of a communication link is not seriously harmed by single errors in any bit position. This high reliability is due to the fact that the signaling information is continually retransmitted. Hence single errors cannot destroy the control information over appreciable periods of time and the system recovers. For example, if an error generates the address code in an idle slot, and if the bit pattern in compartment-2 of that slot happens to match a terminal address, then that terminal will detect its own address. During the next frame cycle that terminal will modify the second-bit value to "0". The slot is idle and therefore its control bit pattern remains unchanged. Thus on the next detection of the slot the terminal in question will not detect the address code and will therefore return to its idle mode. The only effect this error has upon the system is to generate a ring for the duration of one frame cycle.

The proposed control scheme suffices to set-up and disconnect links but additional control codes will be required to set-up external calls and to implement various special features. These additional required control signals are obtained in the following way: if a terminal wants to transmit additional control information to its partner it violates intentionally Laws 1 and 2 by modifying the first-bit value in every frame cycle, i.e., by generating pattern 1D shown in Fig.4. The other terminal is inhibited by Law 2 from modifying the first-bit value. Hence, when it detects that it has had no chance to modify the first bit during four consecutive frame cycles it knows that compartment-2 contains control information.

The control mode just described would allow implementation of various features including transfer, hold for inquiry, multi-way calling, or changing the numbering plan which, due to the highly decentralized nature of the system is only possible by transmitting to the terminals their new numbers.

3.3 Application potential

Up to now we have considered a loop organisation in which the frame rotates around the loop at constant speed and consists of slots of equal size. Obviously one will optimize the channel capacity of the possible communication links for the most important service to be carried. Let us assume the system will be tailored to the needs of speech transmission. Then we may fix the frame cycle time to 125 µs, corresponding to 8 kHz, and the slot size to 10 bits. The channel capacity per slot is then 64 Kbits/sec. Data terminals[4] may also want to use this communication mode. Transmission efficiency will then depend on the speed of these terminals relative to the 64 Kbits/sec transmitted by a single slot. Fig.5 shows that for high data rates efficiency is very good. However, the method is only economic if transmission times are also long compared to the time required to set up communication links. In this set-up time the search for free slots has to be included.

For short data messages or low-speed terminals another switching method is more appropriate which is described in the next section.

4.Control Algorithm for Packet Switched Connections

4.1 Slot structure and signaling

We are now concerned with data terminals which cannot effectively fill one of the circuit-switched links described in the previous section. A reasonable way to satisfy their transmission requirement is to share one or more data slots among these terminals. Sharing a slot among different terminals simultaneously requires the insertion of a priority scheme, if the system is to enable fast response.

This is a requirement, if the system is to service without buffering autonomously clocked terminals, e.g., card punchers or synchronous modems. Furthermore, terminal connectors should never delay the frame by more than one bit period. Considering these requirements the satisfactory sharing of one slot among many terminals becomes rather difficult in a highly decentralized environment.

The proposed method[5] will now be described. The slot structure required for it is shown in Fig.6. A terminal occupies the slot by modifying the free/busy bit value through consecutive frame cycles. Thus any terminal can detect whether the slot is free or occupied by simply observing the free/busy bit during two consecutive frame cycles. If the free/busy bit values are different, the slot is busy; if they are identical, the slot is free. The advantages of this occupation method have already been discussed.

All data terminals are classified into priority classes. The highest priority is associated with the number one; the second highest with the number two, and so on. The lowest priority is associated with the number n. Both priority and priority-reserve compartments contain n bit positions. The i-th (i = 1, 2, 3...n) bit position in each of these compartments is associated with terminals in the i-th priority class.

Assume a terminal belonging to the i-th priority class wants to occupy the slot. For clarity's sake, this terminal will be denoted by Terminal A.

Terminal A begins by observing (i.e., storing) the busy/free bit value and inserting a "1" in the i-th bit position of the priority-reserve compartment. Each time the slot passes the loop controller the entire bit pattern contained in the priority-reserve compartment is shifted into the priority compartment after which all ppriority-reserve compartment bits are set to "0". During the next frame cycle, terminal A observes the priority compartment. If all bit positions up to and including the (i-1) bit have "0" value, no terminal of priority higher than i are requesting the slot. Therefore if the slot is free, terminal A occupies it. On the other hand, if at least one of the bit positions, up to and including the (i-1) bit, have "1" values, terminal A refrains from occupying the slot so as to let some higher-priority terminal use it first.

If terminal A can occupy the slot and the slot is free, it modifies the bit value of the passing busy/free bit. It also retransmits the i-th bit of the priority reserve compartment without modification. On the other hand, if occupation is not possible, terminal A repeats the initial action, i.e., it stores the passing free/busy-bit value and inserts a "1" in the i-th bit position of the priority-reserve compartment. By inserting a "1" in the i-th bit position, terminal A keeps on requesting the slot so that lower-priority terminals do not access the slot before terminal A has satisfied its transmission requirements.

The procedure described enables terminals of a given priority to satisfy their transmission requirements before terminals of lower priority can do so. It also enables a high-priority terminal, requesting the slot, to interrupt the transmission of lower-priority terminals. For example, assume that during occupation of the slot by terminal A a terminal with higher priority, say priority (i-2) undertakes an occupation request. Then terminal A will, during the next frame cycle, detect that the (i-2) bit value of the priority compartment is "1" and will therefore refrain from reoccupying the slot. Using a priority as well as a priority-reserve compartment is a necessary requirement if terminal connector delays are smaller or equal to one inter-bit delay.

Terminals belonging to the same priority category can share the data slot in different ways. We shall consider two cases. In the first and simplest case a terminal of, say, the i-th priority category relinquishes the slot if either a terminal of higher priority requests the slot or the terminal has terminated transmitting. One important characteristic is that this approach has no built-in procedure to limit the occupation duration of a single terminal. Another characteristic is that if priorities are not required, both priority and priority reserve compartments can be eliminated.

In the second case a terminal relinquishes the slot either if a higher-priority terminal requests the slot or the terminal has terminated. But in addition it relinquishes the slot if one or more terminals of the same priority have made occupation requests. The following rules enable this automatic relinquishment to terminals of identical priority.

1. A terminal never modifies any bits in the priority-reserve compartment during the time it actually occupies the slot.
2. When an occupying terminal of priority i detects that the ith bit of the priority compartment is 1 it immediately relinquishes occupation. If the terminal has thereby been interrupted it begins again setting the ith bit of the priority reserve compartment to 1 and continues doing this till it regains access to the slot.

It can be shown that, with this automatic relinquishment mechanism, if a plurality of terminals with equal priority simultaneously overlap in their request for the slot, and higher-priority terminals make no requests, then they share the slot in the following manner. No terminal ever occupies the slot for more than one consecutive frame cycle, but every terminal occupies the slot equal amounts of time. If, on the other hand, only one terminal of say priority i is requesting service and no terminals of higher priority are requesting service then that terminal is able to occupy the slot in a continuous manner. Both priority and priority-reserve compartments each reduce to one bit position if no priority is required.

4.2 Application potential

We have not described how one data slot can be shared among many terminals. The next step would be to consider how data terminals use the slot to communicate and how they use it to initiate communication links. Such a study for the major data terminal types is beyond the scope of this report. We shall therefore limit ourselves to describing the communication facilities which the data slot can offer.

The data slot of Fig.6 contains two address compartments, a data compartment and a flag bit (or bits). The flag denotes whether the contents of the data compartment is text or signaling information. By text, is understood all information which does not get interpreted by the terminal controllers as signaling information. When a terminal controller occupies the slot, it inserts both the destination and its own address into the respective address compartments. It then sets the text/sign flag bit either to "1" or "0" depending on what kind of information it is going to send.

Terminals continually monitor the receiver address compartment. If they detect their own address they proceed to read and store the rest of the data slot content, i.e., the second address and data compartments as well as the text/data flag. The contents of the second-address compartment allows a receiving terminal to know the identity of the source terminal. This second-address compartment is a necessary requirement when initiating a data communication link and when a terminal controller simultaneously supports several communication links. The latter case occurs for example when a time-shared computer is attached to the system.

Different forms of data transmission are possible with this slot. For example, if terminal A and terminal B communicate and terminal A is occupying the slot, then terminal B can acknowledge or transmit new information back to terminal A via the slot's return trip (any terminal occupying the slot does so for at least one complete frame cycle). Alternatively, terminal B can undertake its own slot-occupation request to communicate the acknowledgement or new information back to terminal A.

Finally, different signaling methods and procedures can be implemented with the text/signaling flag-bit approach. This might be an advantage when we consider how heterogeneous data terminals can be.

5. Conclusions

The present paper study endeavors to show that highly decentralized communication systems are indeed technically feasible. Two modes of operation in a single-loop system were considered; circuit-switched synchronous transmission and packet-switched data transmission. It became clear that in both cases the problems can be solved which we considered in this present study most important:

- making, maintaining and breaking of connections and
- system recovery from non-systematic transmission errors.

Stimulated by finding interesting and probably readily implementable solutions to the above problems, we feel that systems utilizing highly decentralized control procedures merit further studies.

References

1. Newhall, E.E., and Venetsanopoulos, A.N., Computer Communications: Representative Systems, IFIP Congress 1971, Ljubljana.

2. Farmer, W.O., and Newhall, E.E., An Experimental Distributed Switching System to Handle Bursty Computer Traffic, ACM Symposium on Optimisation of Data Communication Systems, Pine Mount, Georgia, 1969.

3. Thomas, D.A. and Startford, B., Fernmeldeanlage für Zeitmultiplexbetrieb, Swiss Patent 502043.

4. Dewitt, R.G., An Experiment in Transmitting Computer Data, Bell Laboratories Record, March 1967.

5. Gall, D., and Müller, H.R., Waiting-Time Distributions and Buffer Overflow in Priority Queueing Systems, IEEE International Conference on Communications, Philadelphia, 1972.

6. Closs, F.H., Müller, H.R., Port, E., Rudin, H., and Wild, D., Methods of Time-Division Mulitplex Communication and Switching Unit, Swiss Patent Number 514268.

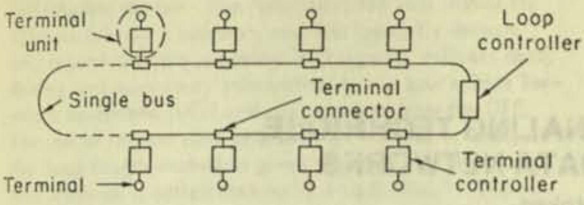


Fig. 1. Basic system configuration.

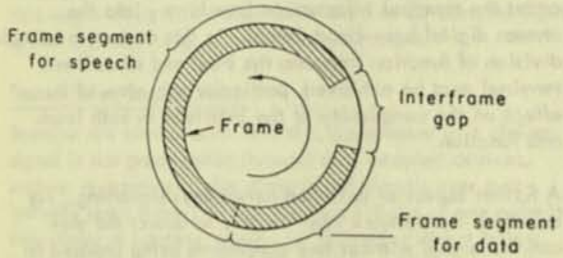


Fig. 2. The basic frame structure.

Fig. 3. Slot structure for circuit switched communication.

		Frame cycle										
		1	2	3	4	5	6	7	8	9	10	11
First-bit pattern	1 A	1	1	0	0	1	1	0	0	1	1	0
	1 B	1	1	0	1	1	0	1	1	0	1	1
	1 C	1	0	0	1	0	0	1	0	0	1	0
	1 D	1	0	1	0	1	0	1	0	1	0	1

Fig. 4. Control patterns described by the first bit of compartment-1 in consecutive frame cycles.

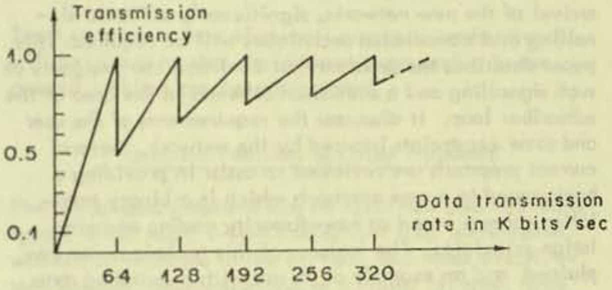


Fig. 5. Transmission efficiency versus transmission rate when using speech slots for data communication. The efficiency is with respect to compartment-2 sections of the slots.

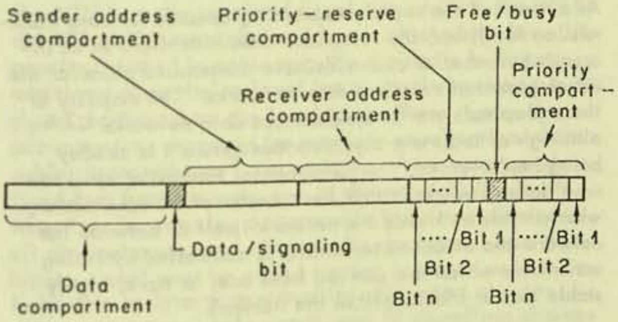


Fig. 6. The data slot structure.

A SUBSCRIBER-LOOP SIGNALING TECHNIQUE FOR SYNCHRONOUS DATA NETWORKS

C.S. Nokes
International Business Machines
Ruschlikon, Switzerland

Abstract

As a result of the rapidly expanding demand for data transmission facilities, the Telecommunications Administrations in an increasing number of countries have announced plans for the development of switched data networks. With the arrival of the new networks, significantly different signalling and transmission techniques will be required. This paper describes the problem that confronts the designers of such signalling and transmission schemes in the area of the subscriber loop. It discusses the requirements of the user and some constraints imposed by the network. Several current proposals are reviewed to assist in providing a background to a new approach which is a binary translation scheme based on low-disparity coding and violation principles. The features of this technique are explained, and an example of its use with a switched data network is given. It is shown that this technique may be simply implemented and could have advantages for both the user and the Telecommunications Administration.

Introduction

As a result of the rapidly expanding demand for data transmission facilities, the telephone administrations in an increasing number of countries have announced plans for the development of switched data networks. The majority of these proposals are for synchronous data networks [1,2,3] although at least one asynchronous network is already being implemented.[4] A fundamental feature of synchronous data networks is that the transfer of all information, whether this be inside the network itself or between the network and attached terminals, is controlled by timing waveforms which are derived from one, or more, highly stable master clocks within the network.

At the present time, most users of teleprocessing equipment lease telephone lines from the national Telecommunications Administration, and so build their own dedicated networks. A smaller number of users employ the switched network for data transmission, particularly if their calling rate is low. Implicit in the use of the telephone network, with either switched or leased connections, is the need for modems, which convert the user's binary signal into a waveform which is suitable, due to its frequency spectrum, for transmission through the telephone network.

With the advent of synchronous data networks, a new situation arises. Base-band transmission over the subscriber loop can now be used since the frequency spectrum of the transmitted signal no longer has to correspond to the requirements of the telephone network. This means that a modem is not needed at the customer's premises, but rather a Network Terminating Unit (NTU) which provides network timing information to the terminal, and possibly codes the terminal information from binary into the chosen digital base-band format. In this respect, a suitable division of function between the NTU and subscriber's terminal must be achieved, particularly in view of its effect on the complexity of the interface in both leads and function.

A further aspect of switched networks is signalling. For example, the network must be able to detect the disconnect signal without any constraints being imposed on the customer concerning the way he transmits data. In essence, there has to be a way of distinguishing signalling, which controls a call, from data, which is simply passed through the network between terminals.

It has been estimated[5] that the costs involved in the local part of a synchronous network, as opposed to the high-level trunk circuits and switches, can be in the order of 70 percent of the total cost. It is not surprising, therefore, that much effort has been devoted to devising schemes for solving these new local signalling and transmission problems. The problems that the designer has to face are strongly interrelated, and he may also be influenced by local or national considerations such as expected distribution of terminal data rates and compatibility with installed systems, for example. Consequently, the various solutions proposed so far are quite different from one another.

The purpose of this paper is to present a further solution. However, this one uses as a starting point what, it is considered, are advantageous features of the solutions already proposed. Since the characteristics of each solution are quite dissimilar, such a technique is not immediately obvious and it would be unreal to expect all the major characteristics of each to remain in it. Nevertheless, it is hoped that the technique to be described may be of assistance in reaching the goal of uniformity in different networks, at least in respect to common signalling procedures and interface specifications.

Description of the Problem

The Local Network System

In Figure 1, the user's access to a synchronous network is shown in its essentials. It is expected that each subscriber loop will consist of two unloaded wire pairs to provide a full-duplex service. The Administration will install its NTU at the user's location, and this typically detects and regenerates the incoming loop signals, delivers data, timing and supervisory information to the user's Data Terminal Equipment (DTE) and accepts data from the DTE. The use of regenerative repeaters in the loop is likely if the loop length exceeds a given distance. In one system this distance is anticipated to be 4 to 6 miles[6]. The repeater function is to regenerate the signals in each direction, to retime these signals and to retransmit them as essentially jitter free data streams.

At the local exchange, or office, the loop signal is translated into an appropriate form for multiplexing. Multiplexors are needed in order to reduce the cost of the network since data switching exchanges can be expected to be relatively thinly spread over a country so that there will be correspondingly long distances between them and the users. The multiplexors combine a number of low speed data channels, together perhaps with some multiplexing control information, into a 64 kbps stream. In this way a channel, derived from a 24 or 32 channel PCM system by a time slot extractor may be used.

Transmission Requirements

Because the transmission of a d.c. component in a digital signal is not practicable through a.c. coupled devices, such as regenerators, the disparity of signals over a relatively short time (i.e. the excess of digits of one polarity over those of another) should be limited if direct baseband transmission is to be allowed. In addition, in order to provide timing information, a reasonable density of transitions between levels should occur, which implies that there must be a restriction on the maximum length sequence of digits of the same polarity. With any scheme capable of transmitting unstructured data, this indicates that the original data should be translated into a more appropriate form.

Any translation of this kind means that redundancy is added. However, the greater the added redundancy, the greater the transmission costs due to local loop regenerators having to be more closely spaced. It should be a design goal, therefore, to keep the redundancy to the minimum commensurate with a simple translation process. If the translation process is purely binary, i.e. the bit rate on the subscriber loop exceeds the data rate of the DTE, then framing information must be incorporated in the loop signal. This framing information will enable the decoders to synchronize with the incoming signal in such a manner that the translation process can be properly in phase.

Finally, the loop signals should be such that they are easily convertible into a binary structure compatible with the multiplexing procedure described in which a PCM channel is assumed. If the subscriber loop signalling scheme does not take the local multiplexing into consideration, extra means would be required at the local exchange to overcome this incompatibility.

User Requirements

Several user requirements have already been mentioned, namely full duplex transmission facilities, the avoidance of restrictions on the data structure or format, and easy signalling procedures. To these should be added a simple interface, in both number of leads and functions, and a continuous bit timing waveform from the NTU to DTE.

Furthermore, there ought to be a method of distinguishing between signalling and data. The association of a status bit with successive groups of digits (i.e. the "envelope" concept [1]) and the use of special signals to indicate the start and the finish of the data mode both meet this requirement. Furthermore, the DTE should be able to recognize when a full-duplex connection has been set up in order to know when to start sending data. This overcomes the problem of bits at the beginning of data mode transmission being lost.

Next, a brief review of several proposals made already will be given, so that the different approaches to the overall problem can be appreciated.

Essential Features of Other Proposals

The techniques proposed can be classified into two major types: those using 3-level subscriber loop transmission without an increase in the bit rate, and those which retain 2-level transmission but at a higher bit rate. This second type can in turn be sub-divided into a group using envelopes and another using low-disparity codes.

3-Level Transmission

When 3-level transmission is used, as for example in the bipolar scheme proposed by AT&T[2], the bit rate on the subscriber loop equals the data rate. It automatically follows, therefore, that the timing waveform derived by the NTU from the received signal can be used directly by the DTE as if it were derived from a clocked modem. The penalty incurred in achieving this similarity is that an additional transmission level has to be distinguished in the signal received from the local loop. There is also the accompanying necessity for accurate equalization of the cable characteristics[7]. In order to accommodate unstructured data from the user, one of the many available alternatives for ensuring a reasonable transition density in all circumstances would have to be employed. Into this bipolar signal must be added framing information in order to be able to determine signalling characters at a receiver. If the user only needs, say, 64 signalling characters but uses 8 bits for each, then the two spare bits per character can be transformed into a special violation pattern. If these two bits are in the same position in each of the signalling characters, the detection of a bipolar violation of this kind not only indicates a signalling character, but provides signalling character synchronization also.

The DTEs are informed at the end of the signalling phase that the desired network connection has been made by the transmission to them both of a "cut-through" signalling character from some central point in the network. Data transmission may therefore take place, but there inevitably exists an indeterminate period between the receipt of this character and the first data from the remote DTE.

2-Level Transmission

i) Envelopes. Two major Administrations, namely the UK Post Office and the French PTT, have published papers [1,3] concerning networks based on the use of envelopes. In this concept data and signalling are divided into n-bit groups. Each of these groups of "information" have two extra bits associated with them. One is the status bit, which denotes whether the information is signalling or is data, and the other is the frame bit, which can be used to obtain envelope synchronization. These (n+2) bit envelopes are transmitted continuously. Due to the status bit in each envelope, there exists a very flexible way of changing between the transmission of signalling and data. Typical values for n are 6 and 8. The instant of the network connection can be easily detected by the receipt of particular signalling envelopes which can only be transmitted to the network by a subscriber.

Since data and signalling are inserted unaltered into the information bit positions of each envelope, it would be possible, especially in the cases of long envelopes and of frame bits not alternating in successive envelopes, for the transition density to fall below the minimum required to ensure accurate timing extraction from a stream of envelopes. For this reason, most proposals based on the envelope principle envisage diphase coding in addition in order to guarantee at least one binary transition per envelope bit occuring in the subscriber loop signal.

Should the NTU rather than the DTE form the envelopes for transmission to the network, and obtain envelope synchronization from the received signal, the envelope information bits may be passed between DTE and NTU in one of at least two ways. The first is in association with a continuous clock waveform at the data rate over the interface Timing lead. However, the NTU can be more simply constructed if the interface Timing lead carries a "burst-isochronous" clock at the bit rate of the envelope stream, and with two pulses missing in every n+2 positions. These two missing pulses correspond to the status and frame bit positions in the envelopes which will be on the Data leads. This interface technique is illustrated in Figure 2.

ii) Low-Disparity Coding. A binary line signal is not only less sensitive to equalization errors compared to a bipolar signal, but also more resistant to interference and simpler to regenerate. Unfortunately, it is often difficult to transmit the d.c. component normally present in a binary sequence. Binary translation techniques can overcome this problem. One code which translates a binary stream into a more suitable stream is described by Griffiths [8]. Using this code, binary blocks of length n bits are translated into code words of length n+1 bits in such a way that the difference between the number of 1's and the number of 0's in any sequence of code words never exceeds one.

Another similar scheme was developed by the Swiss PTT [7]. In this, the d.c. component is not totally eliminated, but kept to a sufficiently low value. Distinction, however, is made between signalling and data to enable switched network use. Signalling characters are so translated that the resulting code words contain a unique bit pattern that cannot be imitated in the code words resulting from the translation of data. Code word alignment is achieved from uniquely occuring bit sequences in the subscriber loop signal.

The Binary Translation Scheme

Basis

The bipolar, envelope, and low-disparity coding techniques already proposed each have advantages and disadvantages. Three criteria which have been chosen to indicate differences in their performance are:

1. Redundancy imposition
2. Signalling limitation
3. Indication of mode change instant

Judged from the redundancy aspect, the low-disparity translation approach is clearly preferable. Nevertheless, the Swiss proposal limits the available 8 bit signalling characters to 33 when the translation is from 8 to 10 bits, and would limit them to 17 if the translation were from 6 to 8 bits, which is clearly unacceptable. The bipolar and envelope techniques do not have this limitation but do impose much greater redundancy with 3-level and diphase transmission respectively. In addition, the bipolar scheme can not provide a clear indication of the start of the data mode as a network "cut through" character is used, whereas the other two schemes can.

The above characteristics of the proposed schemes enable us to state desirable features for a new scheme. The first is that a low-disparity translation should be adopted in order to have low redundancy while at the same time enabling direct transmission and simple regeneration over subscriber loops. Nevertheless, a complex translation matrix ought not to be needed. Secondly, there should be an acceptable signalling character set available for communication between the DTE and network. Thirdly, the DTEs should be given a clear indication of the start of the different transmission modes quite separately from that of the network path "cut through" instant. A binary translation with these features is next described, after first making certain assumptions.

Assumptions

The term "information" will be used from now on for both signalling and data. The translation process causes m-bit information blocks to be converted into low-disparity code words of length m+n bits. However, before one can produce a translation suitable for switched data network operation, certain assumptions must be made in order to choose values for m and n. In addition, the length of the signalling characters must be decided and also, the signal to exist on the subscriber loop in the no-call state.

i) Choice of m and n. The choice of values for m and n is directly influenced by the wish to be compatible with the AT&T local multiplexing principle[9]. This envisages 48 kbps of a 64 kbps channel being used for information and the remaining capacity being for control purposes. Six bits of each byte in the channel would be reserved for information. If, for example, the information rates are each 2.4 kbps, every twentieth byte would correspond to a given sub-channel. The sub-channel bit rate in the multiplex is therefore one third greater than the information rate. In order to avoid having to derive timing waveforms at other rates just for subscriber loop transmission, there can be a straightforward compatibility with the multiplexing process if the ratio m:n is 3:4. In practice, this means choosing between a 3 to 4 and a 6 to 8 bit translation process; the former is selected since a much smaller translation table is required.

ii) Signalling Character Length. 64 different characters should give sufficient signalling flexibility. This is considered to be the case by AT&T, although only 33 exist in the Swiss scheme. Accepting that a 6 bit character suffices, signalling characters can be inserted directly into the 6-bit information field of a sub-channel byte in the multiplex. This simplifies the translation process, since 6-bit signalling characters can correspond to two 4-bit code words.

iii) No-Call State. One further remark concerns the subscriber loop signal in the no-call state. The DTE must always receive a continuous timing waveform from the network so that it can initiate a call at any time. It should not be necessary, though, to translate "idle" characters when no call is in progress. A continuous binary alternating signal is assumed to be better in this phase, and it can be simply generated.

Implicit here is that at the start of a call synchronization must be rapidly acquired so that the initial characters of the signalling phase are properly decoded.

Determination of the Translation

There are 16 possible 4-bit code words. Of these, 0000 and 1111 must be excluded as they contain no transitions and a series of either of these code words would be purely d.c. The remaining 14 code words are classified into 3 types: N (Negative), Z (Zero) and P (Positive) according to whether they contain one, two or three binary ones as follows:

N-Type	Z-Type	P-Type
0001	0011	0111
0010	0101	1011
0100	0110	1101
1000	1001	1110
	1010	
	1100	

Code words have been selected to correspond with the eight 3-bit information blocks in the following manner:

Information Block	Code Word	Type
001	0011	Z
010	0101	Z
011	0110	Z
100	1001	Z
101	1010	Z
110	1100	Z
111	1011 or 0100	P or N
000	1101 or 0010	P or N

When transmitting, the appropriate choice of code word for the information blocks 111 and 000 is a P-type if the previous $\bar{Z}$-type transmitted was an N-type and an N-type if the previous $\bar{Z}$-type was a P-type code word. This is the basic scheme as proposed by Griffiths, and ensures that neither positive nor negative disparity accumulates. This paper does not present a description of how this conversion takes place. However, there are several straightforward approaches, and in no case is a conversion matrix required.

Because the N-type and P-type code words chosen to correspond with the 3-bit information blocks do not contain three consecutive binary digits of the same polarity, the longest sequence of consecutive digits of the same polarity which can arise is of length 4 and occurs in the following situations:

Information Blocks	Code Words
001 110	0011 1100
110 001	1100 0011
001 000	0011 1101 (*)
110 000	1100 0010 (*)
111 110	1011 1100 (*)
111 001	0100 0011 (*)

Since neither adjacent N-type nor adjacent P-type code words are allowed, the information block pairs 111 000 and 000 111 are never translated into the code word pairs 1011 1101 and 0100 0010 respectively and are hence not included in the above list. An asterisk above denotes the fact that this code word pair is only one of the two possible translations, both of which have an equal probability of occuring with random information.

The code word pairs containing the sequences of four consecutive binary digits are of the form X011 11XX and X100 00XX. This observation is made use of for determining code word synchronization, as the fourth consecutive binary digit of the same polarity always occurs in the second slot of a code word.

Translations, therefore, of 6 of the 64 different information block pairs include a 1111 or 0000 code word synchronization sequence. Allowing for the fact that four of these pairs are only so translated in half of the occasions when they occur, there will be a synchronization sequence, on average, once every 32 code words.

Translation Violations

Violations to the normal translation rules are used whenever it is wished to indicate a change of transmission into the signalling and into the data modes. Special signalling characters are reserved for these instances. Let us initially consider the obvious case of the start of the signalling mode following the no-call state. As soon as the Call Request (to network) or the Incoming Call (from Network) signal is transmitted, not only code word synchronization but also signalling character synchronization have to be immediately derived. The signalling characters from then on can then be properly decoded. The violation principle is used in this instance.

In normal operation, information is never translated into adjacent N-type or P-type code words. NN- or PP-type code word pairs are a violation of the normal rules and can provide signalling character synchronization. It will be recalled that code word synchronization is obtained from code word pairs of the forms X100 00XX and X011 11XX in which the fourth consecutive binary digit of one kind indicates the second slot in a code word. Now, if these code word pairs are also NN-type or PP-type, one can additionally determine that a code word is the translation of the first or of the second half of a signalling character.

This can be explained more clearly in the description of such signalling characters, which are denoted SYN-type. There exist two NN-type and two PP-type code word pairs which both violate the normal translation rules and contain a code word synchronization sequence. These are:

Name	Code Word Pair	Type
SYN 1	0100 0001	NN
SYN 2	0100 0010	NN
SYN 3	1011 1110	PP
SYN 4	1011 1101	PP

Such violations can be easily produced. If the NTU translates between code words and information, it should be designed to recognize any of the four SYN-type characters from the DTE and any of the four code word violations pairs from the subscriber loop.

Whenever a 0000 or 1111 synchronization sequence is received in the bit stream from a subscriber loop, code word boundaries can be identified. If the code word pair containing such a sequence is also found to be a violation pair, then it is determined that this is the translation of a signalling character. Now, as one can suppose signalling characters to be transmitted contiguously, the boundaries of the subsequent signalling characters can be correctly determined. This process is assisted by the frequently occuring code word synchronization sequences within the translated bit stream. In this way it is possible to check that the code words are being decoded in phase, although there is one case in which the loss of signalling character synchronization would not be detected. This is when the code word synchronization is correct but where the code words corresponding to the first half of signalling characters are assumed to correspond to the second half, and vice versa. Such a condition would only be caused by a gross error burst, which one can expect be correspondingly rare.

As there are both PP-type and NN-type code word pairs available, it is convenient to use one type, say NN, at the start of the signalling mode and the other, PP, prior to the start of the data mode. Thus, whenever an NN-type code word pair is received, an indication is given that transmission is now in the data mode. Similarly, whenever a PP-type pair is received, one is prepared for the fact that the first $\bar{P}$-type code word received thereafter belongs to the data mode, because the normal translation process will have been resumed. It is now appropriate to give an example of how this scheme might be used in practice with a switched data network.

Example of the Scheme's Usage

Initial Considerations

With any switched data network, only some of the terminals will be calling at any one time. For this reason it has been presumed that there will be some form of scanning mechanism which will search for a change from the no-call state at the network end of each subscriber loop. The network then only requires sufficient capability to service the expected maximum number of simultaneously active subscriber loops. There will consequently be a delay between the DTE sending the Call Request signal and receiving the Proceed to Select response; this delay is dependent on the current activity of the network, the subscriber loop scanning rate, the processing time of the Call Request signal, etc. Because of this indeterminate delay, it is assumed that the DTE repeatedly transmits the Call Request signalling character until the Proceed to Select is received.

A called DTE is notified of an incoming call by the Calling Indicator. So that the procedures for calling the network and calling the terminal should be consistent, the Calling Indicator character is also transmitted repeatedly until the network receives a response from the terminal.

There inevitably exists an "idle" part of the signalling mode between the end of the call selection phase and the start of the data mode. In order not to interrupt the translation process, signalling characters with an "idle" function are transmitted in this period. These are denoted IDLE-A and IDLE-B, and are transmitted by the DTE and network respectively. The use of different IDLE characters in each direction can simplify the recognition of the instant when the path through the network has been completed as will be shown shortly.

Application of SYN-Characters

The SYN-type characters indicate that transmission is in the signalling mode and enable signalling character synchronization as well as code word synchronization to be acquired. These features though, result from the regular translation rules being violated and do not preclude the SYN-type characters from having normal signalling functions. Indeed, there can be benefit in representing the Call Request and Calling Indicator signals at the beginning of a call by a SYN-type character, and the Disconnect and Disconnect Acknowledge signals at the end of a call by another SYN-type character. The advantage lies in the fact that those signalling characters which start a new transmission mode are automatically translated into code word pairs which provide the information needed for recognizing the new transmission mode and for synchronization to enable correct decoding.

The SYN-type characters are used in this example as follows:

Characters	Function	Direction
SYN 1 (NN)	Call Request Calling Indicator	To Network From Network
SYN 2 (NN)	Disconnect Request Disconnect Acknowledge	To or From Network To or From Network
SYN 3 (PP)	Data Mode Following	To Network
SYN 4 (PP)	Spare	

Although the signalling characters representing Call Request, Calling Indicator, Disconnect and Disconnect Acknowledge are repeated several times while waiting for a response, the accumulation of disparity in these short periods ought to be acceptable.

Signalling Sequence

Only the simplest sequence of events will be described, in which a DTE calls, then is connected to the called DTE, transmits data and disconnects. However, the flexibility of the scheme allows more complicated signalling procedures to be carried out if so wished. Among these, for example, would be network recall, terminal identification, etc. The procedure to be outlined is thought to be sufficient to clearly illustrate how the scheme could be used. This is depicted in Figures 3 and 4.

i) Call Selection. The calling DTE sends SYN 1 (Call Request) repeatedly until the Proceed to Select is received. After receiving the Proceed to Select, which is preceded by one SYN 1 to ensure that the signalling mode is recognized and that synchronization is obtained, the DTE sends the dial characters followed by repeated IDLE-A characters. When all the dial characters have been sent the network replies with repeated IDLE-B characters.

After the dial digits have been processed by the network, it will send SYN 1 (Calling Indicator) repeatedly to the called terminal which, if it can accept the call, responds with a SYN 1 followed by repeated IDLE-A characters. The network recognizes this acceptance and responds with repeated IDLE-B characters.

ii) Call Connection. All that now remains for the network to do is to effect the full-duplex path between the called and calling subscribers. As soon as the outward path is connected, the called DTE will receive IDLE-A characters, which will have come from the calling DTE through the network, instead of IDLE-B characters. The called DTE then knows that the outward path has been connected and replies with Connected characters. After the return path from the called to the calling DTE has been completed, these Connected characters will be received by the calling DTE. This can now start data mode transmission.

iii) Entry into the Data Mode. The data mode is preceded by the transmission of one SYN 3 signalling character, which is translated into a PP-type code word pair. The first code word of the data mode must, therefore, be a $\bar{P}$-type as there are no violations in the data mode. When the $PP\bar{P}$ sequence is received, it is recognized that the data mode has begun. The called DTE can then send a SYN 3 character and data also.

iv) Disconnect. The sequence of events for the disconnect procedure is shown in Figure 4. It is assumed that either DTE can initiate the disconnect so that reference is made to initiating and responding DTEs rather than to calling and called ones. The disconnecting DTE transmits SYN 2 characters to request call disconnection. Each SYN 2 is translated into an NN-type code word pair which indicates that transmission is in the signalling mode and which gives synchronization information to enable correct decoding of signalling characters. These NN-type code word pairs will be decoded back into SYN 2 characters both by the network and by the responding DTE since the path through the network has not yet been broken. The responding DTE replies with the Disconnect Acknowledge signal by sending SYN 2 characters also.

The situation is now reached in which each DTE is sending SYN 2 characters. The network acts upon these SYN 2 characters and breaks the full-duplex path between the DTEs. It then puts the no-call waveform onto the subscriber loops. This waveform is the continuous binary sequence 010101... and as soon as a DTE receives this after SYN 2 characters it knows that the network has disconnected the call. The DTE then similarly sends the continuous 010101... sequence to represent the no-call state.

In Figure 4, the network is shown as sending a couple of SYN 2 characters to the DTEs between the call disconnection and the 010101... sequence. This is purely as a safeguard to make sure that the DTEs never miss SYN 2s before the 010101... waveform and to guarantee it being recognized as the no-call waveform.

NTU and Interface

Three NTU alternatives are shown in Figure 5. In its simplest form, the NTU would only be responsible for protecting the network from any highly abnormal electrical signals which might come from the DTE in extreme circumstances. An NTU which carries out the timing extraction process and regenerates the incoming waveform from the subscriber loop is more suitable for many applications. With this kind of NTU the user would most probably be made aware of any fault condition in the subscriber loop as the interface timing waveform would be interrupted. This is a valuable diagnostic aid. In this case the NTU is still simple, and the interface only comprises three leads plus common return.

The third NTU alternative contains logic. This is for synchronization, translation and decoding, and transmission mode control. The division of functions between NTU and DTE in this way means that logic which could be more simply added to the DTE logic is separated from the DTE and contained within the NTU, with the consequent increase in cost and in maintenance problems. In addition, the NTU and DTEare connected by an interface which now includes at least one control lead in each direction, with the corresponding greater expense of implementation.

Conclusion

This paper has attempted to summarize briefly the subscriber loop transmission and signalling problems which arise in the planning of synchronous data networks. It has shown how several proposed techniques approach these problems, and the deficiencies associated with these techniques.

The objective of designing a translation scheme based on low-disparity coding and violation principles has been to overcome the limitations of other techniques. With this scheme the binary transmission rate on the subscriber loop is only one third above the data and signalling rates, the d.c. component is eliminated and a high density of transitions is achieved in order to allow the use of regenerative repeaters. A set of 64 signalling characters is available for communication between the network and user's terminal, and the user may transmit unrestricted data in the full-duplex mode. This translation may be simply achieved, and in no case is a complicated translation matrix needed.

The translation scheme has been developed as a result of considerations of the local area plant from the data switching exchange to the terminal, including the terminal interfacing. No attempt has been made to discuss network control and switching aspects although a number of assumptions have been necessarily made in this domain. It is, however, thought that this scheme represents an advance in solving the subscriber loop signalling and transmission problems, and one which may appeal to teleprocessing systems designers and Telecommunications Administrations alike.

References

1) Allery G.D. and Chapman K.J., "Features of a Synchronous Data Network for the United Kingdom", International Conference on Communications, Montreal, June 1971 Conference Record p. 31-10.

2) USA Contribution GM/NRD No. 66 to CCITT "Interface between Data Terminal Equipment and Data Communications Equipment", October 1971.

3) Jousset A., "Etudes de Reseaux de Commutation de Données en France", ICC, Montreal June 1971, Conference Record p. 31-17.

4) Gabler H.G., "Data Network Planning in the Federal Republic of Germany", ICC, Montreal June 1971, Conference Record p. 31-14.

5) Hartley G.C., "Opportunities and Problems of Synchronous Networks", ICC, Montreal June 1971, Conference Record p. 31-5.

6) Atkinson D.M. and Strahlendorf U.C., "The Probable Future of Canadian Long Haul Digital Data Network Connections with the USA", ACM/IEEE Second Symposium on Problems in the Optimization of Data Communications Systems, October 20-22 1971, Palo Alto, California.

7) Swiss PTT Contribution GM/NRD No. 64 to CCITT "Envelope Structure and Digital Subscriber Line", October 1971.

8) Griffiths J.M., "Binary Code Suitable for Line Transmission", Electronics Letters 5 No. 4, 20th Feb. 1969, pp. 79-81.

9) AT&T Contribution GM/NRD No. 75 to CCITT "Intermediate Level Multiplexor Speed and Format", October 1971.

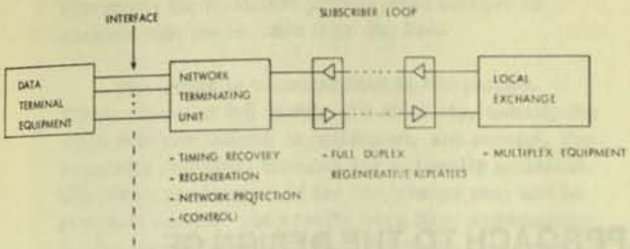


FIGURE 1: THE LOCAL NETWORK SYSTEM

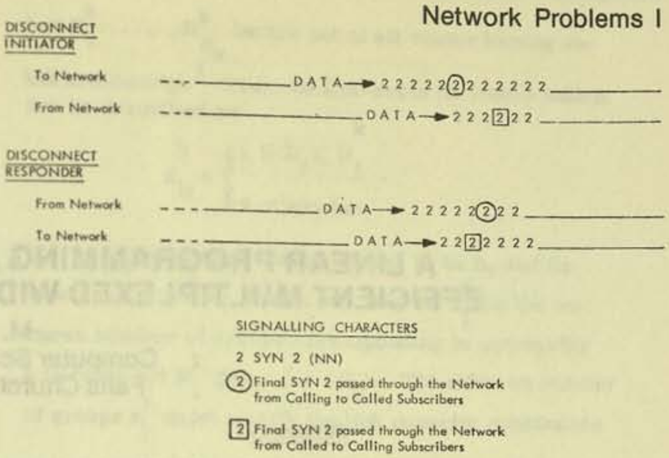


FIGURE 4: CALL DISCONNECTION SEQUENCE

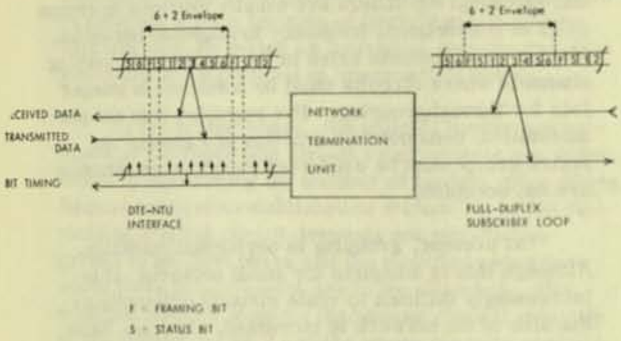
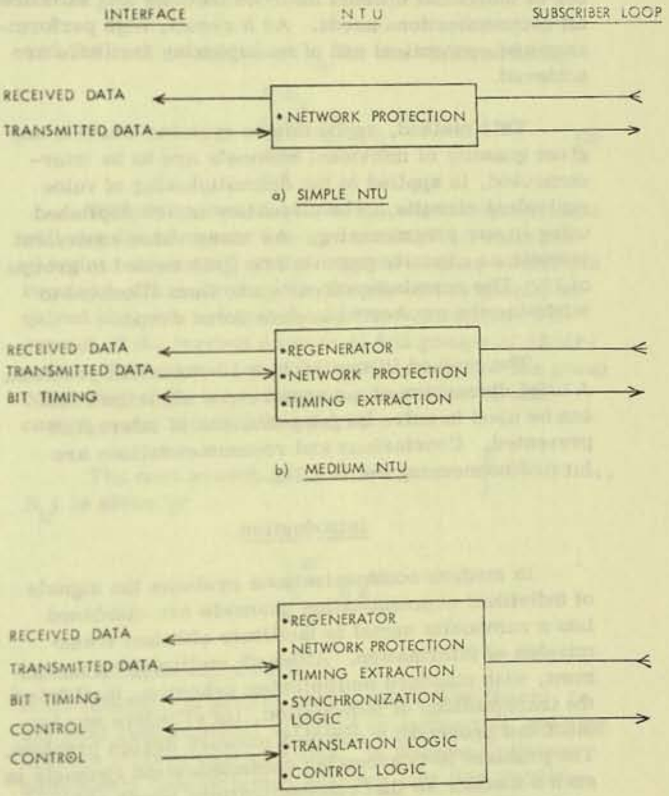


FIGURE 2: THE BURST-ISOCHRONOUS INTERFACE AS USED WITH ENVELOPE SIGNALLING



a) SIMPLE NTU

b) MEDIUM NTU

c) COMPLEX NTU

FIGURE 5: 3 NTU CONFIGURATIONS

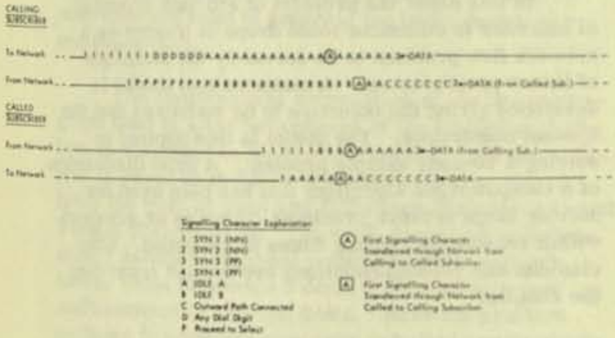


FIGURE 3: EXAMPLE OF A CALL CONNECTION PROCEDURE

A LINEAR PROGRAMMING APPROACH TO THE DESIGN OF EFFICIENT MULTIPLEXED WIDEBAND TRANSMISSION SYSTEMS

M. Millman
Computer Sciences Corporation
Falls Church, Virginia, U.S.A.

Summary

A practical way has been developed to allocate the multiplex facilities of large scale wideband communications systems in a way that minimizes the number of individual channel interconnections and satisfies all communications needs. As a result, high performance and economical use of multiplexing facilities are achieved.

This method, applicable to systems wherein any given quantity of individual channels are to be interconnected, is applied to the demultiplexing of voice equivalent circuits. The allocation is accomplished using linear programming. As many voice equivalent circuits as capacity permits are first routed in groups of 12. The remaining circuits are then allocated to minimize the number of tandem voice drops.

The method is applied to a 10-node test problem. A brief discussion of a computerized algorithm that can be used to solve larger problems of interest is presented. Conclusions and recommendations are formed concerning the results.

Introduction

In modern communications systems the signals of individual communication channels are combined into a composite signal to facilitate efficient transmission of information. Although multiplexing equipment, with standard multiplexing schemes, facilitates the transmission of information, its effective employment has presented a complex network design problem. The problem is to group the communication channels in such a manner so that communications needs are met with a minimum number of individual channel interconnections (e.g., tandem voice drops). The problem is further intensified in transmission systems which are designed with a large number of radio links and multiplexing equipment to facilitate a wide choice of circuit routing.

Tandem voice drops can have a significant effect on signal quality. Demodulation of a signal from carrier group level to voice level and then remodulation back to group level to route that signal through an intermediate node degrades the signal-to-noise ratio. Consequently, it is desirable to minimize the number of tandem voice drops used to satisfy known circuit requirements. In planning a practical network, or improving an existing one, elimination of all tandem voice drops is usually precluded by limitations placed on available facilities. The problem is, therefore, that of grouping channels to obtain the least number of tandem voice drops while utilizing the most desirable transmission paths. Furthermore, since channels and throughgroup filters are usually provided in groups of 12 in conventional frequency multiplexed networks, integer requirements arise in the problem of grouping channels where circuits must be combined in integer lots for throughgrouping. For example, it is not permissible to demodulate a fraction of a group. An entire group must be used even though some channels are not occupied.

At present, grouping is performed manually. Although this is adequate for small networks, it is increasingly difficult to route circuits efficiently as the size of the network is increased. Because there are many routes that circuits can take in large networks, automatic methods are desirable. Linear programming, a technique for allocating resources (routing circuits and grouping channels) when a large number of choices are available, has been applied to this problem. The function to be optimized, and all constraints, must be expressible as linear equalities or inequalities.

In this paper the problem of efficient grouping of channels to minimize voice drops is treated as a network flow problem to be solved by the technique of linear programming. A mathematical model is developed giving the objective to be optimized and the system constraints. The model is then applied to solving a 10-node sample problem. A brief discussion of a computerized algorithm that has been used for solving large network problems (70 nodes or greater) within reasonable running times is presented. Conclusions and recommendations are formed regarding the results.

Technical Approach

General Considerations

The application of linear programming techniques to the network problem discussed in the Introduction requires expressing the restrictions of the problem as a series of linear equalities or inequalities. The restrictions on the network being considered are the capacity limitations of the interconnecting transmission facilities (branches or links) and the communication requirements (circuits) that are desired between each pair of nodes. The nodes are identified when the transmission facilities are interconnected and multiplexing and demultiplexing may take place. The capacity is expressed in voice equivalent channels or groups and represents the maximum permissible number of channels that can be used over the link.

The design is accomplished in two phases. First, the use of full groups (12 channels, usually for voice communications) is optimized, and second, the remaining individual circuits are optimally satisfied. Whereas a combination of the two phases may not be precisely optimal, the results have been satisfactory for the cases studied. In the first phase, the required circuits are grouped in multiples of 12 for each user pair and the appropriate number of spare circuits added to the last group to achieve a full 12-channel group. These groups are then weighted according to the number of required circuits. Next, a linear program is formulated to provide efficient routing of the maximum number of satisfied, weighted, and grouped requirements subject to the constraints of group capacity. Those requirements that cannot be satisfied in complete groups because of limitations on group capacity must be routed as individual circuits and demodulated and modulated at intermediate points.

In the second phase, the optimal routing assignments of the remaining circuit requirements are determined by minimizing the number of tandem voice drops. Requirements are established to include all of the remaining original circuit demands not satisfied as groups in the first phase and the modified network to which these requirements are to be allocated. The modified network includes the unused channel capacity of the original network and the spare capacity carried by the satisfied rounded groups. A second linear program is then used to determine the maximum number of remaining satisfied circuit requirements, subject to capacity constraints, while utilizing the least number of channels in the allocation process. This is tantamount to routing the remaining circuits so that the resultant voice frequency interconnections are minimized. The routing information obtained from the solution outputs of both linear programs can then be combined and used to generate the basic design of a multiplex plan.

The mathematical formulation of the linear program used to solve both phases of the multiplex problem is discussed in the next paragraph. The formulation is based on methods of network flow developed from the work of Ford and Fulkerson (Reference 1) and Tomlin (Reference 3). The work described in Reference 1 refers to maximal multicommodity network flows, while Reference 3 deals with minimum cost multicommodity network flows. Since the problem pertains to achieving maximum satisfied requirements (flows) while obtaining minimum utilization (cost), both formulations are combined into one general model.

Mathematical Model

Assume an enumeration of links $L_1, \ldots\ldots, L_m$. Let $R^k_1, \ldots\ldots, R^k_{N_k}$ be the set of all routes joining the kth commodity.[1] Then the link-route incidence matrix can be described as

$$a^k_{ij} = \begin{cases} 1 \text{ if } L_i \in R^k_j \\ 0 \text{ otherwise} \end{cases}$$

Let the group capacity of link L_i be b_i and its associated cost[2] c_i $(i=1, \ldots m)$, and let x^k_j be the unknown number of groups corresponding to commodity k in route set R^k_j $(j=1, \ldots\ldots, N_k)$. The unknown number of groups x^k_j must satisfy the link capacity constraints

$$\sum_{k=1}^{k=q} \sum_{j=1}^{j=N_k} a^k_{ij} x^k_j \le b_i \quad (i=1, \ldots\ldots, m) \qquad (1)$$

and the user requirement r_k for the kth commodity

$$\sum_{j=1}^{j=N_k} x^k_j \le r_k \qquad (2)$$

With respect to Equation (2), user requirements will consist of one or two commodities for each pair of nodes, depending on the original number of circuits required. If more than 12 circuits are originally required between two nodes, the first commodity will consist of the largest number of full groups of 12 circuits. The second commodity will comprise one group consisting of the remaining required circuits and enough spare circuits to total 12.

The cost associated with any route R^k_j $(j=1, \ldots\ldots, N_k)$ is given by

$$\sum_{i=1}^{i=m} c_i a^k_{ij}$$

---

[1] A commodity, in generalized network flow theory, is any flow requirement, directed or undirected, existing between a given pair of nodes. In frequency division multiplexed transmission systems, the flow requirement is usually expressed in circuits or groups of circuits. A commodity may consist of the total flow requirement or any part of it. Thus, the total requirement can be made up of more than a single commodity.

[2] Cost can be expressed in dollars, channel miles, number of relay stations needed to maintain proper signal levels, or any number that implies a relative desirability of using one resource over another.

and hence the total cost for all x_j^k $(j=1,\ldots\ldots,N_k;$ $k=1,\ldots\ldots,q)$ is

$$\sum_{i=1}^{i=m}\sum_{k=1}^{k=q}\sum_{j=1}^{j=N_k} c_i a_{ij}^k x_j^k$$

Let p_k be the revenue or weighting factor[1] associated with satisfying requirement r_k. Thus, for each requirement r_k, the revenue is

$$\sum_{j=1}^{j=N_k} p_k x_j^k$$

The objective to be maximized is profit, which is the difference between total revenue and total cost. Another way of expressing the objective is to satisfy the greatest number of weighted grouped requirements in the most efficient way. Thus, the function to be maximized is

$$Z = \sum_{k=1}^{k=q}\sum_{j=1}^{j=N_k} p_k x_j^k - \sum_{k=1}^{k=q}\sum_{j=1}^{j=N_k}\sum_{i=1}^{i=m} c_i a_{ij}^k x_j^k \qquad (3)$$

The objective function, Equation (3), with the constraints (1) and (2), and the implicit restriction of each $x_k^j \geq 0$, form a linear program that solves the initial problem of efficiently throughgrouping the largest number of circuit requirements. The remaining circuits that are not throughgrouped because of capacity limitations must now be satisfied with use of the remaining network capacity. These circuits constitute a new set of requirements while the remaining capacity is composed of a residual network of two parts. The first part consists of the unused link capacities of the original network. The second part is spare capacity carried by the throughgroups used in satisfying the grouped requirements. These spare circuits, which always number less than 12 for each satisfied grouped requirement, generate new "links" in the network since they now provide a direct connection between the various user endpoints.

The residual requirements, together with the modified network, constitute a second allocation problem to be solved by linear programming. The objective is to satisfy the largest number of residual requirements within the limits of capacity by using the least number of channels. This solves the identical problem of allocating the remaining circuit requirements to achieve the fewest number of tandem voice drops. The solution to this problem can be obtained by using the same linear programming formulation developed for the throughgrouping problem. The basic differences are in the units involved; circuits and channels are used rather than groups. In addition, each residual requirement is represented by just one commodity since the need for expressing fully utilizes and partially utilized groups as separate commodities no longer exists.

The values chosen for the p_k merit serious consideration, especially for the case involving the maximization of grouped requirements. If all p_k are assigned equal values, a requirement having one or two circuits is as likely to be throughgrouped as one having 11 or 12 circuits. As a consequence, tandem voice drops used in satisfying residual circuit requirements will be far greater than if the more completely fill groups were satisfied first. To avoid this, a weighting factor for each grouped requirement is assigned in direct proportion to the number of required circuits contained in each group. A weighting factor of 100 times the number of circuits is suggested. These values are made large intentionally with respect to link costs so that the "revenue" received from satisfying a given requirement is always greater than any "cost" associated with the links that make up the routing for it. If the total cost to satisfy a grouped requirement were larger than the assigned revenue simply because the weighting factor was not large enough, the requirement would not be satisfied despite the existence of sufficient capacity.

The preceding linear program contains m + q rows (excluding the objective function) and a large number of columns, each corresponding to a generated route of the link-route incidence matrix. Since the number of permissible routes, and hence the number of variables, is usually large in practical networks, it might seem computationally infeasible to solve anything but small problems. Fortunately, there is no need to use the entire link-route incidence matrix because the special structure of this program can be utilized in algorithms for its solution. One such algorithm is the "shortest chain" which is used in conjunction with the simplex method to bring new vectors into the current basis. This can be done without explicit knowledge of the nonbasic column vectors of the link-route incidence matrix. Although a description of the mechanism by which this is accomplished is beyond the scope of this paper, the mathematics describing this technique can be found in the literature (see References 1, 2, and 3). Of major importance however, is that using the shortest chain algorithm to solve the multicommodity network flow problem proposed here enables the solution of large problems within reasonable limits of computation.

Computerized linear programs using the simplex computation for link-route formulation of the multicommodity network flow problem have been developed

---

[1] A more detailed discussion of weighting factors as they apply to the multiplexing problem is presented later in this section.

by several investigators. For example, some computational experience has been acquired recently with the use of TRAN10 developed by Dr. Mandell Bellmore of Johns Hopkins University. TRAN10 is a computerized multicommodity transshipment algorithm which uses the mathematical theory developed in References 1 and 3. The program was provided to the Defense Communications Agency and Computer Sciences Corporation for use in solving large scale network optimization problems of the Defense Communications Systems (DCS).

The actual size of the network that can be solved using TRAN10 is a function of the combined number of links and commodities for a given machine capacity. Networks having as many as 500 links and commodities have been solved on the UNIVAC 1108 in less than 2 minutes. The computer running time varies approximately as the cube of the number of links and commodities. For example, if a problem with 500 links and commodities takes 7 minutes to run, a similar problem with three times as many links and commodities might require several hours. TRAN10 is written in FORTRAN IV and, therefore, is adaptable to most scientific computers.

Although the output of the TRAN10 program does not guarantee that solution values are integer, experience has shown that when the coefficients of the matrix are 1, -1, or 0, as in the multicommodity network flow problem, the solution is nearly always in terms of integers. Moreover, whenever fractions occur in the solution, the number of variables having fractional values is relatively small compared to the total number of structural variables involved. When this is the case, the rounding of the few fractional answers should not affect the desired results adversely provided no constraints are violated and the deviation from the original objective value is not too great. The latter is usually no problem since, in practical networks, there exist many links having large capacities.

In the next section, a sample 10-node network problem is formulated in terms of the mathematical model developed in this section and solved using TRAN10. Two LP computer programs were run on the Philco 2000/212, each requiring about 1 minute of running time. The first run solved the grouping problem and the second run solved the minimum number of tandem voice drops. The results were combined to yield a basic multiplex plan. A summary of utilized system capacity was subsequently obtained from this plan.

In addition to the 10-node problem, a 27-node test problem was successfully run. The initial network consisted of 55 links and 205 commodities and required about 20 minutes on the Philco 2000 for each phase. Further, at the time of this writing, a major study has begun in the generation of efficient multiplex plans for the DCS within the entire European area. The network is expected to consist of about 200 nodes, 400 links, and 1000 commodities, which will be the largest practical problem attempted with this method.

Illustrative Example

Consider the transmission network whose connectivity is shown in Figure 1. The network consists of 16 transmission links and 10 nodes that serve as connecting points for the transmission facilities. The nodes are used to originate, terminate, or serve as via points for circuit requirements between any user pair. The upper limit on link capacity and the associated link costs[1] are given in Table 1.

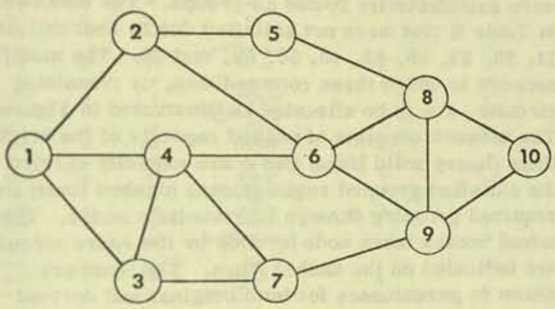


Figure 1. Network Connectivity of the 10-Node Illustrative Example

LINK NO. (i=)	FROM NODE	TO NODE	CHANNEL CAPACITY	GROUP CAPACITY (b_i)	LINK COST (c_i)
1	1	2	120	10	1
2	1	3	60	5	1
3	1	4	120	10	1
4	2	5	60	5	1
5	2	6	120	10	1
6	3	4	24	2	1
7	3	7	60	5	1
8	4	6	60	5	1
9	4	7	120	10	1
10	5	8	60	5	1
11	6	8	60	5	1
12	6	9	120	10	1
13	7	9	120	10	1
14	8	9	60	5	1
15	8	10	60	5	1
16	9	10	60	5	1

Table I. Upper Limit on Link Capacity and Associated Link Cost for the 10-Node Illustrative Example

---

[1] Different values of link cost to reflect desirability measures are immaterial to the demonstration of the technique used in this example. For the purpose of simplicity, however, all values are assumed the same throughout.

The user requirements between stations are furnished in Table II and are expressed as both circuits and groups. Also included are the weighting factors (pricing) associated with each grouped requirement. Each requirement is identified by at least one commodity. When two commodities are needed to describe the complete user requirement, the first commodity always contains some integer multiple of 12 circuits.

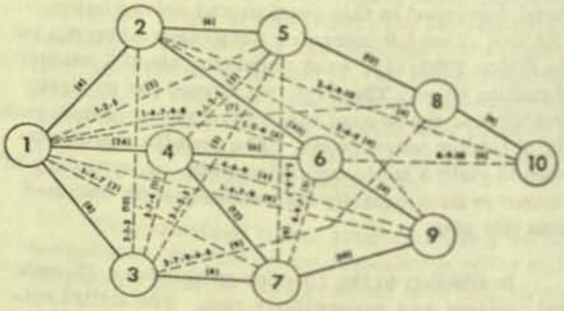
Sufficient information is available to ascertain the largest number of grouped requirements that can be routed through the network by use of the linear program (LP) developed in the previous section. The results of the program show that all but 10 commodities were satisfactorily routed as groups. The commodities in Table II that were not satisfied due to shortfall are 14, 35, 38, 39, 43, 46, 50, 53, and 58. The modified network to which these commodities, or remaining circuits, are to be allocated is illustrated in Figure 2. The network consists of unused capacity of the original links (heavy solid lines) and spare capacity carried by the satisfied grouped requirements (dashed lines) that required grouping through intermediate nodes. The actual routes taken node by node by the spare circuits are indicated on the dashed lines. The numbers shown in parentheses for both original and derived links indicate the available capacity left after the first LP run.

Using the remaining unsatisfied circuits as new input requirements with the connectivity shown in Figure 2, a second LP run was made. The objective was to allocate the remaining circuits in such a way that the least number of remaining channels was used (minimum tandem voice drops). The same formulation was used except that units were expressed in circuits and channels instead of groups. In addition, all commodities or remaining circuits were assumed to have equal weight (all $p_R = 10$) and all link costs were assumed equal (all $c_i = 1$). These last two assumtions, however, were made solely for simplicity and generally cannot be made for all problems.

The results of the second LP run indicated that all remaining requirements were satisfied. Since all of the derived links and most of the original links of the modified network had capacities less than 12 channels, voice drops were required. The routes taken by the satisfied circuits were originally expressed in terms of the physical links of the network shown in Figure 1. Where derived links were used in these routes, they were expressed in terms of the original routes indicated on the dashed lines. Thus, in terms of the modified connectivity in Figure 2, voice drops occurred at all intermediate nodes included in the routes.

COMMODITY NO. (k=)	SOURCE NODE	DESTINATION NODE	CIRCUIT DEMAND	GROUP DEMAND (r_k)	WEIGHTING FACTOR (p_k)
1	1	2	12	1	120
2	1	2	4	1	40
3	1	3	12	1	120
4	1	3	4	1	40
5	1	4	12	1	120
6	1	4	18	1	180
7	1	5	7	1	70
8	1	6	12	1	120
9	1	6	4	1	40
10	1	7	10	1	100
11	1	8	9	1	90
12	1	9	12	1	120
13	1	9	2	1	20
14	1	10	3	1	30
15	2	3	12	1	120
16	2	3	2	1	20
17	2	4	12	1	120
18	2	5	12	1	120
19	2	5	6	1	60
20	2	6	24	2	120
21	2	8	8	1	80
22	2	9	9	1	90
23	2	10	9	1	90
24	3	4	24	2	120
25	3	4	1	1	10
26	3	5	7	1	70
27	3	7	24	2	120
28	3	7	4	1	40
29	3	8	2	1	20
30	4	5	10	1	100
31	4	6	24	2	120
32	4	6	4	1	40
33	4	7	36	3	120
34	4	8	12	1	120
35	4	8	3	1	30
36	4	9	12	1	120
37	4	9	6	1	60
38	4	10	6	1	60
39	5	6	2	1	20
40	5	7	8	1	80
41	5	8	24	2	120
42	5	10	12	1	120
43	5	10	1	1	10
44	6	7	9	1	90
45	6	8	24	2	120
46	6	8	2	1	20
47	6	9	24	2	120
48	6	9	8	1	80
49	6	10	9	1	90
50	7	8	2	1	20
51	7	9	12	1	120
52	7	9	3	1	30
53	7	10	3	1	30
54	8	9	24	2	120
55	8	10	24	2	120
56	8	10	2	1	20
57	9	10	24	2	120
58	9	10	4	1	40

Table II. User Requirements and Associated Weighting Factors for the 10-Node Illustrative Example



NOTES:
(a) NUMBERS IN PARENTHESIS INDICATE REMAINING CHANNEL CAPACITY
(b) 1-2-5, 2-4-6-10, ETC. ARE ACTUAL ROUTES TAKEN BY THE ADDED LINKS

Figure 2. Modified Network Used to Allocate Remaining Circuit Requirements

A summary of circuit routing, required throughgroups, and voice drop locations by source-to-destination pairs for the 10-node sample problem is furnished in Table III. The first column lists the required user (S/D) pairs. The second column shows the original demand for each user pair. This demand should equal the sum of the commodities for each user pair shown in Table II. The third column shows that portion of the original demand (second column) carried over each route shown in the fourth column. Those routes containing circled nodes indicate the location of voice drops determined in the second LP run. The fifth column furnishes information relating to the number of throughgroups needed to route grouped demands (first LP run) that required indirect routing. The sixth column contains a list of the voice drops for each user pair. The number of voice drops is determined by multiplying each circuit by the number of circled nodes in each route. Thus, to obtain the number of required groups per user pair, each satisfied group is multiplyed by the number of intermediate nodes used by each route. The number of throughgroups for all requirements is 33, the minimum needed to obtain the maximum number of weighted grouped requirements for this problem.

The results furnished in Table III are all that are required to generate the multiplex plan for the 10-node sample problem. The procedure is carried out manually in the following steps:

a. Position the nodes on a large sheet of paper (or several sheets depending on problem size) using the original network connectivity as a guideline and allowing sufficient space between nodes.

b. Using the satisfied grouped requirements first, draw lines between each pair of nodes using the links that describe the route. The appropriate number of throughgroups is specified at each intermediate node and the number of utilized spare channels is indicated for each link in the route.

c. Continue this procedure until the lines have been drawn and labeled for all satisfied grouped requirements. An important consideration in this step is to ensure that all lines sharing the same link are kept reasonably close to one another for easy identification of all links and to facilitate the tabulation of utilized system capacity.

d. Next, route the satisfied residual circuits using the spare capacity created in the previous steps. When a residual requirement has been routed, ensure that the utilized capacity of each link involved in the routing is updated by reducing the amount of spare capacity and correspondingly increasing the utilized portion within the previously routed groups.

e. Continue this procedure until the complete multiplex plan is drawn.

S/D PAIR	ORIGINAL CIRCUIT DEMAND	CARRIED DEMAND	ROUTE	NO. THROUGHGROUPS	VOICE DROPS
1-2	20	20	1-2		
1-3	16	16	1-3		
1-4	22	22	1-4		
1-5	7	7	1-2-5	1	
1-6	18	18	1-2-6	2	
1-7	10	10	1-4-7	1	
1-8	5	5	1-4-7-9-8	3	
1-9	15	15	1-4-7-9	4	
1-10	2	2	1-4-7-9⑧10		2
2-3	14	14	2-1-3	2	
2-4	12	12	2-1-4	1	
2-5	18	18	2-5		
2-6	32	32	2-6		
2-9	8	8	2-6-9	1	
2-10	8	8	2-6-9-10	2	
3-4	25	24	3-4		
		1	3-7-4	1	
3-5	7	7	3-1-2-5	2	
3-7	28	28	3-7		
3-8	3	3	3-7-9-6-8	3	
4-5	10	10	4-1-2-5	2	
4-6	30	30	4-6		
4-7	36	36	4-7		
4-8	14	12	4-6-8	1	
		2	4-7③7-9-6-8		2
4-9	20	12	4-7-9	1	
		8	4-6-9	1	
4-10	5	4	4⑥9-10		4
		1	4①4-7-9⑧10		2
5-6	2	2	5-2①2-6		2
5-7	6	6	5-8-9-7	2	
5-8	24	24	5-8		
5-10	13	12	5-8-10	1	
		1	5⑧10		1
6-7	9	9	6-9-7	1	
6-8	38	36	6-8		
		2	6-2①4-7-9-8		2
6-9	44	44	6-9		
6-10	8	8	6-9-10	1	
7-8	2	2	7-4①4-7-9-8		2
7-9	26	26	7-9		
7-10	3	3	7③7-9-6⑧10		6
8-9	36	36	8-9		
8-10	39	39	8-10		
9-10	40	36	9-10		
		4	9-6②6-9-10		4
TOTALS	675	675		33	27

Table III. Summary of Circuit Routing, Required Throughgroups, and Voice Drop Locations by Source to Destination Pairs for the 10-Node Network

This procedure was followed for the 10-node network, the results of which are shown in Figure 3. Essentially, each line represents a group of 12 channels or a multiple thereof. Further, one or two numbers appear on each line between two interconnected nodes. The first represents the number of utilized channels, while the second, where applicable, depicts the number of spare channels also carried in the same group. The small rectangles inside each node show the number of throughgroups required.

One additional form of output that is useful to the network planner is the summary of utilized system capacity provided in Table IV. This output is readily obtained from the results furnished by the multiplex plan. For completeness, the maximum stated channel and group capacity of each link shown in Figure 1 is furnished in the second and third columns. Columns 4 through 6 provide information as to the actual number of groups used, the number of utilized channels within the groups, and the number of spare channels carried with the utilized groups. This information is obtained

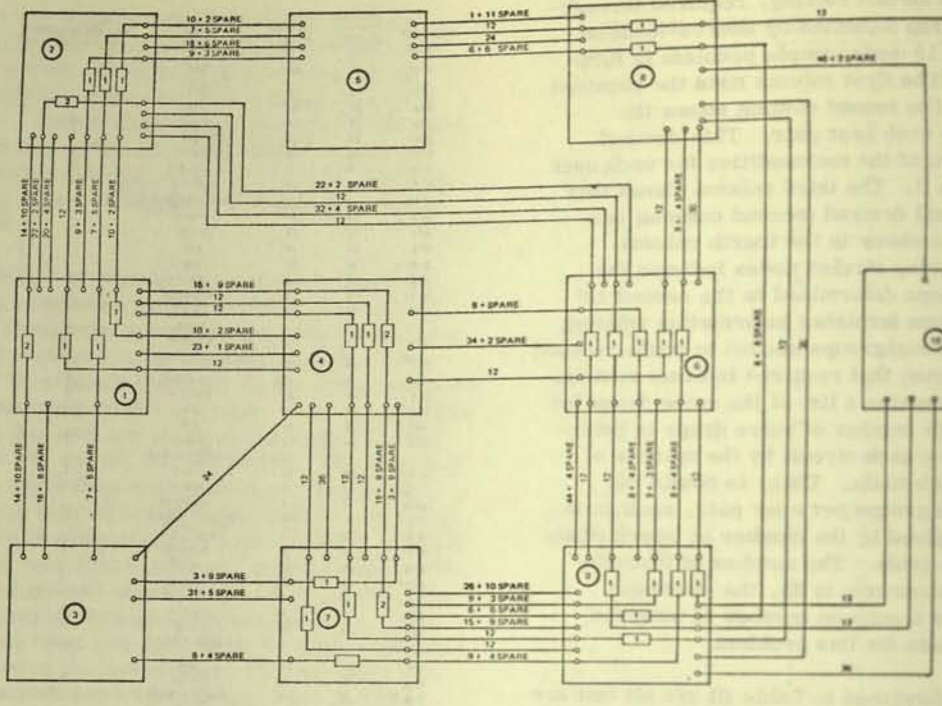


Figure 3. Multiplex Plan for the 10-Node Network

LINK NO.	MAXIMUM GROUP CAPACITY	MAXIMUM CHANNEL CAPACITY	NO. UTILIZED GROUPS	NO. UTILIZED CHANNELS WITHIN GROUPS	NO. SPARE CHANNELS WITHIN GROUPS	TOTAL NO. SPARE CHANNELS IN LINK
1-2	10	120	10	94	26	26
1-3	5	60	5	37	23	23
1-4	10	120	8	84	12	36
2-5	5	60	5	44	16	16
2-6	10	120	7	78	6	42
3-4	2	24	2	24	0	0
3-7	5	60	5	42	18	18
4-6	5	60	5	54	6	6
4-7	10	120	9	90	18	30
5-8	5	60	5	43	17	17
6-8	5	60	5	56	4	4
6-9	10	120	10	105	15	15
7-9	10	120	10	88	32	32
8-9	5	60	5	54	6	6
8-10	5	60	5	58	2	2
9-10	5	60	5	60	0	0
TOTAL	107	1284	101	1011	201	273

Table IV. Summary of Utilized Capacity for For the 10-Node Network

directly from the multiplex plan by straightforward tabulation of the numbers shown on each interconnecting line. The total number of spare channels present in each link is shown in the last column. These include the spare channels within the utilized groups and the remaining unused groups (not shown in Figure 3).

A summary of utilized and spare system capacity is useful to the network planner because it provides information needed for refinement and future planning after the basic multiplex configuration is determined. Since the technique outlined in this paper provides a multiplex design based primarily on achieving a high level of performance, some manual refinement is usually necessary to satisfy additional restraints not easily handled by linear programming. Network expansion due to anticipated changes in user requirements, identification of potential bottlenecks, flexibility, and reliability are but a few of the system factors that may result in additional refinement. Consequently, the ability to identify idle facilities after a basic design has been implemented is paramount if such refinements are to be made.

Conclusions and Recommendations

An approach for determining a high performance and efficient multiplex design for wideband communications systems was developed using linear programming techniques. The method was applied to a 10-node test network and gave excellent results for the case investigated. Furthermore, the model was found to be highly flexible in its ability to accept inputs that reflect engineering judgment as shown by the use of weighting and cost factors.

Although the model was developed primarily to eliminate the need for the intuitive approach in the grouping and routing of circuits, it still does not provide full automaticity with respect to generating the multiplex plan. This is especially true in the case of manually transforming the output data (Table III) into the actual drawing (Figure 3) for larget networks. Automated procedures can and should be developed that will reduce the time required to transform the routing and throughgrouping information into a practical multiplex drawing.

The technique described in this paper has provided a major improvement in generating routes for the majority of requirements that can be maintained at group levels. It is apparent, however, that the method could be carried further by proceeding from group to supergroup levels with some modifications of the basic approach. It should be noted that although a cost savings may result in maintaining supergroup levels, additional capacity is often necessary to satisfy the same number of circuit requirements than if only group levels were maintained. The procedure, nevertheless, warrants further investigation.

References

1. Ford, L.R., Jr. and D.R. Fulkerson, "A Suggested Computation for Maximal Multicommodity Network Flows," Management Science, Vol. 5, No. 1 (January 1958, pp 97-101.
2. Ford, L.R., Jr. and D.R. Fulkerson, "Maximal Flow Through a Network," Can. J. Math., Vol. 8, (1956) pp 399-404.
3. Tomlin, J.A., "Minimum-Cost Multicommodity Network Flows," Operations Research, Vol. 14, No. 1, (Jan-Feb 1966) pp 45-51.
4. Kalaba, R.E., and M.L. Juneosa, "Optimal Design and Utilization of Communication Networks," Management Science, Vol. 3, No. 1, (October 1956).

COMPARISON OF NETWORK TOPOLOGY OPTIMIZATION ALGORITHMS

V. Kevin Moore Whitney
GM Research Laboratories
Warren, Michigan, U.S.A.

Abstract

Analytical comparisons of procedures to optimize centralized communications networks of multi-drop lines are impossible except in a few very special cases. To facilitate comparison, several topology optimization procedures have been uniformly coded and applied to a variety of test configurations. The test cases were chosen to represent real communication networks having between fifty and five hundred terminal locations. The procedures selected are compared with respect to final network cost, procedure execution time, and procedure flexibility.

Introduction

Many recent papers have described communication network optimization procedures.[1-7] Each of these papers presents an algorithm and a few examples of its use, but comparisons with previous algorithms are seldom given. Centralized communication networks typically consist of a central node joined to hundreds of remote terminal nodes. The terminal nodes are connected to the central node with several multi-drop communication lines, as illustrated in Figure 1. Topology optimization procedures attempt to configure the terminals and the lines joining them to minimize cost while maintaining a satisfactory performance.

This paper will discuss only the optimization of centralized communication networks with tree topologies. While decentralized computer-communication systems may become more common in future years, few such systems are currently in use. The most significant decentralized system, the ARPA network, has been extensively studied in other papers.[1] Other networks with several central nodes are actually combinations of separate centralized networks joined by a few high speed links. For these systems, the component subsystems may be analyzed with procedures for centralized systems.

A very strong case can be made for considering only tree topologies for centralized communication networks. For certain communication cost functions, the tree can be shown to be less expensive than any non-tree topology.[6,9,10] Whenever the network cost includes a high charge for each additional edge of the topology, or when the cost of an edge is independent of the traffic, then the system cost will be minimized by minimizing the number of edges. Queueing studies indicate that the system delay is minimized when there are as few independent channels as possible.[6,8] In addition to these reasons, there are already many tree topology communication networks in use, and many manufacturers and users are well acquainted with their design and operation. These trends may be expected to continue.

Many different tree topologies may be used to link the remote terminal nodes to the central node of a communication system. The topology illustrated in Figure 1 for a system with 40 terminal nodes is the Minimal Spanning Tree (MST). This is the topology which has the least expensive set of edges and thus the absolutely lowest line cost when the line cost is the sum of the edge costs. Another tree topology which may be used when cost is no constraint is the star network illustrated in Figure 2. Although it is not the most expensive tree topology, a star network will typically cost several times as much as the MST (3.2 times in this example).

The minimal spanning tree and the star tree are two extreme cases. The star tree is expensive and high performance; the MST is least expensive and usually a poor performance topology. Many other topologies exist having better performance than the MST but are also more expensive. We will find it convenient to display the cost-performance tradeoff on a cost-performance graph as illustrated in Figure 3. A good topology optimization procedure will generate a series of topologies whose points on this graph lie on a curve extending from the MST toward the star graph.

The optimal topology for a communication system depends on the cost and performance measures used for the network. In all our examples, we use the standard multi-drop leased line tariffs. Performance will be quantized by line utilization, the ratio of line traffic (bps) to line capacity (bps). This performance measure is compared to several others by Whitney.[6] There it is shown that other commonly used performance measures lead to similar results.

In all of the topology optimization algorithms discussed in this paper, the performance constraint is an upper bound on the maximum line utilization. For a given maximum line utilization, the optimal topology is the topology with the lowest line cost. Thus the optimal topologies lie toward the lower right-hand corner of the cost-performance graph, the low-cost and low-utilization corner.

Assorted claims of optimal, sub-optimal, or low-cost network topologies generated by various algorithms have been made. Selecting the optimal topology from among the n^{n-1} possible trees is impossibly time consuming using the fastest computers available for values of n greater than about 10. Analytic tools to find the optimum are not available except for the minimal spanning tree or other suitably restricted problems. Hence it is necessary to search among all possible tree topologies, using heuristic methods to keep the search confined to a small number of good topologies. Since these design techniques are heuristic in nature, evaluation techniques are limited.

The best method for evaluating topology design algorithms is to try the proposed procedure on a variety of test examples. The network topologies produced may then be compared with the topologies produced by other optimization procedures. Other comparisons such as algorithm speed, flexibility, and convenience provide additional data for algorithm evaluation. These comparisons help select an algorithm which may be used dependably, confident that a good network will be designed. It is less important that an algorithm sometimes gives a good topology than that it never produces a really poor topology. Other desirable properties for a network topology optimization algorithm are speed of execution and the generation of a range of solutions.

Topology Optimization Algorithms

The five algorithms compared in this paper are the Minimal Spanning Tree,[5,10] the Constrained Minimal Spanning Tree,[6] Sectoring,[3,6] Reversed CMST,[3] and Steepest Ascent Hill Climbing.[1,2,6] These algorithms are specified in a common format to show similarities and differences. Each algorithm generates tree topologies joining n terminal nodes (or terminals or nodes) to a single central node (or center). The cost of a tree is the sum of the costs of its edges; performance of a tree is the maximal line utilization of a multi-point segment represented by some branch of the tree. The "optimal" topology for a specific line utilization is the lowest cost topology which has that maximal line utilization.

For comparisons of algorithm speed, it is convenient to assume that there are k branches in the generated tree, each branch containing $m = (n/k)$ nodes. The computation speed of the algorithms are expressed as a function of the number of nodes of the tree, the number of branches, and the number of nodes on a branch. These functions show the relationship between problem size and algorithm time.

The Minimal Spanning Tree topology is helpful in evaluating topology optimization procedures because it has the lowest possible line cost for the specified nodes. Often the MST is not a feasible topology because there are too many nodes on a single branch. In the example of Figure 1, there are 18 nodes on the upper branch. If the performance constraint of the network is that no branch is to carry traffic from more than 8 users, then this topology is unacceptable. It still gives a lower bound on the cost of networks with feasible topologies.

The most efficient procedure for generating the MST topology from an array of inter-node distances[5] requires only n^2 comparisons. The algorithm is elegantly simple. To begin, the node nearest the center node is joined to the center node. The node nearest to either of these two nodes is then joined to the nearer of the two. At each step, the node nearest to some node of the partially formed tree is adjoined to the partial tree. This process is continued until all nodes are in the tree. Thus the tree grows from the center node until it joins all the terminal nodes. The steps of the algorithm are shown below:

MST Algorithm

1. Let set P contain all the terminal nodes. Let set T contain only the center node.
2. Find the shortest edge from any node of P to any node of T. Let it be edge from node p to node t.
3. Add edge (p, t) to the MST. Remove node p from set P and add it to set T.
4. If there are any nodes left in set P, go to step 2.

The minimal spanning tree algorithm may be slightly modified to incorporate a performance constraint. The resulting Constrained Minimal Spanning Tree is guaranteed to be a feasible topology for the specified performance level. This CMST algorithm differs from the MST algorithm only in step 2.

CMST Algorithm

1. Let set P contain all the terminal nodes. Let set T contain only the center node.
2. Find the shortest edge from any node of set P to any node of set T for which the performance constraint will remain satisfied if the edge is added into the tree. Let this edge be from p to t.
3. Add edge (p, t) to the CMST. Remove node p from set P and insert it in set T.
4. If there are any nodes left in set P, go to step 2.

This algorithm grows the tree outward from the center node, at each step adding the nearest node which will not violate the performance constraint.

The computational requirements of this algorithm are greater than the MST algorithm but the resulting topology is guaranteed to satisfy the performance constraint. Unfortunately, the topologies generated often have high cost compared to alternate algorithms.

Another simple technique, also dependent on a fast MST algorithm, is Sectoring or pie-cutting. This algorithm divides the area surrounding the center node into sectors and joins all terminals in each sector with a single multi-point (MST topology) line. This is Algorithm 1 of Sharma and El-Bardai.[3]

Sectoring Algorithm

1. Point a ray from the center toward node b. Put the node b in set S.
2. Rotate the ray to the next node, say k.
3. If adding node k to set S will violate the performance constraint, then place the MST of the nodes of S in the final topology and reset the set S to contain only new node k. If adding the node k to the set S does not violate the performance constraint, add node k to set S.
4. If node k is not node b, then go to step 2 to start another branch of the tree.

By using different starting nodes, b, various topologies result from this algorithm. If there are m terminal nodes in each sector, then most of the distinct topologies which can be generated by this algorithm will be generated using m successive starting nodes. This sometimes results in considerable time savings.

If there are m terminals on each line, and every terminal is used once as a possible starting node, then the amount of computation is proportional to n (initial nodes) n/m (MST computations) m^2 (MST computation) = n^2m. If only m terminals are used as starting nodes, the amount of computation is proportional to n m^2.

A second technique advocated by Sharma and El-Bardai[3] is the Reversed CMST procedure, a variant of the CMST algorithm. This algorithm grows feasible MSTs starting with the farthest node from the center. When a branch cannot be extended, another is started from the farthest remaining node. More specifically:

Reversed CMST Algorithm

1. Let set P contain the terminal nodes.
2. Place in set T that node from set P which is farthest from the center node. Remove the node from set P.
3. Find the shortest edge from any node of set P to any node of set T for which the performance constraint will remain satisfied when that edge is added into the tree. Let it be the edge from node p to node t.
4. If any edge was feasible, add node p to set T, remove node p from set P, and go to step 3.
5. If no feasible edge was found, then place the MST of the nodes in set T in the final topology and remove all the nodes from set T.
6. If there are any nodes left in set P, go to step 2 to start another branch of the tree.

The computation for this algorithm is proportional to n^3 with the most time spent in step 3 checking for the nearest feasible node.

An extremely effective heuristic for network design, steepest ascent hill climbing (SAHC), was pioneered by Esau & Williams.[2] The essential steps of that algorithm are as follows:

SAHC algorithm, Esau & Williams variant

1. Select initial topology to be a STAR with center node b.
2. Find that pair of nodes i and j for which removing edge (i, b) and adding edge (i, j) reduces the network cost the most without violating the performance constraint.
3. If no such topology transformation exists, quit.
4. Make the change, and go to step 2.

Thus the algorithm improves a star topology, successively removing a central edge and replacing it with a less expensive non-central edge, until no further improvements are possible. Efficient coding of this algorithm results in an amount of computation proportional to n^2 where there are n terminal nodes of the network.

Closely related to this algorithm are a wide variety of other SAHC algorithms using other starting topologies or other sets of topology transformations. These variants have included considerations of stochastic message flow in the performance constraint, of network reliability as another performance constraint, modem and multiplexer as well as line cost, and non-centralized networks. The most general procedures are those of Frank et al.[1,4] This method is considered the most likely candidate for continued development of heuristic network optimization algorithms. There are several intuitive reasons for the success of this algorithm. Probably most important is that cost reductions are made in order of decreasing size so that longer edges are removed from the network first. With the CMST algorithm, the opposite is true. There the short central edges near the center node are added first, often causing problems when more distant nodes cannot be joined into an already started high utilization line. While the hill climbing nature of the algorithm may leave the design stranded on an isolated bump far from the ultimate hilltop, using a variety of initial topologies reduces the likelihood of this possibility.

In addition to these algorithms for designing a feasible network configuration, there are a variety of improvement procedures which perturb the network to produce a better solution. Attention has been concentrated in this paper on the basic procedures which can produce a variety of reliable network topologies for the designer. While perturbation techniques are important for fine tuning the final design, in the initial design states it is more important to have a procedure to produce good designs consistently.

The next section of this paper contains a discussion of some simulation experiments with these design algorithms. By trying the algorithms on a variety of problems, certain ones are shown to be consistently better than others.

Comparison of Algorithms

Each of the five algorithms described in the preceding section has been coded in FORTRAN IV and executed for a wide variety of possible networks. The results of the comparisons have been surprisingly consistent. For each of the algorithms, the network cost function has been the standard ATT leased multipoint line tariff. The performance constraint was that no line's utilization should exceed a specified value.

The maximum line utilization performance constraint was chosen as a realistic performance measure. Use of a more specific performance measure such as the average delay time of messages in the networks requires more detailed information about the message traffic than is generally available to the system designer. Utilization averaged over several lines results in uneven performance where heavily loaded

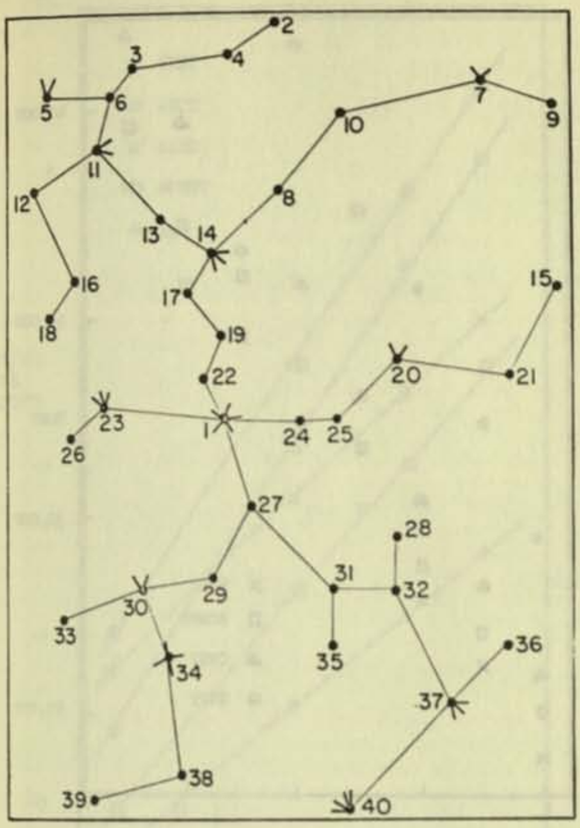


Figure 1. Minimal Spanning Tree Topology. Cost = 1430.

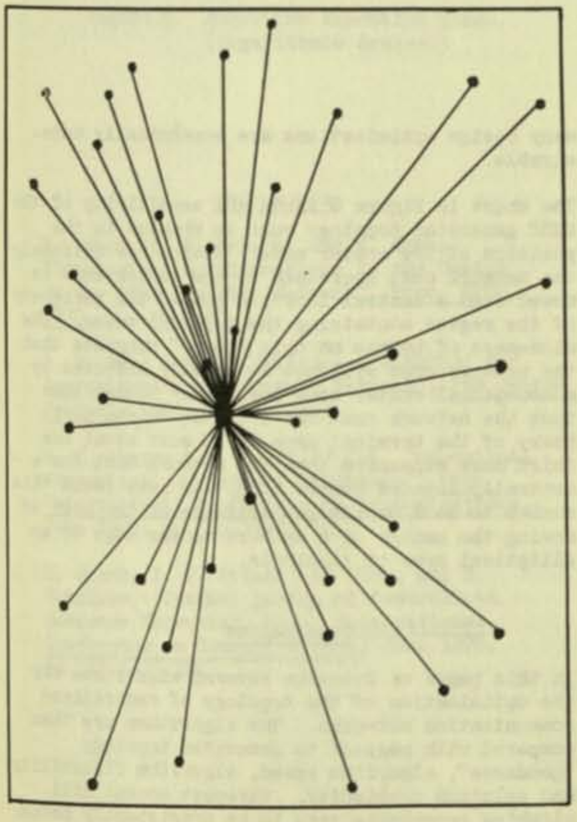


Figure 2. Star Topology. Cost = 4557.

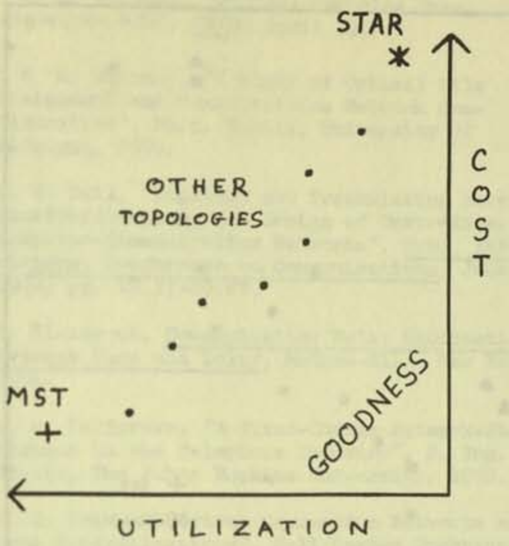


Figure 3. Cost-Performance Graph.

lines would serve the connected terminals poorly while lightly loaded lines would serve their terminals well.[6]

The most important criterion for evaluating network design algorithms is the quality of the generated topologies. The cost-performance values for each of the algorithms applied to the network of 40 nodes with a variety of maximal allowable line utilizations are shown in the cost-performance graph of Figure 4. A similar graph for a network of 100 nodes is shown in Figure 5 and one for a network of 200 nodes is shown in Figure 6. In all three cases, the SAHC procedure is superior to RCMST, CMST, and Sectoring algorithms. The MST algorithm is not truly comparable since it produces only a single topology independent of the performance constraint.

Another important evaluation criterion is the execution speed for the optimization procedures. Also of importance is the relationship of this speed to the size of the network to be optimized. The amount of computation required for each algorithm was discussed in the previous section of this paper and does not need to be repeated here. Execution time for each of the algorithms in seconds of 370/165 CPU time for several different network sizes is shown in Figure 7. That graph shows MST and SAHC algorithms to require time proportional to the square of the number of terminal nodes, while Sectoring, CMST, and RCMST require more time. A reasonable extrapolation of the trends of this chart shows the practicality of using the SAHC algorithm on networks of many hundreds of terminal nodes.

Another valuable characteristic of the SAHC algorithm can be extracted from the graphs of Figures 4, 5, 6. This is the closeness of points along the cost-performance "curve" generated by applications of the SAHC algorithm. The closeness of these points indicates that a small change in the performance constraint will result in only a small change in the total cost of the network topology. Thus the system designer need not worry that a slightly different performance constraint would have resulted in a network topology with significantly higher or lower total cost. This can be very important when the network is so large that

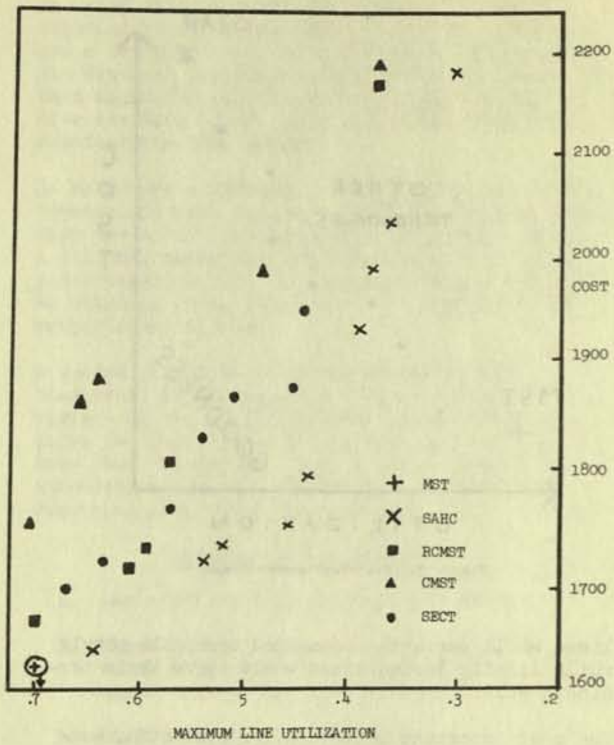


Figure 4. Cost-Performance Graph. 40 Node Network.

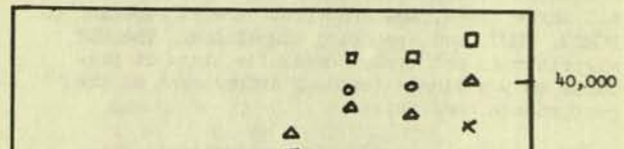


Figure 5. Cost-Performance Graph. 100 Node Network.

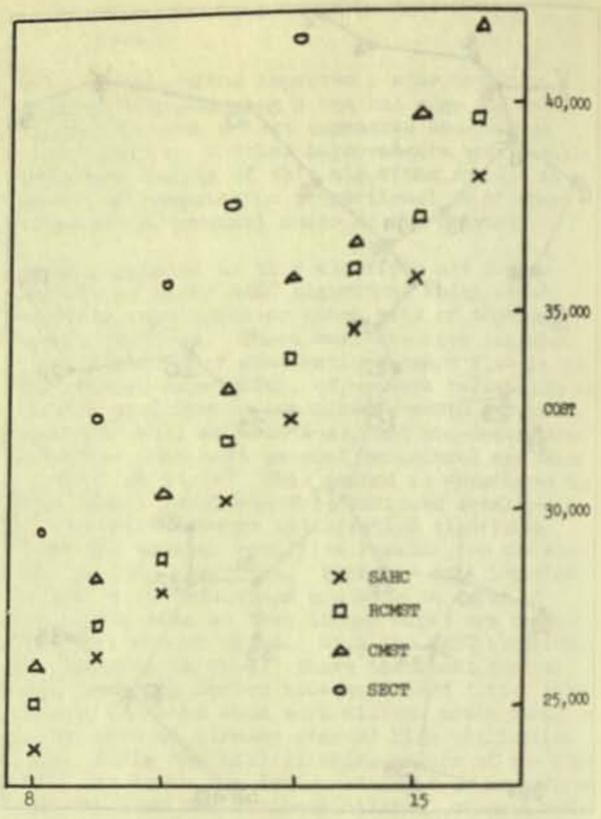


Figure 6. Cost-Performance Graph. 200 Node Network.

many design optimizations are economically undesirable.

The chart in Figure 8 shows the sensitivity of the SAHC generated topology cost to changes in the position of the center node. Notice how uniformly the network cost increases as the center node is moved from a central location toward the periphery of the region containing the terminal nodes. The closeness of points on this "curve" suggests that the network cost will not be greatly affected by a nonoptimal center node location. Notice also that the network cost for a center on the periphery of the terminal area is at most about one third more expensive than the network cost for a centrally located center node. We have found this number to be a consistent estimate of the cost of moving the center of a network to the edge of an elliptical area of terminals.

Conclusions

In this paper we describe several algorithms for the optimization of the topology of centralized communication networks. The algorithms are then compared with respect to generated topology "goodness", algorithm speed, algorithm flexibility, and solution continuity. Steepest ascent hill climbing procedures seem to be consistently better than the other algorithms.

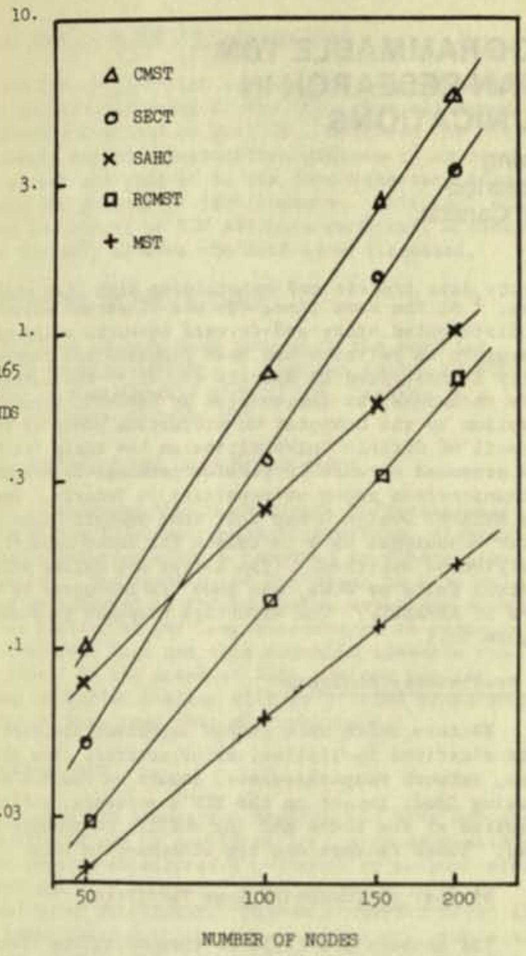


Figure 7. Algorithm Execution Times. (Logarithmic Scales.)

References

1. H. Frank, I. T. Frisch, and W. Chou, "Topological Considerations in the Design of the ARPA Computer Network", Proc. SJCC, 1970, pp. 581-587.
2. L. R. Esau & K. C. Williams, "A Method for Approximating the Optimal Network", IBM System Journal, 1966, pp. 142-147.
3. R. L. Sharma & M. T. El-Bardai, "Suboptimal Communications Network Synthesis", Proc. International Conference on Communications, June 1970, pp. 19.11-19.16.
4. H. Frank, I. T. Frisch, W. Chou, and R. VanSlyke, "Optimal Design of Centralized Computer Networks", Proc. International Conference on Communications, June 1970, pp. 19.1-19.10.
5. V K. M. Whitney, "Minimal Spanning Tree, Algorithm 422", CACM, April 1972.
6. V K. M. Whitney, "A Study of Optimal File Assignment and Communication Network Configuration", Ph.D. Thesis, University of Michigan, 1970.
7. D. R. Doll, "Topology and Transmission Rate Considerations in the Design of Centralized Computer-Communication Networks", Proc. International Conference on Communications, June 1970, pp. 19.17-19.27.
8. L. Kleinrock, Communication Nets; Stochastic Message Flow and Delay, McGraw-Hill, New York 1964.
9. W. W. Hardgrave, "A Fixed-Charge Network-Flow Problem in the Telephone Industry", D. Eng. Thesis, The Johns Hopkins University, 1962.
10. R. C. Prim, "Shortest Connection Networks and Some Generalizations", Bell System Technical Journal, Vol. 36, 1957, pp. 1389-1401.

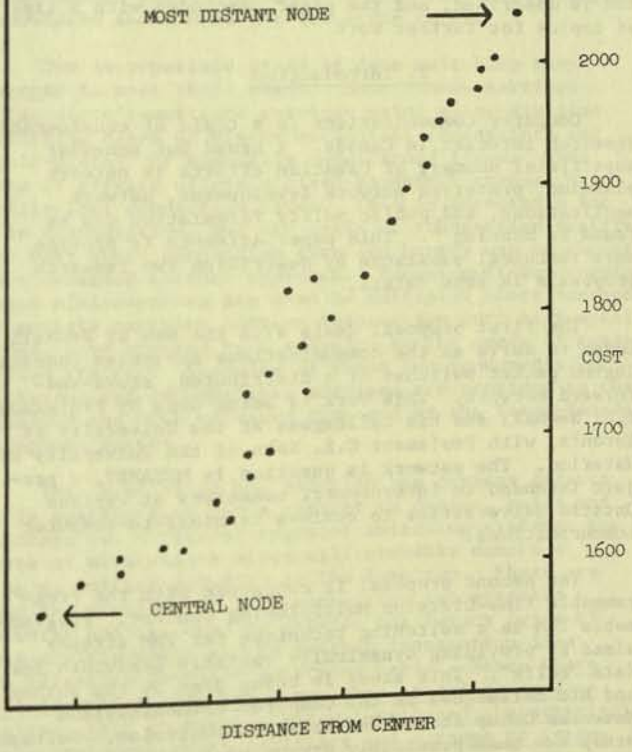


Figure 8. Network Cost vs. Center Location.

NEWHALL LOOPS AND PROGRAMMABLE TDM TWO FACETS OF CANADIAN RESEARCH IN COMPUTER COMMUNICATIONS

Eric G. Manning
University of Waterloo
Waterloo, Ontario, Canada

Abstract

Computer communications is a topic of considerable research interest in Canada. Communications is clearly of prime importance in a large, thinly-populated nation and computer communications research has received considerable attention and support. (A summary of Canadian efforts in network planning, prototype network development, network application, and public policy formulation can be found in Manning [1].) This paper describes two research proposals by Canadian workers - the development of the Newhall Loop concept to serve as the communications subnetwork interconnecting the switches of a distributed store-and-forward network; and the Programmable TDM concept as a method of dynamically varying the bandwidth assigned to data "calls".

The suitability of Newhall Loops to interconnect small packet switches is examined, and some simple calculations relating to response time and loop capacity are presented. The Programmable TDM concept is proposed as a desirable development in switching technique for future public networks, and examples are given of its potential usefulness. A model or "conceptual design" of a switching matrix for Programmable TDM is described, and the paper concludes with a list of topics for further work.

I. Introduction

Computer communications is a topic of considerable research interest in Canada. A broad but somewhat superficial summary of Canadian efforts in network planning, prototype network development, network applications, and public policy formulation can be found in Manning [1]. This paper attempts to provide more technical substance by describing two research proposals in some detail.

The first proposal deals with the use of Newhall Loops to serve as the communications subsystem connectingthe packet switches of a distributed, store-and-forward network. This work is being done by Professor E.E. Newhall and his colleagues at the University of Toronto, with Professor C.E. Køhn of the University of Waterloo. The network is question is METANET, a project intended to interconnect computers at various Ontario universities to achieve terminal-to-computer communications.

The second proposal is concerned with the Programmable Time-Division Multiplexing concept. Programmable TDM is a switching technique for TDM streams aimed at providing dynamically variable bandwidth for data "calls". This study is being done by the author and his colleagues in the Computer Communications Networks Group at the University of Waterloo. Neither study has been previously discussed in the literature.

II. NEWHALL LOOPS AND METANET

1. Background

The Newhall Loop concept was first introduced in a paper by Newhall and Farmer[2]. Also known as the Pierce Loop concept, it has received considerable attention as an attractive technique for dealing with bursty data traffic and maintaining high line utilization.[3] At the same time, the use of packet switching in distributed store-and-forward networks using minicomputers as switches has been proposed and successfully demonstrated by Roberts et. al.[4] The success of this technique, as implemented by ARPANET, prompted its adoption by the Computer Co-ordination Group of the Council of Ontario Universities as the basis for METANET - a proposed network to provide terminal-to-computer communications among universities in Ontario. However, the METANET Design Group felt that Newhall Loops could offer advantages as a technique for interconnecting the distributed switches. (The latter are called Node Control Units or NCUs, and they are analogous to the IMPS of ARPANET.) The structure proposed is shown in Figure II.1.

2. Preliminary Analysis

Factors which were judged important are cost of communications facilities, error control, line utilization, network responsiveness, impact on the NCU's processing load, impact on the NCU's software, buffering required at the nodes and the ability to measure traffic load. These factors are now discussed in turn.

a) Cost of Communications Facilities

The numbers and costs of communications lines are little affected by the choice of loop transmission. (Costs are based on Trans Canada Telephone System tariffs.) However the loop repeater/interface will be inexpensive; perhaps $900 per node.

b) Error Control

Something will be lost in giving up overconnection and alternate routings; however, the use of full-duplex lines has been exploited to give, in effect, a nested pair of counter-circulating loops. Certain troubles might be handled by moving traffic to the "other" loop, and the interface hardware will be very simple and thus reliable.

c) Line Utilization and Response Time

If we assume a six-node network, each serving 100 bps terminals, we can serve 50 terminals per node by specifying a 50-kilobit per second loop. For,

100 bps x 50 terminals per node x 6 nodes
= 30,000 bps total data rate;
leaving 20,000 bps or 40% of loop capacity
for framing characters and control packets.

Moreover, if packets are limited to 1,000 bits maximum a packet can be dispatched in 20 milliseconds. The passing of control from node to node requires 2 msec.; hence, each node will receive control at 122-millisecond intervals.

Thus a surprisingly simple calculation indicates that line utilization and response time may be quite favourable. Professor Newhall's group intends to do detailed simulation studies, similar to those reported recently by Newhall et. al.[5], to further investigate loop characteristics in this application.

d) Impact on NCU Processing Load

The loop repeater at a given NCU recognizes incoming packets addressed to its NCU, thus eliminating considerable overhead in the NCU. Moreover, an NCU need merely append a destination address to an outgoing packet and pass it to its loop repeater, further reducing NCU processing requirements. This fact also reduces the amount of NCU software required, although not as markedly as does the next item discussed.

e) Impact on NCU Software

The loop concept does away with the need for alternate-routing software. Thus the design of NCU software is considerably simplified, leading to a cheaper and more reliable NCU.

f) Buffering and Load Measurement

It is believed that double-rank buffering will suffice at the NCU-loop repeater interface, although detailed simulation will be required to substantiate this belief. The question of buffering is also related to load measurement in the following way. Any network requires facilities for load measurement in order to control offered load and thus maintain adequate response time. In the proposed loop system, regular sampling of buffer loading will be a prime measurement on which to base load control techniques.

3. Further Work

The need for simulation studies has been mentioned. Before METANET can be committed to loop communications, we also require experimental evidence to support simulation and analytic studies. This evidence is being gathered by an experimental two-node Newhall Loop, the nodes being located at the Universities of Toronto and Waterloo. Reliability, cost, line utilization and response time are under careful observation. We therefore expect to reach a decision vis-a-vis METANET in the near future.

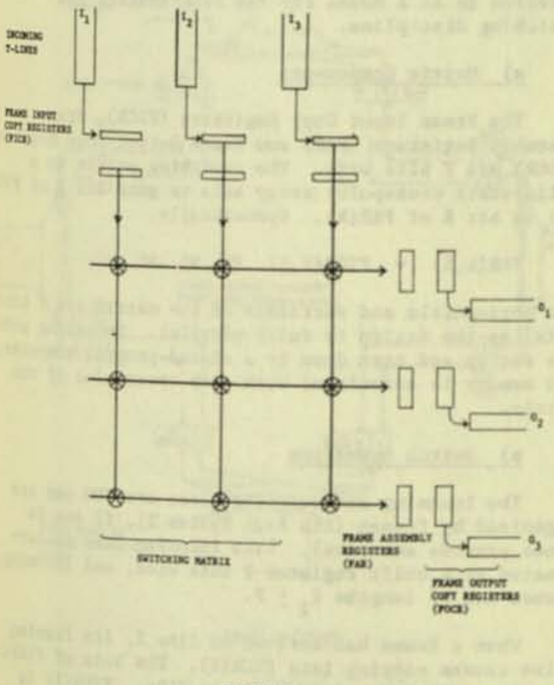


FIGURE III.1

III. PROGRAMMABLE TDM - A PROPOSAL

1. Background

Data traffic is growing rapidly and most of it is carried on the public telephone network. This has caused problems with connect times, error rates and user charges, as the switched voice network was not originally designed for data. In the United States these problems are being partially solved by the licensing of specialized data carriers such as DATRAN and MCI. However, in less wealthy countries, this option is impractical - rather, we must develop the switched voice network into an instrument for all types of telecommunications. The conversion of the network to digital transmission is a partial solution and this is gradually being done in Canada. However the area of switching also requires attention.

2. The Problem

Two distinct types of data traffic have already emerged. Short messages or transactions evoking rapid responses are one type, and they are characterized by the need for very short connection time, bursty use of channels, small volumes of data per call and the need for low error rates. The volume of transaction traffic generated by POS devices, data bank queries and related applications is expected to grow explosively.[6]

The other traffic type is file transmission traffic, such as occurs when a remote computer ships large files to another computer. File traffic is characterized by no great need for short connection time, smooth use of channels, a need for fairly large bandwidths and a need for low error rates.

Two incompatible kinds of data switching have emerged to meet these needs. Some common carriers offer circuit-switched services using space-division switching based on electromechanical crosspoints and "gold-plated" or specially conditioned lines. This type of service offers a fixed-bandwidth channel and fairly long connection times. This is acceptable for file transmission, but the needs of transaction traffic for very short connection time and bursty use of channels have demanded another approach. Packet-switched services where minicomputers are used to multiplex short messages or packets carrying address information onto high speed lines, have emerged as a solution to the needs of transaction traffic. Examples of these two fundamentally different and incompatible services are provided by the MULTICOM and SCCS services proposed by the Trans-Canada Telephone System.[7]

This kind of ad hoc solution has serious defects. It is unattractive for a carrier to design, build and maintain two different types of switching systems; the costs of maintenance alone will probably demand a single, unified solution in the long run. There are also the problems of interconnecting incompatible networks and the need to provide for future types of data traffic which may be neither transaction-like nor file-like in their switching needs. Hence there is evidently a need for a new type of switching discipline, able to provide circuit-switched characteristics, packet-switched characteristics and a range of intermediate characteristics. This paper proposes such a switching discipline, called Programmable Time-Division Multiplexing.

3. Assumptions

The assumptions germane to the rest of this paper are as follows:

1) Digital transmission links using end-to-end pulse regeneration techniques will gradually replace the present analog links. Digital transmission offers much improved error performance and the ability to handle data, facsimile, and digitally encoded voice in a uniform, economical manner.

2) Digital transmission links will use a format of fixed-length frames, much like the present T-1 digital link.

3) Packet switching may be impractical as a universal solution. It has not been demonstrated, for example, that the packet-switched ARPA Network can be expanded to include 8,500 switches - the number of wire centres in the American Bell System. More fundamentally, the central processing unit of a packet switch has to participate in the processing of each character transmitted. Despite the falling costs of CPU's, this may prevent the handling of video signals or large data files on public networks due to saturation of the switch CPU's. Therefore, a solution to the problem posed must not require that the switch's CPU process individual data characters.

4) Time-Division Multiplex switching has become feasible. After years of anticipation, the problems of coping with timing variations (pulse-stuffing) and of designing a fast, economical crosspoint are being solved in an economic manner. The active development of TDM switches in the U.S.A. (ESS-4)[8] illustrates this point.

5) It will be necessary to largely retain the present topology of the switched voice network. Thus, although we may gradually put new switching machines into the central offices, we will not be able to dig up and reroute very much cable. The cost of local distribution (70% of network cost) dictates this assumption.

4. Functional Performance Specification of a Programmable TDM Switch

This section is devoted to a description, from the user's point of view, of the capabilities offered by a programmable TDM switch. The key concept is that user can negotiate with the system to obtain the transmission bandwidth (data rate) desired. Moreover, the bandwidth assigned to a call in progress can be dynamically varied over a wide range. This mode of operation subsumes present-day circuit-switched performance and packet-switched performance as special cases.

1) Circuit-Switched Performance

To obtain circuit-switched performance, the caller A asks for a connection to B at a fixed bandwidth. Thus he issues control signals which specify "connection to B, bandwidth of n kbps". After the network responds ("go ahead") a file is transmitted and the called terminates the call. This is precisely the service provided by present circuit-switched systems.

2) Packet-Switched Performance

For packet-switched performance, the caller A asks for a connection to B at a high bandwidth, say 50 kbps, and sends a short packet, followed immediately by a command to change the bandwidth assigned to the call to zero bps. This mode of operation is identical in performance to packet-switching as the user gets exclusive use of a high-bandwidth channel for short periods of time.

3) Other Types of Performance

Modes of operation are also possible which fall between circuit and packet-switched performance. For example, a process A_2 running on Computer A could ship a sub-task to process B_{26} running on Computer B as follows:

(i) Process A_2 issues control signals asking for a connection to Computer B at 2 kbps.

(ii) The supervisory process B_1 of B answers and advises that B_{26} will be scheduled in twenty minutes.

(iii) Process A_2 reduces the call bandwidth to zero and becomes suspended (goes to sleep).

(iv) When process B_{26} enters the running state, it asks for a call bandwidth of 2 kbps and signals A_1 A_1 resumes A_2 and interprocess communication begins.

(v) A_2 and B_{26} agree via protocols to ship a file to B_{26}. Either of them increases the call bandwidth to 40 kbps, the task is transmitted and processed by B_{26}. The results are returned to A_2 and B_{26} is suspended. The call is terminated.

Points to note here are that:

(i) The switching performance required is neither packet-like nor circuit-like;

(ii) The call receives exactly as much of the communications resource-bandwidth-as is required, thus optimizing utilization.

(iii) This kind of operation is not readily available via existing telecommunication networks.

5. Conceptual Model for a Programmable TDM Switch.

This section describes a conceptual design or model for the switching matrix or network of a Programmable TDM switch. The design is sketched in Figure III.1. Considerable work will be required to verify the cost/effectiveness of this design or to modify it for practical development. Hence it is referred to as a model for the Programmable TDM switching discipline.

a) Matrix Components

The Frame Input Copy Registers (FICR), Frame assembly Registers (FAR) and Frame Output Copy Registers (FOCR) are F bits wide. The switching matrix is a solid-state crosspoint array able to gate bit j of FICR (i) to bit ℓ of FAR(k). Symbolically,

$$FAR(k,\ell) \leftarrow FICR(i,j) \quad \forall i \quad \forall j \quad \forall k \quad \forall \ell.$$

The horizontals and verticals of the matrix are F bits wide, as the design is fully-parallel. Switching paths are set up and torn down by a stored-program computer, and memory is associated with each crosspoint of the matrix.

b) Switch Operation

The incoming and outgoing lines are TDM and are organized by frames (the Bell System T1, T2 and T4 lines provide examples). Each incoming line is terminated at a shift register F bits wide, and incoming frames are of lengths $f_i \leq F$.

When a frame has arrived on line I_1 its framing pulse causes copying into FICR(1). The bits of FICR(1) are then switched and sent to the FARs. FICR(1) is now free to receive the next incoming frame from I_1. Concurrently, the same process is occurring on the other input lines. Periodically, at a rate R, a master output event occurs which copies all FARs simultaneously to the FOCRs and output begins on the outgoing T-lines. CPU intervention occurs only when some input stream's bandwidth is to be changed, or a call is to be set up or torn down.

c) Switch Parameters

Let the switch repetition rate be R.

Let $B = FR$

be the switch bandwidth.

Let b_i be the bandwidth of incoming line I_i.

Then

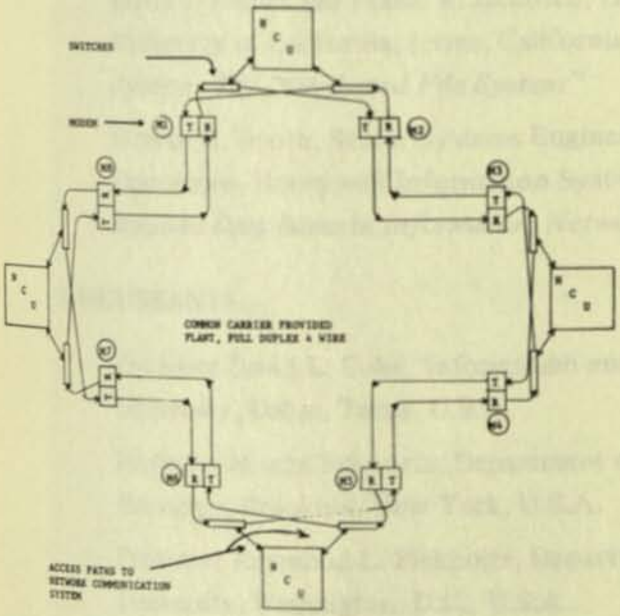
$$b_i \epsilon \frac{j}{Fm} B = \frac{j}{m} R$$

where

$$j \epsilon \{1, 2, 3, \ldots, F\}$$

and m is any positive integer.

This result is established by noting that incoming frames on I_i can be of any fixed length j up to F bits, and that the frame arrival rate on I_i can be any integral submultiple R/m of R. Moreover, the bandwidth b_i of incoming line i can be apportioned among the calls carried in exactly the same manner. A given call can occupy any subset of the j bits of frames of I_i and then bits can be occupied in every frame, every second frame and so forth. Limits to this theoretical flexibility will be set by the difficulty of matching new requests for bandwidth to the free slots available, and by the extra complexity of operating the switch in this manner.



DUPLEX LOOP SYSTEM

FIGURE II.1

6. Steady-State Operation

A hardware frame detector associated with each FICR signals the presence of a complete incoming frame. The "frame complete" signal of FICR (i*) is sent to all crosspoints $M(i^*,j,k,\ell)$ on the ith vertical of the switching matrix. Under the steady-state assumption the memory bits of all crosspoints are currently set; routing of the frame bits to the correct bits of the correct FARS therefore occurs. The transmission of data to a FAR sets a "data present" control bit associated with the FAR; this control bit AND the master output event causes the FAR to be copied to its FOCR. Outpulsing on all output lines which have received data therefore begins simultaneously.

Thus, all steady-state control of the switching matrix can be accomplished in hardware; processor intervention is required only when calls are set up, torn down or require variations in assigned bandwidth. Requests for such changes of network state would be delivered to the control processor via a dedicated control channel. The control processor would compute new values for the crosspoint bits $\{M(i,j,k,\ell)\}$ and would copy these into the crosspoint flipflops in order to effect the desired change in matrix operation.

7. Further Work

The present discussion is only a proposal, and falls far short of making an unchallengeable case for Programmable TDM. To make such a case, we must evaluate switching matrix costs and throughput for present MSI technology, propose refined designs as necessary, develop algorithms for matrix control to be implemented in the control processor, do simulation studies on the amount of inter-office signalling required, and evaluate such items as connect time, blocking probability, switch bandwidth and switching delay. Most of these topics are under investigation at Waterloo.

ACKNOWLEDGEMENTS

The idea of using Newhall Loops in conjunction with a distributed, packet-switched network was put forward by E.E. Newhall; the preliminary analysis reported here was done by E.E. Newhall, C.E. Køhn, and N. Housley. The experimental two-node loop is being developed by Professors Køhn, Newhall and the author, using loop interfaces provided by E.E. Newhall Associates Limited. Thanks is due to the Computer Communications Region of Bell Canada who have contributed lines for the two-node loop.

The Programmable TDM concept was suggested independently by the author and W.M. Gentleman, and is being studied by ourselves, D.E. Morgan and C.E. Køhn. Support of the National Research Council of Canada and the Defence Research Board is gratefully acknowledged.

REFERENCES

1. Manning, E.G.; Computer Networks North of the Border; EDUCOM Spring Conference, Washington, D.C., U.S.A.; 13 April 1972.

2. Newhall, E.E. and Farmer, W.D.; An Experimental Distributed Switching System to Handle Bursty Computer Traffic; Proc. ACM Conference, Pine Mountain, Georgia, U.S.A.; October 1969.

3. Pierce, J.R.; Network for Block Switching of Data; Bell Systems Technical Journal.

4. Roberts, L.G. and Wessler, B.; Computer Network Development to Achieve Resource Sharing; AFIPS Conference Proceedings, Volume 36, 1970 SJCC, pp. 543-549.

5. Newhall, E.E., et. al.; Traffic Flow in a Distributed Loop Switching System; Polytechnic Institute of Brooklyn, Symposium on Computer Communications; New York, N.Y., U.S.A.; April 1972.

6. The DATRAN Company; Comments to the (U.S.) Federal Communications Commission re: docket no. 18920; Washington, D.C.; U.S.A.; 10 October 1970.

7. Trans Canada Telephone System; Computers, Communications and Canada; document issued by TCTS Headquarters, Place Bell Canada; Ottawa, Canada; November 1971.

8. Vaughan, H.E.; An Introduction to No. 4 ESS; International Switching Symposium Record; IEEE, New York City, N.Y.; U.S.A.; June 1972.

9. Manning, Eric, Gentleman, W.M., Kóhn, C.E., and Morgan, D.E.; Programmable TDM - A Proposal; University of Waterloo Research Report, Waterloo, Ontario, Canada.

DATA NETWORK DESIGN PROBLEMS II

CHAIRMAN:

George J. Lissandrello
IBM World Trade Corporation
New York, New York, U.S.A.

PROGRAM COORDINATOR:

Dr. Carl Hammer, Director,
Computer Sciences, UNIVAC,
Washington, D.C., U.S.A.

PAPERS:

Rémi F. Despres, Centre National D'Etudes des Telecommunications, (CNET), Issy les Moulineaux, France, *"A Packet Switching Network with Graceful Saturated Operation"*

Joseph F. Marchese, Manager, Telecommunication Systems, and W. Gerhard, Data Communications Center, IBM World Trade Corporation, IBM Zurich Research Laboratory, Rueschlikon, Switzerland, *"Some Effects of Switched Network Time Delays and Transmission Speed on Data Based/Data Communication Systems"*

John M. Husted, COMSAT Laboratories, Clarksburg, Maryland, U.S.A., *"Current and Near Future Data Transmission via Satellites of the Intelsat Network"*

David J. Farber and Frank R. Heinrich, Department of Information and Computer Science, University of California, Irvine, California, U.S.A., *"The Structure of a Distributed Computer System–The Distributed File System"*

Grayce M. Booth, Senior Systems Engineer, Advanced Systems Design, Phoenix Computer Operations, Honeywell Information Systems, Phoenix, Arizona, U.S.A., *"The Use of Distributed Data Bases in Information Networks"*

DISCUSSANTS:

Professor David L. Cohn, Information and Control Sciences Center, Southern Methodist University, Dallas, Texas, U.S.A.

Professor Mischa Schwartz, Department of Electrical Engineering, Polytechnic Institute of Brooklyn, Brooklyn, New York, U.S.A.

Professor Raymond L. Pickholtz, Department of Electrical Engineering, George Washington University, Washington, D.C., U.S.A.

THEME: Two distinct approaches have been employed during the past years as data network designers either adapted to existing facilities or proposed new facilities to meet special requirements. It is still too early to draw definite conclusions about the relative merits of these and related concepts but their continued study is expected to provide us with the analytical tools needed to design for the future.

A PACKET SWITCHING NETWORK WITH GRACEFUL SATURATED OPERATION

Rémi F. Despres
Centre National D'etudes Des
Telecommunications (CNET)
Issy Les Moulineaux, France

ABSTRACT

More experience on packet switching appears to be necessary before a large scale data transmission network featuring this switching mode can be developed. Saturated operation in particular, has to be considered from the outset as a normal state because of computer traffic flexibility.

A good control of transmission resources allocation must therefore be exercised by the network, so as to give rapid service to high priority users without slowing down too much low priority communications.

A network structure, compatible with this requirement, is proposed. It comprises fixed routing, private buffering, packet handshaking, variable length packets and multirate scheduling. An evaluation of this design will be available when an experimental network, which is being developed along these lines, becomes operational.

INTRODUCTION

The need for a data transmission network adapted to computer trafic has now been recognized for a long time[3 5] ; several administrations and compagnies are developing such networks[1 4 7 9 12].

Two main switching techniques have been considered for this purpose, and arguments have been exchanged regarding their relative suitability to data transmission. Both have successfully been applied to real networks and therefore must be examined attentively[11 14].

On one hand we have the well known circuit switching organization where a physical path is established between two correspondants for the complete duration of their call.

On the other hand we have the less known packet circuit organization where :

1) Data are stored into buffers at each exchange traversal ;

2) Transmission lines are time-shared among active communications ;

3) Data input into the network is subordinate to the corresponding output rate or to transmission capacity, whichever is most restrictive.

Our purpose here is to discuss one particular aspect of packet switching, namely network behaviour in case of saturation. It will be argued that there are economical and technical reasons for considering saturated operation as normal. Then a network structure will be proposed for achieving smooth operation when user requests are competing for the same saturated transmission or switching resource.

This structure is being experimented on a small scale network to be opened in 1973. A better knowledge of packet switching should result from this project before a larger scale packet switching system can be developed.

SATURATED VERSUS CONGESTED OPERATION OF A TIME-SHARED RESOURCE

For any time-shared resource, be it a computer CPU or a transmission line or any other, a distinction must be made between saturation and congestion ; the former is indicative of good resource utilization, the latter has to be avoided.

In order to give precise definitions of these terms in the present context, let us introduce a few notations.

R is the time-shared resource

Ui with $1 \leq i \leq N$ is one of the N users of resource R which are active, i.e. waiting for their turn to be served.

Mi is the minimum percentage of resource R which must be allocated to user Ui to satisfy his needs. Mi may be a function of t but we assume that its fluctuations are slow enough to be negligible.

Di is the extra percentage of R which user Ui can use if so much is made available to him. Di/Mi is a measure of the flexibility of Ui's needs.

ei is the percentage of R which is effectively allocated to user Ui. Whereas Mi and Di are user dependant parameters ei depends not only on the sum of user needs but also on allocation rules applied to resource R.

L is the load of R defined as the sum of all minimum needs of active users, i.e.

$$L = \sum_{1 \leq i \leq N} Mi$$

E is the effective utilization of R defined by

$$E = \sum_{1 \leq i \leq N} ei$$

Resource R is said to be "saturated" if E = 100 %. It is said to be "congested for user Ui" if $ei < Mi$ and it is said to be "congested per se" if at least one of its users Ui doesn't receive its minimum claim.

It is clear that, with normal allocation, no congestion can appear before saturation (the unused percentage of resource R, if there is any, can be allocated to users whose needs are not satisfied in order to improve their service). But the converse is not necessarily true. If users have flexible needs, resource R can be saturated while being far from congestion. If the average value of Di/Mi is represented by λ, R can remain saturated with loads L varying from $\frac{100\ \%}{1+\lambda}$ to 100 %.

Optimal resource allocation must prevent congestion as long as the load is less than 100 %. In practice a safety margin is mandatory as users cannot specify their minimum needs with a good precision, and because allocation rules can hardly distribute resource R in exact proportion to these needs.

Figure 1 gives a typical load curve of a shared resource where user needs have no flexibility. A telephone network can be taken as example of such a resource [13] : most of transmission capacity is unused at night and cannot be used to improve the bandwith of the few active communications.

Figure 2.a gives a typical utilization curve of a time-shared resource whose users have flexible needs. While load varies full resource utilization can be maintained until the load is so low that every user gets as much service as he can use.

Figure 2.b presents the case of a resource with permanent saturation ; it can be the case if either the load is stable or users have very flexible needs when the load is light.

FLEXIBILITY OF COMPUTER TRANSMISSION NEEDS

Several facts concur to show that computer traffic can be very flexible.

1) In general computers can transmit data at much higher rates than transmission lines capacity and can adapt to any lower rates.

2) Base band transmission allows for cheap fast transmission on local loops.

3) Computers don't have to cease operation at night and can therefore benefit by transmission network availability at this time.

4) Even interactive users have rather flexible needs, as it results from a study by JACKSON and STUBBS on multiaccess computer communications [10] ; service rate reductions of the order of 50 % are tolerated by users of scientific time-sharing systems.

NETWORK DESIGN PHILOSOPHY

The preceding remarks justify the general philosophy we adopted for packet switching network design. It can be summarized as follows :

If a packet switching network involving computer to computer communications were to never reach saturation, it would have to be largely oversized so as to resist bursts and shifts of computer traffic. A sounder approach is to build the network for permanent, or at least frequent, saturated operation.

A primary objective must then be good

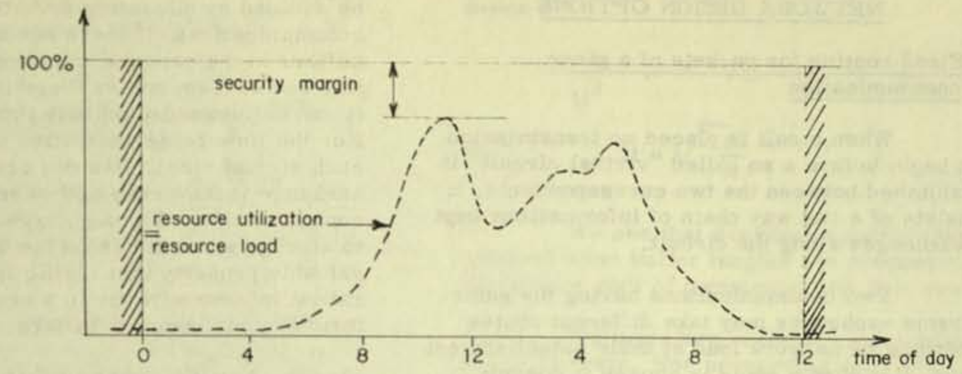


Figure 1 : Typical Resource Utilization with Rigid User Needs

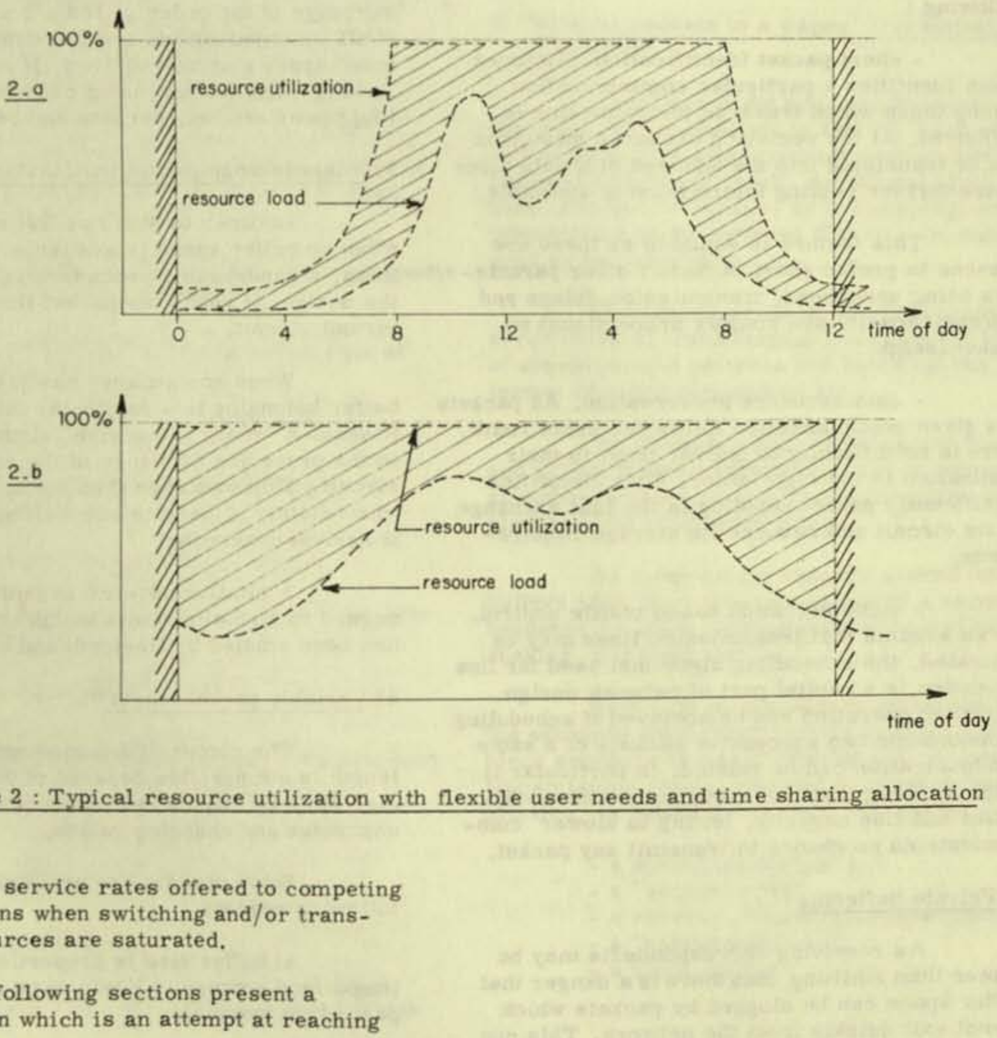


Figure 2 : Typical resource utilization with flexible user needs and time sharing allocation

control of the service rates offered to competing communications when switching and/or transmission resources are saturated.

The following sections present a network design which is an attempt at reaching that goal.

NETWORK DESIGN OPTIONS

1) Fixed routing for packets of a given communication

When a call is placed no transmission can begin before a so called "virtual circuit" is established between the two correspondents. It consists of a two way chain of informations kept in exchanges along the circuit.

Two communications having the same extreme exchanges may take different routes depending on network load at their establishment times. But once a route is chosen it doesn't change, except possibly in case of interexchange line failure.

Advantages of fixed routing are the following :

- short packet identification. An index which identifies a particular communication among those which traverse the same line is sufficient. At the recieving exchange this index can be translated into the address of a data block where further routing information is available.

This feature is valuable as there are reasons to prefer short packets : other parameters being unchanged, transmission delays and storage capacity are roughly proportional to packet length.

- data sequence preservation. As packets of a given communication follow the same route there is no difficulty to deliver them to their destination in the right order. This simplifies significantly packet handling in the last exchange of the circuit and reduces its storage requirements.

- communication based traffic control. As we assume that transmission lines may be saturated, the scheduling algorithm used for line allocation is a capital part of network design. Smoother operation can be achieved if scheduling decisions for two successive packets of a same communication can be related. In particular it helps avoiding that fast communications monopolize full line capacity, letting to slower communications no chance to transmit any packet.

2) Private buffering

As receiving correspondents may be slower than emitting ones there is a danger that buffer space can be clogged by packets which cannot exit quickly from the network. This can be avoided by allocating private buffers to communications. If there are also common buffers in the network, no packet must be sent into one of them unless there is a guarantee that it can be forwarded quickly into a private buffer. For the time being we settled on a solution where each virtual circuit has one private buffer in each exchange it traverses and where there is no common buffer. Although expensive with regard to storage space it has on the other hand the valuable property that traffic in one direction cannot influence traffic in a cross direction through their common buffers.

Assuming that an exchange handles up to 100 simultaneous communications, that on the average each communication involves three exchanges, and that an average packet buffer is 20 bytes long, we get a buffer storage per exchange of the order of 100 x 3 x 20 = 6000 bytes. If all communications are full duplex, twice as much space may be required. If storage space is lacking while transmission capacity and computing power are not, packets can be made shorter.

3) Interexchange packet handshaking

In order to avoid packet transmission when no buffer space is available to receive them, a handshaking procedure is included in the design. It regulates packet flow on each virtual circuit.

When an exchange has just freed one buffer belonging to a particular communication it sends a "ready to receive" signalling packet to the preceding exchange of the same virtual circuit. This exchange then sends a data packet, immediately if there is one waiting, as soon as it arrives otherwise.

A similar network organization, with regard to virtual circuits and packet handshaking, has been studied by Beeforth and others in [9].

4) Variable packet length

The choice of a unique optimum packet length is not possible because of the many parameters to be considered, most of which have unprecise and changing values.

Short packets are preferable for the following reasons :

a) buffer size is proportional to packet length (and storage is likely to be an expensive part of the system).

b) transmission delays, both due to buffer filling and due to queue lengths, are proportional to packet length.

Long packets on the other hand are better for the following reasons :

a) packet processing time is practically independent of packet length, so that a given CPU has a switching throughput in inverse ratio to packet length.

b) packet headers and other service data, like "ready to receive" signals, constitute a fixed overhead associated with each packet transmission. Total overhead for a given data throughput is thus in inverse ratio to packet length.

In a situation where most of the traffic would be due to a few rapid communications, while the majority of communications would contribute little to the network load, it would clearly be unwise to give equal buffer space to all communications. Very short buffers should be allocated to all the slow communications and a few large ones should be reserved for rapid communications.

The following quantitative result supports the preceding intuitive remark :

let ri denote the transmission rate of communication Ci

for i = 1, 2, ...,n ;

let li denote the length of the private buffer allocated to Ci ;

let L be the total buffer space.

The number of transmitted packets per unit of time is then :

$$p = \sum_{i=1}^{n} \frac{ri}{li}$$

It must be minimized with the constraint that

$$\sum_{i=1}^{n} li = L$$

This yields :

$$\frac{\partial}{\partial li} \left(\sum \frac{ri}{li} - \lambda \left[L - \sum li \right] \right) = 0$$

hence

$$\frac{-ri}{li^2} + \lambda = 0$$

or $$li = \sqrt{\frac{ri}{\lambda}}$$

We see that overhead minimization is reached when buffer lengths are proportional to the square root of communication data rates.

We therefore chose to have variable length packets. For preserving good line utilization when packets are short, while using efficient detection techniques, a transmission procedure allowing for packet grouping had to be designed. Its description follows.

5) "Several packets in a wagon" transmission procedure

Usual error detection techniques require that data are transmitted in blocks, each of which is preceded by a synchronizing pattern and is followed by a block check. Additional block overhead is incurred because an acknowledgment is transmitted on the reverse direction to signal good or bad reception.

The choice of block length depends on bit error rates of transmission lines, on the length of synchronizing patterns and block checks, on the format of acknowledgments etc.

We chose to make transmission block and packet length independent, and to associate acknowledgments related to one direction with blocks of the other direction.

As these blocks must be stored into buffers when they enter an exchange a second handshaking level, independent of packet handshaking, has been included.

Rejection of non-corrupted blocks by the exchange they reach is thus avoided and block sequence is easily maintained. A complete transmission frame, as used for packet transmission, is called a "wagon". It includes :

- a synchronizing pattern
- a "sequence tag"
- a reverse channel acknowledgment
- a "handshaking tag"
- a block check.

The "sequence tag" makes it possible to detect lost wagons. It comprises a sequence identification, which is zero or one, and a wagon index giving the rank of the wagon in the current sequence. The latter is taken modulo n, where n is such that an exchange never sends more than n-1 consecutive wagons without having received any acknowledgment. For most situations n = 4 seems to be sufficient.

The receiving end of a line keeps track of wagon indices and negatively acknowledges a wagon with index k if it receives in its place a wagon with index h.

A positive or negative acknowledgment comprises the sequence index of the wagon being acknowledged and a sequence identification. The latter is identical to the sequence identification of the wagon being acknowledged if acknowledgment is positive ; it is opposite if acknowledgment is negative.

When a new wagon is sent, the only reverse channel acknowledgment to be inserted is the most recent one (positively or negatively acknowledging a wagon implies a positive acknowledgment for all previous ones).

After having negatively acknowledged a wagon i of sequence s, an exchange ignores all wagons the same sequence and waits for the first one of the new sequence $\bar{s}$. If it comes with index i it is positively acknowledged otherwise negatively. A transmission error thus entails in general several wagon retransmissions but never leads to data loss, permutation or duplication.

The "handshaking tag" gives the number of data characters for which there is buffer space.

If eventually no buffer space is available or if there are no data to be sent, empty wagons are sent on the line. Their purpose is to convey acknowledgments and handshaking information concerning the reverse channel. When an empty wagon is sent, it bears the same sequence index as the next non empty wagon to be sent ; no sequence index incrementation results and empty wagons are not acknowledged.

The transmission procedure above is applicable to any full duplex transmission with non negligible error rates (typically of the order of 10^{-4} or 10^{-6}).

6) Multirate scheduling

Since we have assumed that CPU's or lines could be permanently saturated, the scheduling algorithm is clearly a central part of the design. Two objectives must be fulfilled.

- Under fixed load, each communication gets a stable service rate

- The service rate of a communication which is slowed down by the network can be increased by changing its service class.

The algorithm being experimented is a simplified version of a more general one implemented on the Ramsès II computer at the CNET [6]. The general version has given satisfaction to users for more than a year, providing them with stable service rates and several service classes.

The main characteristics of the simplified algorithm are the following :

- Each request for service is put at the end of the queue of its service class. A "deadline", obtained by adding a constant to a so called "resource time", is attached to it.

- Each selection of the next request to be served involves a comparison of queue heads ; the one with the earliest "deadline" is selected.

- "Resource time" is increased, at each new selection, by a quantity depending on the resource load.

After being processed a request goes to another queue where it stays until "resource time" reaches it deadline. A new request cannot be presented by a communication before its previous one has left the second queue.

Such a sophisticated scheduling is of course somewhat time consuming. But we expect that its counterpart, namely graceful saturated operation, will legitimate its use. If it does, some specialized hardware can be built in the future to improve switching throughput.

CONCLUSION

Saturated operation has been considered as an essential aspect of packet-switching networks, and design approachs have been proposed to cope with it.

The experimental network on which these techniques will be field tested has, on purpose, a limited capacity. It will consist of a triangle with 16 bits mini-computers at its vertices and one or several 9600 bauds lines on each edge. Each exchange will handle up to 200 customer lines with speeds ranging from 50 bauds to 9600 bauds. Programming work started at the end of 1971 and first users are expected to be connected before the end of 1973.

BIBLIOGRAPHY

1) G. D. ALLERY and K. J. CHAPMAN
Features of a Synchronous Data Network for the United Kingdom
ICC 71, pp 31-10 to 31-13

2) T. H. BEEFORTH, R. L. GRIMSDALE, F. HALSALL, D. J. WOOLLONS
Aspects of a Proposed Data Communication System
IFIP 71, pp TA4 - 74 to 78

3) D.W. DAVIES
The Principles of a Data Communication Network For Computers and Remote Peripherals
IFIP 68

4) D. W. DAVIES
Packet Switching in a Public Data Network
IFIP 71, pp TA4 - 69 to 73

5) J. B. DENNIS
A Position Paper on Computing and Communications
Comm. ACM n° 5, may 1968, pp 370 to 377

6) R. DESPRES et A. BACHE
Algorithme d'Allocation d'Unité Centrale pour Calculateurs Exploités en Temps Partagé
Annales Telecom., Tome 27 n. 1-2, pp 19-30

7) J. M. GAUTHIER
Les Réseaux de Téléinformatique
L'Echo des Recherches, oct. 1970, pp 4-11

8) P. B. HANSEN
An Analysis of Response Ratio Scheduling
IFIP 71, pp TA3 - 150 to 154

9) F. J. HEART, R. E. KAHN, S. M. ORNSTEIN, W.R. CROWTHER, D. C. WALDEN
The Interface Message Processor for the ARPA Computer Network
SJCC 70, pp 551 to 567

10) P. E. JACKSON, C. D. STUBBS
A Study of Multiaccess Computer Communications
SJCC 69, pp 491 to 504

11) A. JOUSSET, A. PROFIT
Le Projet Caducée
L'Echo des Recherches, april 1970, pp 10-17

12) A. JOUSSET
Etudes de Réseaux de Commutation de Données en France :
Structure du Réseau Hermès
ICC 71, pp 31-17 to 31-20

13) R.R. MINA
The Theory of Teletraffic Engineering
Telephony, april 71, pp 32 to 37

14) L. G. ROBERTS, B.D. WESSLER
Computer Network to achieve Resource Sharing
SJCC 70, pp 543-549

SOME EFFECTS OF SWITCHED NETWORK TIME DELAYS AND TRANSMISSION SPEED ON DATA BASED/DATA COMMUNICATION SYSTEMS

J.F. Marchese and W. Gerhard
IBM Research Laboratory, Zurich, Switzerland

Abstract

In many countries throughout the world, there are plans for development of Switched Telecommunications Networks designed to efficiently accommodate data processing systems. Most of todays data processing systems that require telecommunications facilities operate over private, dedicated networks which are leased from the Telecommunications Administrations. These users will be confronted with a significantly new environment, which will pose many problems and questions, ranging from the need for new procedures and techniques, to the cost and performance effects of these new switched networks.

This paper examines two of these questions: the possible effects on data processing cost and performance of switching delays and line transmission speed. Based on the relatively select user characteristics and overall assumptions, it is concluded that these new networks will be confronted with various demands which are dependent on the users system requirements. These demands range from the need for very fast network switching for large data processing systems (many terminals connected to a computer), to high transmission speeds for the large system and the batch data processing user, to relative insensitivity to both of these parameters for the small data processing user. In all cases, these users have optimized their cost/performance in todays environment and will require at least an equivalent from these new facilities.

Introduction

In many countries throughout the world, there are plans for development of Switched Telecommunications Networks designed to efficiently accommodate data processing systems. These new switched facilities will attempt to attract data processing users who presently operate over dedicated, private, leased communications networks as well as those systems now using switched telephone facilities. Since the private network data based/data communications systems have been relatively optimized for a nonswitched environment, there is an important question of what will be the effect on these users of this new switched network environment. It is obvious that many users will continue to find private networks to be more desirable for their particular needs.

The questions and problems that will confront the data processing user of this new environment include: (1) what are the optimum line control procedures and access methods for this switched network environment; (2) will concentrators and multiplexors provide cost and performance advantages, and under what circumstances; (3) what signalling techniques should be employed between terminal and network; (4) what are the effects of the switch and signalling delays on the system; (5) what line transmission speed should be used; and (6) a host of other similar questions.

Since the problem is extremely complex, it is desirable to simplify the analyses and examine the issues separately. This paper considers the switching speed effects (time delays) and line transmission speed effects on several representative classes of potential users of these new switched networks. In particular, several classes of existing inquiry/response data processing systems were selected as the basis for the study.

In any study of this type, the results are quite sensitive to the assumptions, and therefore the conclusions may change if the assumptions prove to be invalid or unreasonable. During this paper, an attempt will be made to indicate what the effects may be if the assumptions differ from the ones indicated. Certain conclusions will be drawn which hopefully will be of use to planners and designers of Switched Telecommunications Networks, as well as data processing system designers.

Characterization of Network Users

In data based/data communications systems, there are a wide variety of applications. One of the types in wide use today (and expected to continue in the future) is the inquiry/response system. In these systems, the operator at a remote terminal wishes to obtain information from a central computer as well as input some data into the Central Processing Unit (CPU). These systems will be among the most sensitive to data network functional characteristics, and will be used as one of the basis for this analysis. Examples are, Airline Reservation Systems, Banking Systems, Management Information Systems, etc. These users characteristically impose some response time requirement on the data processing system in order to satisfy the overall application demand (e.g., many Airlines require a 3 second response time from the time the inquiry is made at the terminal to the time the response is made from the CPU). In addition many remote terminals are usually connected to the CPU.

Since there are a wide range of parameters that classify data network users, it was decided to establish only a few general system categories in order to simplify the analysis. Thus, three classes of users were defined. They were based on averaging many existing IBM users. In addition, it is assumed that no Data Processing Concentrators would be used in this new environment, although this choice obviously depends on the users system configuration and requirements. The key characteristics are indicated in Table 1.

Table 1

USER SYSTEM CONFIGURATION CHARACTERISTICS

Characteristic	Large User	Small User	Special
Number of Terminals	800	30	100
Peak Terminal Call Rate (in busy hour)	0.01/s	0.02/s	0.03/s
Response Time Requirement (90th Percentile)	3-4 s	3-4 s	4-10 s
Mean Message Lengths			
In	24 Characters*	20 Characters	400 Characters
Out	120 Characters	50 Characters	400 Characters
*A character is 8 bits			

The study is presented in three sections: (1) the effects of switching speed on the user (in terms of communications lines and cost versus connect time); (2) the effects of transmission speed on the user (in terms of buffer size and cost versus transmission speed); and (3) the combined effects of switching speed and transmission speed on the user (in terms of total communication cost including transmission, line control and buffering).

The Effects of Switching Speed

Several assumptions were made in order to perform the analysis. These include:

1) Definition of communications tariffs
   - a) The user is tariffed on the basis of holding time (message length) with the minimum metering interval not longer than the actual message time. This may be an ideal solution from the users point of view, but some telecommunications administrations have indicated it may be impracticable. Thus, if the metering interval is longer than the message (e.g., one second), the results could change, although this is a function of the comparative costs of line communications and the Transmission Control Unit (TCU).
   - b) The user pays more for higher speed lines in the ratio of $(X)^{1/3}$. That is, a 9600 bps line costs $(\frac{9600}{2400})^{1/3}$ times as much as a 2400 bps line (in this case 1.6 times)(1). The factor 1/3 is a result of a study conducted in the U.S. (See reference 1). However, some administration have indicated the factor should be closer to 2/3. This would cause a reduction in the cost advantage of high-speed lines compared to low-speed lines.
2) Establishment of line control routines
   - a) The terminal initiates a call to the CPU only when he has data to send or obtain. It disconnects from the network after its call, and the computer redials when it has a reply for the terminal. Computer to terminal "handshaking" routines are assumed to either be included in the signaling routines, or to be revised in the future so as to have no effect on the results.
   - b) Similar to (a), but the remote terminal holds the line until the computer makes the reply.
3) The network is used as a concentrator, and lines are connected to the computer through a TCU. The TCU cost is related to the number of line attachments in a linear fashion.

With these assumptions, and employing conventional analytical techniques (Erlang formulas), it is possible to relate the number of communication line appearances at the computer as a function of the network switching speed and the total message traffic for the various classes of users (since the message lengths affect the line hold time at the CPU end of the network). The number of lines is approximated by the following equation[2] (assuming exponential message lengths and exponential call arrival rates):

$$L = BT + 2.4\sqrt{BT} \tag{1}$$

where:

L	is the number of lines required between the computer and network
B	is the mean call rate in calls per second (for the busy hour)
T	is the mean line hold-time in seconds
2.4	results from a call-blockage probability of 1%.

Figure 1 shows the results for the three classes under consideration.

From Figure 1, it can be concluded that the Small System and the Special System (with long messages) are relatively insensitive to the network switching speed. However, the Large System is significantly affected, and line requirements begin to increase rapidly when the network switching speed exceeds around 100 milliseconds.

The results of Figure 1 were then extended to obtain a possible cost relation. In addition to the assumptions previously described, appropriate communications tariffs were employed. To obtain a representative tariff cost per message transmitted, an existing IBM system operating over private lines in England was chosen, and the cost per call computed. This system was felt to be reasonable since the terminals were geographically disbursed throughout the entire country. The cost per call included both line tariffs and modem costs. The cost of an existing IBM TCU was also used. The results are indicated in Figure 2. Only the Large System was plotted since the other two were not critically sensitive to switching speed (except for the users response time requirement which is sensitive to the switching speed in a linear fashion). In these results, it is assumed that the Network Terminating Unit (future equivalent to todays modems) costs are independent of speed (up to 9600 bps).

From Figure 2, we observe that the user can reduce his total communications cost (for TCU and lines) by operating at higher line speed. Furthermore, this class of customer would desire switching speeds on the order of 100 to 200 milliseconds (or less).

In this computation, one potential major factor has not been considered: the cost of switching as a function of switching speed. It is possible that switch costs are inversely proportional to switching speed, which could change the results of Figure 2. This is an area that requires further study. (It is also theoretically feasible that TDM switching techniques provide fast switching at no extra cost since the primary requirement of the switch is to service a large number of lines with an aggregately high calling rate). This switching cost does not exist in a private, nonswitched network, and may be one of the factors that could cause continuing support of leased facilities by some users.

The Effects of Network Transmission Speed on the Data Processing System

The number of communication lines between the network and the computer is determined by the traffic characteristics and the switching speed as mentioned in the previous section. In this section, the relationships between the network transmission speed and the communications line requirements as well as computer buffer (memory) requirements will be described.

Line Requirements

As noted earlier, the number of line appearances at the computer is affected by the message characteristics, since these characteristics are directly related to the line holding time, T. The line holding time is a variable, depending on the line speed, message length, signaling time, and switching speed. Thus, we can indicate the line requirements as a function of the transmission speed by the following:

$$L = B\ (x + \frac{Im}{s} + 2.4\ \sqrt{B\ (x + \frac{Im}{s}} \qquad (2)$$

where:

L and B are as before

x is the sum of the call connect-time and disconnect-time plus signaling time

Im is the mean message length in bits

s is the transmission speed in bits per second.

Assuming x is a constant (of appropriate value) Figure 3 describes the line requirements as a function of the speed and message characteristics for the three classes of systems under investigation.

Buffer Requirements

In general, a message arriving at a computer (for processing and other appropriate actions) requires storage at the computer for the following periods of time:

1) While the message is being received and assembled
2) When the message is received and being queued for processing
3) While the message is being processed
4) After processing while the reply is being queued for output
5) While reply is being transmitted and disassembled.

Queuing for processing and output can be done either in the main storage of the computer or at a peripheral storage device (e.g., disk, tape, or drum) while buffering for message processing must be performed in the main storage at the computer. Buffering for message assembly and message disassembly can be performed either in the main storage of the computer or in a storage within the Transmission Control Unit at the computer center. The buffer requirements, for these two purposes, depend on several factors, including the transmission speed of the data over the line.

There are at least two methods for allocating storage to a communication line:

1) Static buffer allocation, where a fixed amount of storage corresponding to the longest message is permanently assigned to each line.
2) Dynamic buffer allocation, where the storage is pooled and each line requests buffer from the pool as the need arises. For this method two cases can be distinguished:
    a) The unit size corresponding to the longest message is employed within the buffer pool,
    b) A unit size smaller than the average message length is used within the buffer pool and these units are chained to blocks.

The total buffer requirements in a teleprocessing system are reduced by employing the dynamic core allocation technique with a unit size smaller than the average message length (method 2b). This method of buffer allocation is used for the investigation of the buffer requirements at the computer site, and the availability of adequate software support is assumed.

The buffer requirements are calculated under the following assumptions:

1) The buffer allocations for the various message streams are mutually independent
2) The message length (Im) has an exponential distribution
3) The call rate (B) during peak hour has an exponential arrival time
4) Prespecified probability of buffer overflow is 1%.

In order to determine the buffer pool size, one has to derive the mean and the variance of the buffer units required at the computer to accommodate the exponentially distributed message traffic flowing into and out-of the computer.

The procedure to calculate the mean number of units is to assess the requirements (statistically) on a per-line basis and then sum the total requirements over the number of lines employed in the system (which was determined earlier). It can be shown that the mean number of buffer units for m lines is given by:[2, 3]

$$U_m = B\,(\frac{Im}{b})^2 \cdot \frac{Im}{s}\,(e^{b/lm} - 1) \qquad (3)$$

where:

U_m is the mean number of buffer units required to accommodate the total message flow at the computer

b is the number of data characters in a buffer unit (for data assembly and disassembly).

<u>Note</u>

b is a key parameter and will be discussed further.

In a similar fashion, the variance can also be determined. It is given by:

$$U_{var} = U_m\,(2 \cdot \frac{Im}{b} - u^2)$$

where:

U_{var} is the variance of the buffer unit size

U_m is the mean (from above)

u is the mean number of buffer units required on a per-line basis.

From these two expressions, one can compute the total buffer size necessary for the system. As indicated earlier, the probability of buffer overflow was assumed to be permitted to reach 1%. With this assumption, an approximation for the total buffer size can be made.

It is:

$$U_{total} = U_m + 2.4\sqrt{U_{var}} \qquad (4)$$

where:

U_{total} is the total buffer number of buffer units necessary to satisfy the system.

As noted earlier, the parameter b, which is the number of data characters assigned to each buffer unit (in other words, the block size for data within the dynamic buffer), is a key parameter in these computations. Obviously, if one stores segments of a message in a buffer, it is necessary to allocate some overhead information along with these segments in order to chain them together (for total message reassembly). Thus it is desirable to optimize this process so that the buffer is not overloaded with redundant overhead information (chaining characters), and still keep the unit size, b, small enough to reduce the total buffer requirements. From Equations 3 and 4, the optimum buffer size can be determined. It is:

$$b_{opt} = \sqrt{2C \cdot Im} \quad (5)$$

where:

b_{opt} is the optimum buffer unit size in bits for data

C is the number of bits necessary to chain segments together

Im is the mean message length in bits.

The parameter C varies according to software system applications, but in one general usage control program, it is four characters (32 bits). Thus b can easily be computed for the three data processing systems under evaluation.

The total buffer size, U_{total} is then compared as a function of transmission speed and shown in Figure 4. As shown, the Large System and the system with long messages can realize an appreciable reduction in buffer requirements when the transmission speed is increased (in this example, at least up to 9600 bps).

Combining the Switching Speed and Transmission Speed Effects

It is now possible to observe the combined effects of the network switching speed and transmission speed on the data processing systems being considered. The information contained in Figures 2 and 4 (and using an appropriate cost per bit for the buffer) is combined into one figure. The variables are, relative cost, switching speed, and transmission speed.

Figure 5 shows the relative total cost of communications (including transmission costs, TCU costs, and buffer costs) as a function of transmission speed for several selected values of network switching speed. Only the Large System is shown to avoid a confusing amount of data in the figure. It can be seen that the total cost can be reduced significantly by operating the system at higher transmission speeds (from 4800 to 9600 bps) and with relatively fast network switching (around 100 to 200 milliseconds).

The results for the Small System do not show the same sensitivity to either transmission speed or switching speed. No significant savings can be achieved by increasing either parameter.

For the Long Message System, a cost improvement can be obtained by operating at higher transmission speeds since this reduces buffer requirements. However, these systems are relatively insensitive to switching speeds (unless they become excessively long, say greater than 1.5 to 2 seconds).

Conclusions

This study examined, in a rather limited manner, some of the effects on a data processing user of communication network switching speed and line speed.

One potentially significant factor has not been considered, and remains to be examined when further quantitative cost information is obtained: that is, the sensitivity of switching cost to switching speed. The results indicated in this study could be altered if the switching costs increase significantly as the switching speed is made faster.

Three basic conclusions can be drawn for the classes of data processing users examined when operating over switched data networks:

1) The large inquiry/response type of data processing user, with communications requirements, will impose the most significant functional demands on the switching and transmission speed of the network to achieve minimum cost and satisfactory performance.

2) Users operating with rather long messages (e.g., batch applications, etc.) are relatively insensitive to switching speed, but normally desire high transmission speed to minimize cost and optimize performance.

3) Many small data processing users (few terminals connected to a CPU) do not generally achieve any significant cost advantages from fast switching or high transmission speed. They can generally be satisfied by medium speed switching and transmission which provides sufficient performance for their response time requirements. These users are more sensitive to excessive tariffs that may be established for all customers in order to satisfy the two classes above.

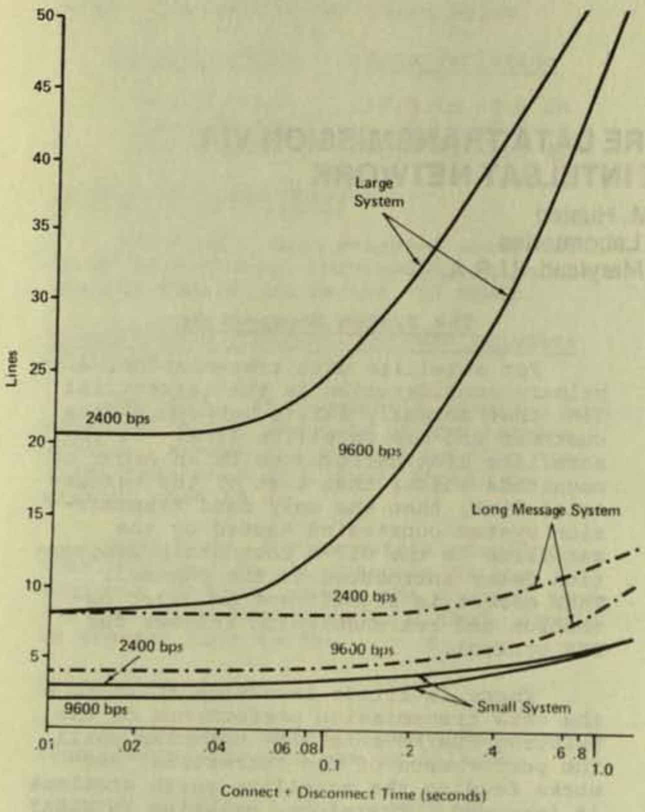
In addition to these conclusions, the many unanswered questions, plus the additional cost for switching, coupled with the need for data processing system changes will probably cause some users to continue to use private, leased, nonswitched communications facilities when they find it to their cost/performance advantage.

References

1. "Patterns of Technology in Data Processing and Data Communications", Stanford Research Institute, Feb. 1969.

2. "Analysis of Same Queuing Models in Real-Time Systems", IBM Manual No. GF 20-0007-1.

3. Martin, James, Design of Real-Time Computer Systems, Prentice Hall, 1969.



Note: Probability of Blockage = 0.01

Figure 1. Number of Lines at the Computer as a Function of Network Connect/Disconnect Time for Different Teleprocessing Systems

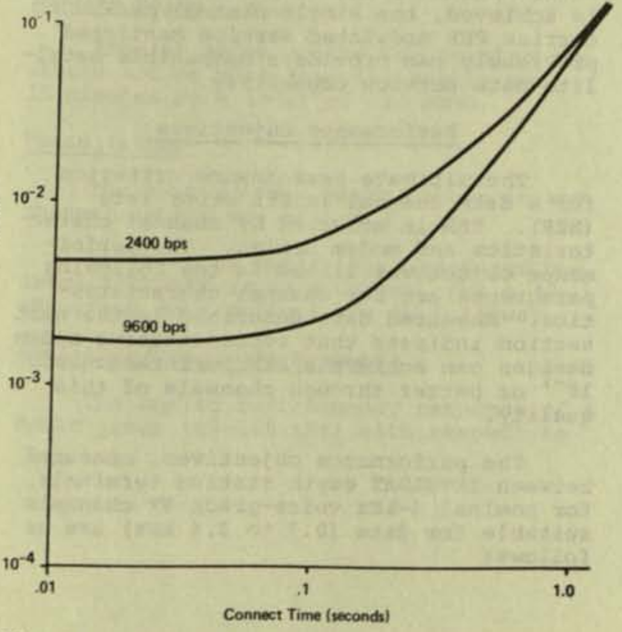


Figure 2. Cost (TCU + Transmission) as Function of Connect Time (For Large Systems Only)

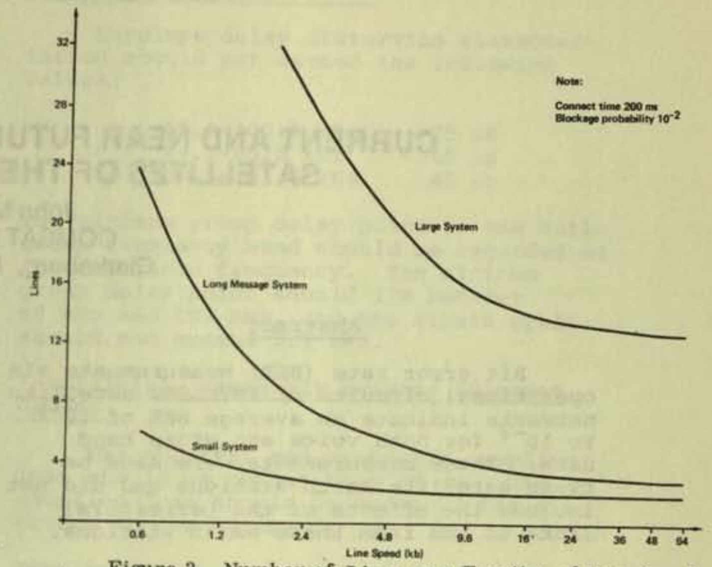


Figure 3. Number of Lines as a Function of Line Speed

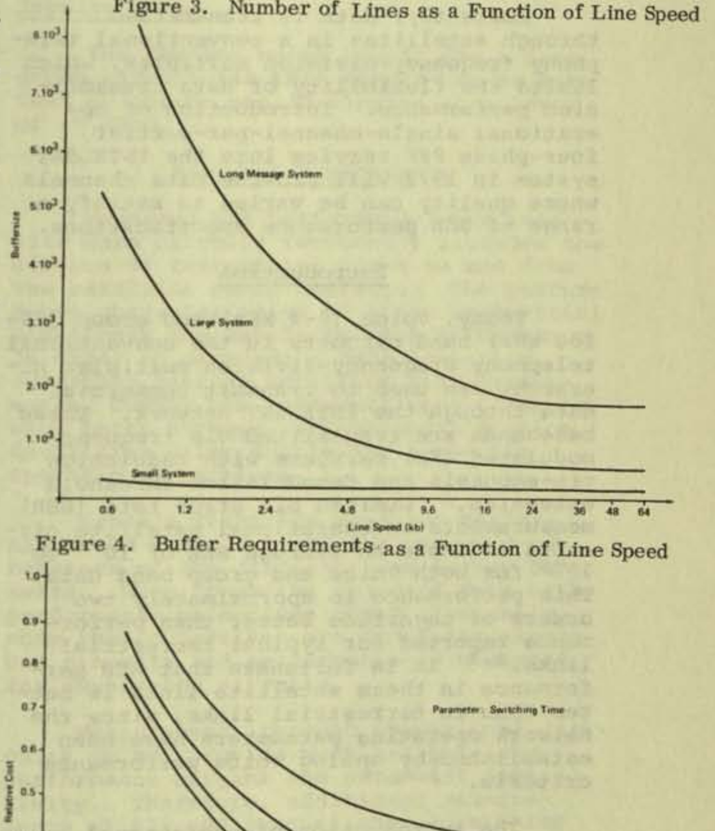


Figure 4. Buffer Requirements as a Function of Line Speed

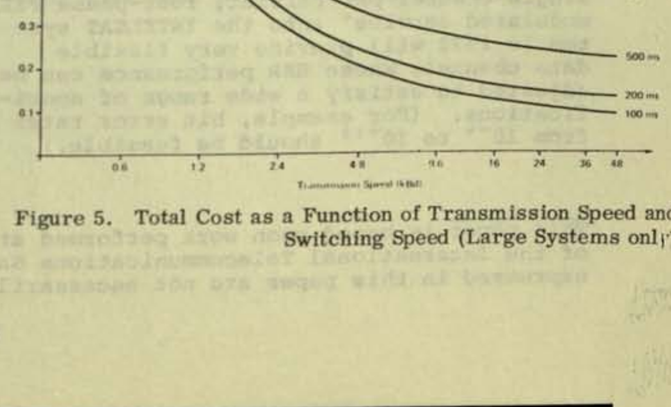


Figure 5. Total Cost as a Function of Transmission Speed and Switching Speed (Large Systems only)

CURRENT AND NEAR FUTURE DATA TRANSMISSION VIA SATELLITES OF THE INTELSAT NETWORK

John M. Husted
COMSAT Laboratories
Clarksburg, Maryland, U.S.A.

Abstract

Bit error rate (BER) measurements via operational circuits of INTELSAT satellite networks indicate an average BER of 10^{-6} to 10^{-8} for both voice and group band data. These measurements were made between satellite earth stations and did not include the effects of the terrestrial links to and from these earth stations.

Currently, data is transmitted through satellites in a conventional telephony frequency-division multiplex, which limits the flexibility of data transmission performance. Introduction of operational single-channel-per-carrier, four-phase PSK service into the INTELSAT system in 1972 will provide data channels whose quality can be varied to satisfy a range of BER performance specifications.

Introduction

Today, voice (0-4 kHz) and group (60-108 kHz) band circuits in the conventional telephony frequency-division multiplex hierarchy are used to transmit commercial data through the INTELSAT network. These basebands are transmitted via frequency modulated (FM) carriers with modulation pre-emphasis and demodulation threshold extension.[1] Limited bit error rate (BER) measurements on operational satellite circuits indicate an average BER of 10^{-6} to 10^{-8} for both voice and group band data. This performance is approximately two orders of magnitude better than performance reported for typical terrestrial links.[2,3] It is fortunate that BER performance in these satellite links is better than in terrestrial links, since the network operating parameters have been established by analog voice performance criteria.

The introduction of operational single-channel-per-carrier, four-phase PSK modulated service[4] into the INTELSAT system in 1972 will provide very flexible data channels whose BER performance can be adjusted to satisfy a wide range of specifications. (For example, bit error rates from 10^{-4} to 10^{-10} should be feasible.)

The System Perspective

For satellite data transmission, a primary consideration is the terrestrial link that normally exists between a data customer and his satellite link. If the satellite link performance is an order of magnitude better than that of the terrestrial link, then the only data transmission system constraint caused by the satellite is the 0.7-s round trip propagation delay introduced in the channel. This effect is significant in error detection and retransmission schemes for BER control.[5]

There is little incentive to improve the data transmission performance of the existing FDM/FM satellite networks until the performance of the terrestrial networks feeding the satellite earth stations is improved. Therefore, existing INTELSAT data transmission facilities should be adequate for the near future. However, the emergence of data carrier facilities separate from the telephony transmission systems holds some promise of improved BER performance in terrestrial data transmission. Whenever this improved performance is achieved, the single-channel-per-carrier PSK modulated service mentioned previously can provide a compatible satellite data service capability.

Performance Objectives

The ultimate performance criterion for a data channel is bit error rate (BER). BER is affected by channel characteristics and modem design. The performance objectives listed in the following paragraphs are for channel characteristics. Measured data described in the next section indicate that representative modem designs can achieve a BER performance of 10^{-6} or better through channels of this quality.

The performance objectives, measured between INTELSAT earth station terminals, for nominal 4-kHz voice-grade VF channels suitable for data (0.3 to 3.4 kHz) are as follows:

This paper is based upon work performed at COMSAT Laboratories under the sponsorship of the International Telecommunications Satellite Consortium (INTELSAT). Any views expressed in this paper are not necessarily those of INTELSAT.

Amplitude/Frequency Response

The amplitude/frequency response of a voice-grade VF channel with respect to a reference frequency of 1 kHz should be better than the values given below.

Frequency Band	Loss Variation
500-2800 Hz	+0.5 to -0.5 dB
300-3000 Hz	+1.0 to -0.5 dB
3000-3400 Hz	+1.5 to -0.5 dB

Channel Weighted Noise

The hourly mean weighted noise level of an idle channel terminating in an earth station should not exceed -50 dBmOp.

Intelligible Crosstalk Between Adjacent Channels

Far End. The crosstalk coupling loss measured on all channels at the receive station, except the channel sending the 1-kHz test tone at 0 dBmO, should be greater than 65 dB.

Near End. The crosstalk coupling loss measured on all channels at the transmit station, except the channel sending the 1-kHz test tone at 0 dBmO, should be greater than 65 dB.

Delay Distortion

The voice channel group delay distortion characteristics are as follows:

1000-2600 Hz	170 μs
600-2600 Hz	500 μs
500-2800 Hz	1000 μs

The minimum group delay point in the voice frequency band should be regarded as the reference frequency.

Impulse Noise

Impulse noise counts for VF circuits should not be more than six counts per 15 minutes at a level of -21 dBmO.

Phase Jitter

The peak-to-peak phase jitter in any channel should not exceed 3°.

Performance objectives, measured between INTELSAT earth stations, for a 48-kHz group data channel are as follows:

Amplitude/Frequency Response

The amplitude/frequency response of a basic group (60-108 kHz) with respect to the 104.8-kHz group pilot frequency should result in a loss variation better than ±1.0 dB between 60.6 and 107.7 kHz.

Envelope Delay Distortion

Envelope delay distortion characteristics should not exceed the following values:

65.5-102.5 kHz	25 μs
64.2-65.5 kHz	65 μs
102.5-103.8 kHz	45 μs

The minimum group delay point in the basic group frequency band should be regarded as the reference frequency. The minimum group delay point should lie between 66 kHz and 102 kHz, and the ripple cycle should not exceed 2.5 kHz.

Intelligible Crosstalk Between Adjacent Groups

Far and near end crosstalk coupling loss should be better than 70 dB. The coupling loss should be measured for all groups in the baseband.

Impulse Noise

The impulse noise counts for a data group circuit (48 kHz) should not be more than 60 counts per 30 minutes at a level of -9 dBmO.

Measured Performance

Measured BER performance over satellite data circuits frequently includes the effects of terrestrial links to and from the satellite earth stations. The performance characteristics of these terrestrial links mask the satellite link performance. In order to generate a data base that describes typical satellite link performance, COMSAT has conducted data transmission tests through various earth station facilities using double-hop loops via operational carriers.

Table 1 lists BER measured performance and Table 2 lists the measured parameters of the circuits used in the BER tests. When comparing measured data with performance objectives cited previously, note that objectives are cited for single-hop circuits and performance is measured for double-hop circuits.

Measurements through operational circuits do not provide much insight into performance margins and parameter sensitivity. Therefore, additional measurements of BER and channel characteristics were conducted via test carriers using

Table 1. Data Transmission Performance Chart

Performance	Tx Station-Rx Station				
	Etam-Cayey Loop (3/71) SG2/G1/Ch11	Etam-Cayey[a] Loop (10/70) SG2/G1/Ch9&11	Andover-Goonhilly Loop, GA/Ch2	Jamesburg-Paumalu Loop, SG1/G3/Ch5	Jamesburg-Paumalu Loop, SG2/G3
Modem Type	Milgo 4400/48	Milgo 4400/48	Milgo 4400/48	Milgo 4400/48	GE Diginet 422
Data Rate (bps)	4800	4800	4800	4800	50 kbps
Test Duration (hr)	44.5	36.5	37	70	12
Data Level (dBmO)	-13	-13	-13	-17	-5
Channel Noise (dBmO)	-49.0	-48.0	-56.5	-59.0	-38
S/N (dB)	36.0	35.0	43.5	42	33.0
Number of Bit Errors	68	4380[a]	1162	36	
Bit Errors per Burst	8, 4	20-30[b]	1000, 79, 42, 14, 18		
Average Bit Error Rate	9×10^{-8}	7×10^{-6}	2×10^{-6}	3×10^{-8}	1×10^{-8}

[a] Etam-Cayey tests of 10/70 were degraded by AGR "hits," which were eliminated for 3/71 tests.

[b] Most errors were in bursts.

Table 2. Channel Parameter Performance Chart

Channel Parameter Performance	Tx Station-Rx Station			
	Etam-Cayey Loop, SG2/G1/Ch4	Andover-Goonhilly Loop, GA/Ch2	Jamesburg-Paumalu Loop, SG1/G3/Ch5	Jamesburg-Paumalu Loop, SG1/G3
Attenuation-Frequency Response (dB), 1000-Hz Reference				
500-2800 Hz	0.4	0.3		--*
300-3000 Hz	0.7	0.6		
3000-3400 Hz	1.8	0.9		
Envelope Delay Distortion (μs)				
1000-2600 Hz	160	370		--*
600-2600 Hz	1160	1100		
500-2800 Hz	1730	1400		
500-3000 Hz	1730	1400		
Impulse Noise Counts in 15 Minutes				
-22 dBmO	2.0	1.4	0.08	9
-26 dBmO	4.6	2.2		33
-30 dBmO	7.8	5.8		
Phase Jitter (deg)				
Peak-to-Peak	1.0	0.5	1.0	
Average	0.3	0.1		
Phase Hits (>15°) in 15 Minutes	1.0	0.7		
Amplitude Hits (>1 dB) in 15 Minutes	1.5	0.6		
Coincidence Hits	0.4	0.5		
Channel Noise (dBmO)	-49.0	-56.5	-59.0	-38.0

*Group band response: amplitude response 1.6 dB, 63-107 kHz; envelope delay 10 μs, 64-105 kHz.

equivalent noise loading to simulate adjacent traffic in the FDM/FM multiplex configurations. C/N ratios in the satellite channel were varied by reducing the up-link carrier power, while BER and channel characteristics were measured at different C/N conditions. Figures 1 through 4 present typical relationships.

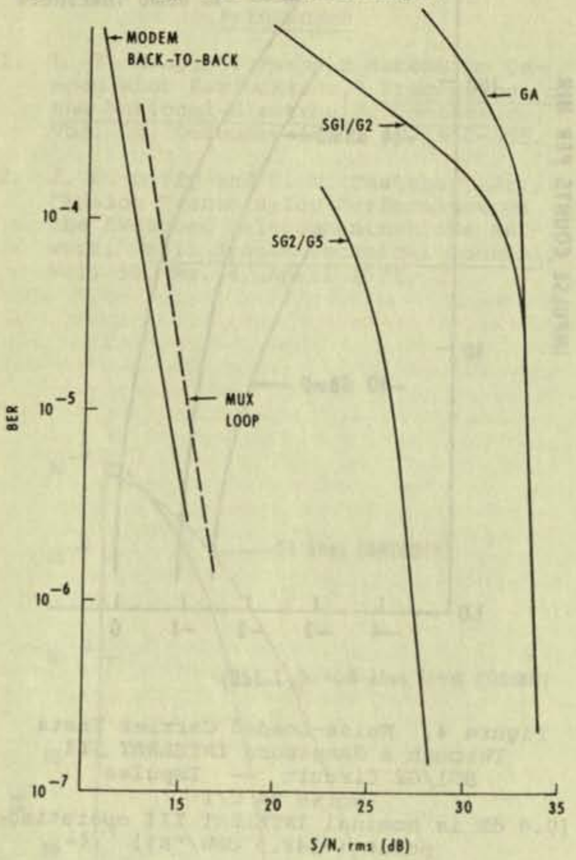


Figure 1. Noise-Loaded Carrier Tests Through a Jamesburg INTELSAT III Group Band Circuit — BER vs S/N (rms)

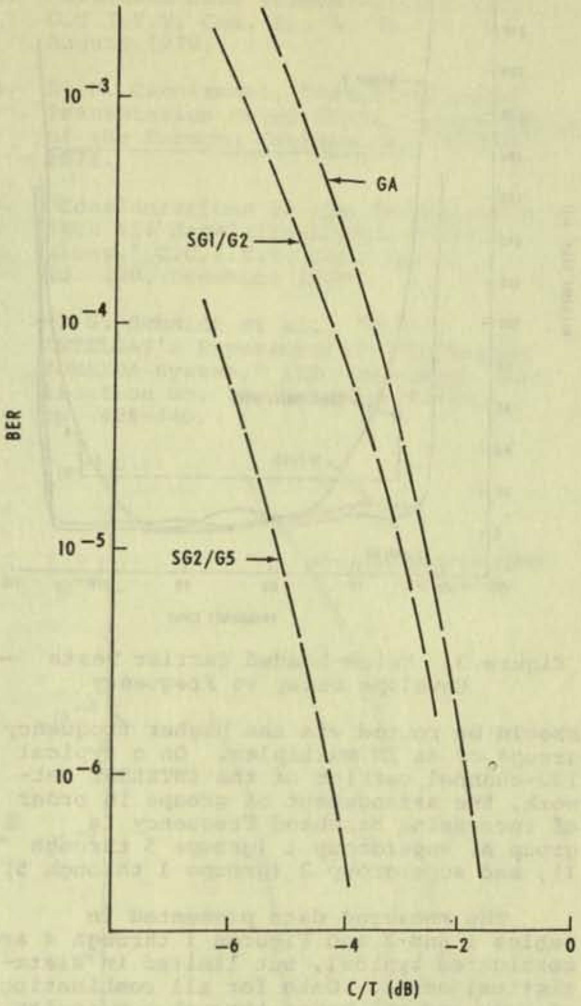


Figure 2. Noise-Loaded Carrier Tests Group Band Circuit — BER vs C/T [0.0 dB is nominal INTELSAT III operating point (-148.5 dBW/°K)]

Figure 1 shows that BER performance through any loop that includes earth station frequency modulators and demodulators is degraded significantly relative to back-to-back and multiplex loop performance. Although these data are for group band transmission, a similar effect is measured for voice band data as well. This performance degradation can be explained as follows. Threshold extension demodulators generate impulse noise when operating in the vicinity of threshold,[1] which is near the operating region selected to meet analog voice performance criteria.[1] The addition of impulse noise to the thermal noise spectrum produces the degraded performance. Hence, the popular conception of a satellite channel as a thermal noise channel is not valid if an FM demodulator operating close to threshold is included in the channel. Nevertheless, the resultant performance through the FM loops is still satisfactory at commercial carrier operating conditions since it exceeds terrestrial link performance. Also, the data presented in Figure 1 is for an INTELSAT III system. INTELSAT IV systems operate earth station FM demodulators several dB further from the FM threshold point than the INTELSAT III operating point. Therefore, INTELSAT IV data channels should have BER performance superior to INTELSAT III data channels.

Figures 1 through 3 support conclusions drawn previously from data transmission via terrestrial telephony multiplex configurations; i.e., group band data

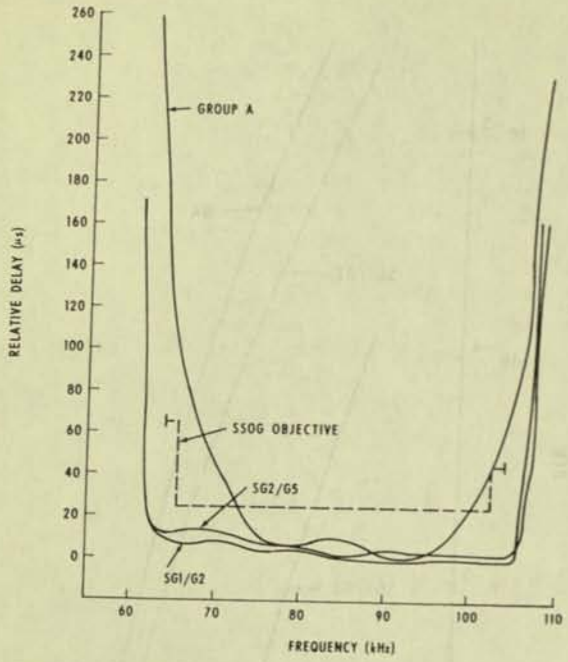


Figure 3. Noise-Loaded Carrier Tests — Envelope Delay vs Frequency

should be routed via the higher frequency groups of an FM multiplex. On a typical 132-channel carrier of the INTELSAT network, the arrangement of groups in order of increasing baseband frequency is group A, supergroup 1 (groups 5 through 1), and supergroup 2 (groups 1 through 5).

The measured data presented in Tables 1 and 2 and Figures 1 through 4 are considered typical, but limited in a statistical sense. Data for all combinations of busy and off hours through a majority of the INTELSAT earth stations are incomplete. Only two types of modems, four-phase PSK and vestigial sideband, have been tested to date. The determination of complete data channel performance statistics for the INTELSAT network is a continuing program at COMSAT Laboratories.

Future Performance

In both terrestrial and satellite data networks, contemporary commercial data transmission is limited by the analog telephony FDM multiplex configuration. Hence, existing satellite data circuits are adequate for this traffic. As separate terrestrial data transmission facilities become available, there should be a requirement for higher quality satellite data channels, however.

In 1972, a single-channel-per-carrier, four-phase PSK modulated system

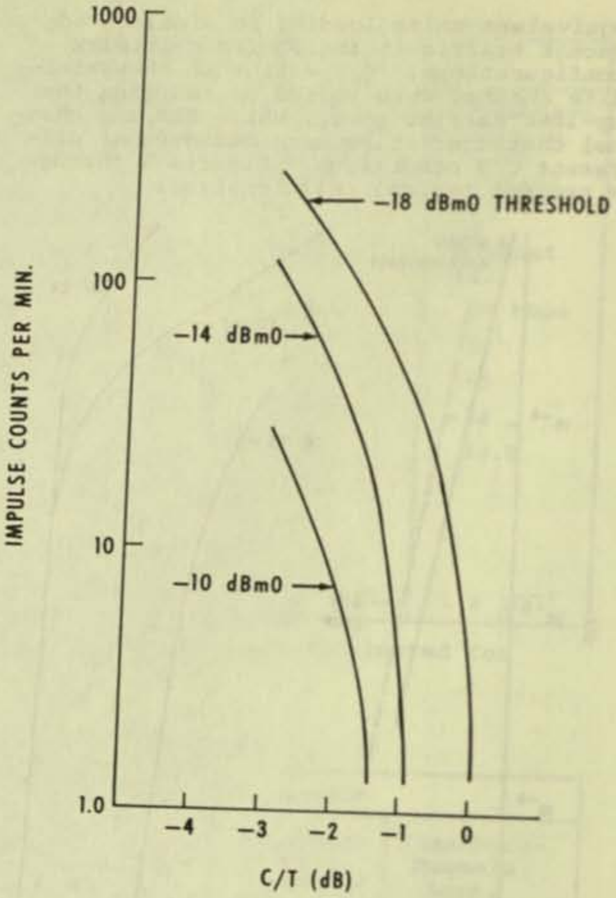


Figure 4. Noise-Loaded Carrier Tests Through a Jamesburg INTELSAT III SG1/G2 Circuit — Impulse Noise vs C/T [0.0 dB is nominal INTELSAT III operating point (-148.5 dBW/°K)]

will be introduced operationally in the INTELSAT network. This system can carry 56-kbps PCM encoded voice traffic or data at any rate from 150 bps TTY to 1.544 Mbps. Reference 5 describes its data handling capabilities. Figures 5 and 6, which were taken from this reference, indicate the range of BER performance which can be provided at different data rates by this single-channel-per-carrier service.

Since PSK modulation is used in this system, the impulse noise associated with FM will be circumvented. Also, since each channel will have its own carrier, the channel transmission parameters established to meet the data transmission criteria will be less constrained by voice transmission requirements.

In the more distant future, all-digital communications networks using time-division multiplexing (TDM) and time-division multiple access (TDMA)[6] should be established. The specifications for a TDMA prototype network are under development by the ICSC/T TDMA Working Group of INTELSAT. Field trials of an operational system tentatively are scheduled for 1976.

References

1. L. F. Gray, "Threshold Extension Demodulator Performance," Proceedings of the National Electronics Conference, Vol. 25, December 1969, pp. 550-555.

2. J. P. Duffy and T. W. Thatcher, Jr., "Analog Transmission Performance on the Switched Telecommunications Network," Bell System Technical Journal, Vol. 50, No. 4, April 1971.

3. "Wideband Data Transmission Tests," C.C.I.T.T. Com. Sp. A, No. 45, August 1970.

4. E. R. Cacciamani, "Satellite Data Transmission Using SPADE," Proceedings of the Eurocon Conference, October 1971.

5. "Considerations in the Transmission Data via Satellite-Linked Connections," C.C.I.T.T. Com. Sp. A, No. 120, December 1970.

6. W. G. Schmidt et al., "MAT-1: INTELSAT's Experimental 700-Channel TDMA/DA System," IEE Conference Publication No. 59, November 1969, pp. 428-440.

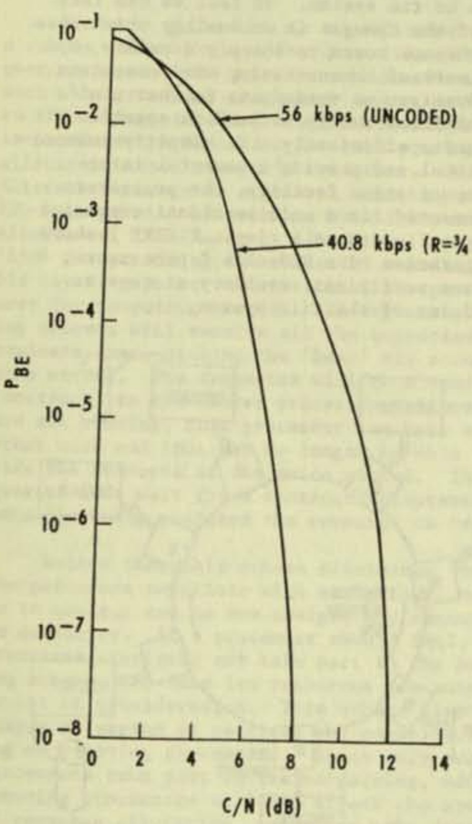


Figure 5. Single-Channel-per-Carrier INTELSAT IV Global Beam — BER vs C/N (56 kbps)

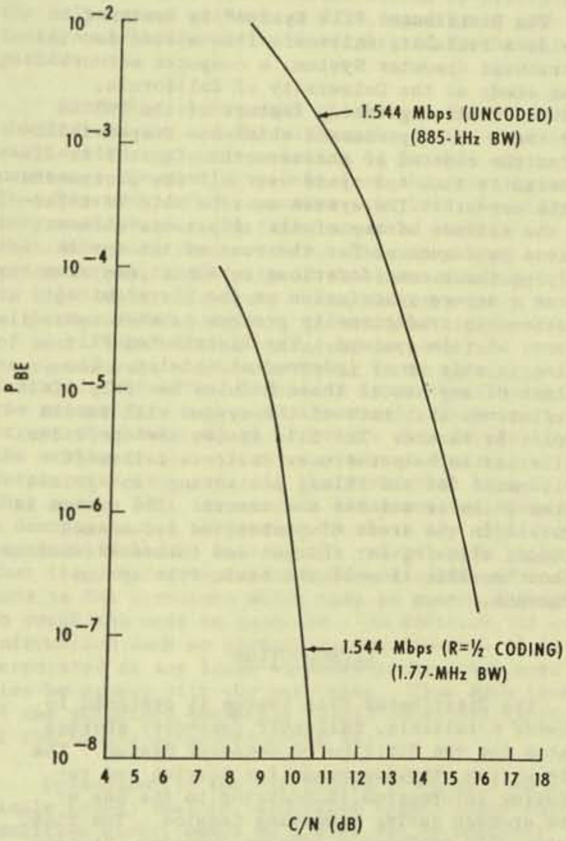


Figure 6. Single-Channel-per-Carrier INTELSAT IV Global Beam — BER vs C/N (1.544 Mbps)

THE STRUCTURE OF A DISTRIBUTED COMPUTER SYSTEM—THE DISTRIBUTED FILE SYSTEM

David J. Farber, Frank R. Heinrich
University of California
Irvine, California, U.S.A.

ABSTRACT

The Distributed File System* is designed to provide a reliable, fail-soft file system for the Distributed Computer System, a computer network under study at the University of California, Irvine. A distinguishing feature of the DCS is that there is no processor which has responsibility for the control of the network. Control is distributed in time and space over all the processors in the network. The system must be able to tolerate the failure of any of its components without serious consequences for the rest of the system. Applying these considerations to the file system places a severe restriction on the hierarchical relationship traditionally present between components of file systems. The Distributed File System is made up of independent modules. The failure of any one of these modules has only limited effect so that much of the system will remain completely intact. The file system also provides facilities to help the user create a fail-soft environment for his files, protecting him against system failures and his own errors. The system is flexible in the areas of protection and access methods, allowing for changes and future extensions without modification of the basic file system structure.

INTRODUCTION

The Distributed File System is designed to provide a reliable, fail-soft secondary storage system for the Distributed Computer System. The file system provides means for storing and retrieving information in addition to the use of core storage during a working session. The file system also performs a library function, storing information between sessions.

The fail-soft criteria implies a system which tolerates partial failure, yet is relatively secure from total failure. In addition, it should be relatively easy to recover from partial failure. If the system is to be fail-soft, the failure of either a processor or storage media device should only affect a portion of the files for a portion of the users. The individual user's environment should also be fail-soft. That is, if a user's file becomes permanently unavailable or damaged either through system failure or his own error, the work necessary to recreate that file should be minimized.

THE ENVIRONMENT - THE DCS

The Distributed Computer System (DCS) is a computer network under study at the University of California, Irvine. As a result of experience in our environment we are interested in developing a computer system which has fail-soft characteristics, will have low initial cost and allow easy expansion or contraction of capacity without much disturbance to the system. We feel we can take advantage of the changes in technology which have allowed processor costs to sharply decrease, by using a network of communicating mini-computers. An added advantage we anticipate is that mini's can be specialized easily to perform specific functions quite efficiently. To simplify communication protocol and provide a powerful interprocess communication facility, the processors will be connected via a unidirectional communications path configured as a ring. FIGURE 1 shows the configuration of a DCS with 7 processors, 4 of which have sufficient secondary storage to support modules of the file system.

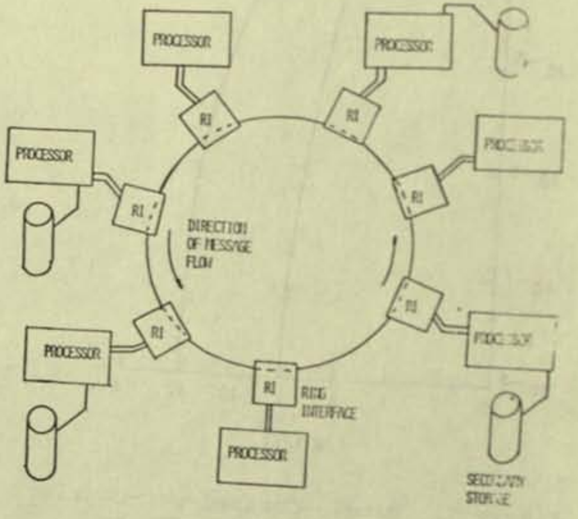


FIGURE 1 AN EXAMPLE DCS

Messages circulate about the ring, returning to where they originated after passing all other processors in the DCS. Since one of the goals is fail-safe operation, no one processor could be given the responsibility of control of the system. The failure of that processor would cripple the system. The cost constraints rule out redundent processors operating in standby status. Thus control will be distributed both in time and space. Control functions will move through the system,

---

*This work has been supported by the National Science Foundation under Grant GJ - 1045 residing on a processor which has little or no other load and moving on when that processor has other tasks to perform. Each processor will also perform many of its own control functions. The communication protocols and resource allocation are designed to eliminate the need for many of the control functions. We have taken a different approach to the communication system than is normally taken in computer networks. Messages on the ring are addressed to process names rather than physical hardware addresses, with the routing done by hardware. An originator of a message need know nothing about the topology of the network nor the physical location of the process he is sending to. In fact both originator and receiver can move without effect on the other. A device called a Ring-Interface (RI) connects the processors to the communication ring. The RI performs the routing as follows. The RI has stored in it the process names of all active processes in the attached processor. It compares the destination addresses of all messages that pass on the ring to those process names. If any messages are addressed to any of the active processes in that processor, the message is routed into the processor as well as being passed on down the ring.

Resource allocation is accomplished through a scheme called Request For Quotation (RFQ). A user or process who requires additional resources, such as a processor and memory space, broadcasts an RFQ message to all Resource Allocators. There is one Resource Allocator for each processor which allocates that processor's resources. All Resource Allocators recognize a general name, so a single RFQ message will be received by all Resource Allocators. The Resource Allocators reply with a 'bid' for providing the requested service. This bid may or may not reflect actual charges to a user for computing service. In any case, a requesting process will receive all the quotations and evaluate them, picking the 'best' bid according to some metric. The requestor will then send a 'contract' to the chosen processor. Since bids are not binding, that processor may have accepted other work and thus may no longer be able to provide the resource at the price quoted. The requestor must wait for a contract acceptance before actually being assigned the resource he requested.

Notice that this scheme eliminates 'control'. The processes negotiate with each other, neither is in control and no one assigns any resources as an authority. If a processor should fail, its processes simply do not take part in the bargaining scheme, and thus its resources are automatically out of consideration. This scheme also makes it simple to expand or contract the capacity by adding or removing processors. Since only active processors take part in the bargaining, adding or removing processors will not affect the operation of resource allocation, except to make resources more or less abundent.

THE STRUCTURE OF THE FILE SYSTEM

To provide the fail-soft characteristics, the file system is made up of modules distributed over several processors in the Distributed Computer System. These modules may reside in special purpose file processors or in more general purpose processors. Each processor should be autonomous and independent, capable of providing its portion of the file service without the assistance of any other particular process; any assistance necessary should be available from more than one source. The system accomplishes this with a minimum redundancy.

Models of the structure of many current and proposed file systems have described their structure as a hierarchy of directories. These directories limit context to allow partially specified file names. They also control access by providing the only access paths to a file, many times protection mechanisms being incorporated in these paths.

Since the system must be able to function with the loss of any module, either the system must be viable with any module missing or be able to reconstruct the missing module from the existing ones. In a hierarchy of directories, the loss of a directory would render all files accessed through that directory inaccessable. This is a high price to pay, especially if the lost directory is high in the hierarchy. The solution is to either provide redundant, distributed copies of each 'node' in the hierarchy or be able to reconstruct the hierarchy if nodes are lost. We reject the strict redundant approach as being too expensive (in use of resources). To reconstruct the hierarchical structure when faced with the arbitrary loss of nodes, it is necessary to retain with each node all the paths that lead to that node. That is, at each level must be stored a description of the structure above it. This can be accomplished by storing all global path names, that is, successive concatination of names of each node in the structure which must be passed through to reach the node in question. In addition, if any information such as protection requirements is incorporated at any level in these paths, that must also be stored with the path name. Thus each level of the structure becomes a single level description of the entire structure above it.

Consequently, the DCS System provides only a single level; that is, all files retain their fully qualified global names so a file can be uniquely identified and located regardless of the availability of other components of the file system. File sharing is accomplished without use of directories and protection is applied to the file itself rather than the access path to it.

There is a hierarchy of processes which assist in locating a file; however, the processes are expendable and if destroyed, can easily be recreated. These processes have access to data bases and provide information contained within those data bases. There is no access to this data except through these processes. In the following, the terms catalog process and volume process will refer to the process and its associated data base, since its external appearance is that of a process which answers requests for information.

FIGURE 2a shows an example of the logical structure the catalog system provides. FIGURE 2b is this same structure with a possible physical organization superimposed. The dotted lines indicate the limits of the physical modules. The logical links between these modules is provided through the communication system by sending messages addressed to module names.

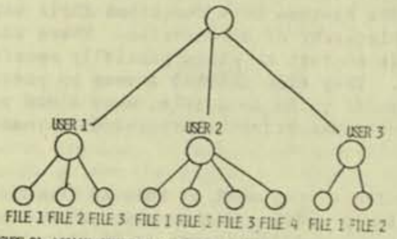


FIGURE 2A LOGICAL STRUCTURE OF THE FILE SYSTEM

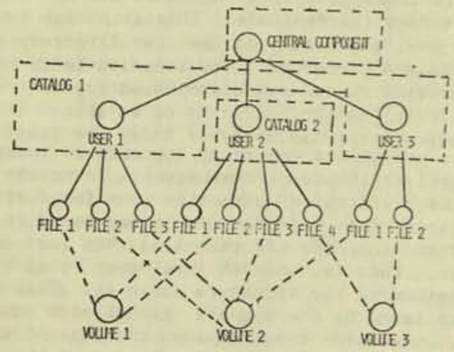


FIGURE 2B LOGICAL STRUCTURE OF THE FILE SYSTEM
DOTTED LINES SUPERIMPOSE A POSSIBLE PHYSICAL ORGANIZATION

The difference between the catalog and directory systems is subtle. A directory controls access and provides the only path by which a file is located. A file cannot be accessed except though the directory structure. On the other hand, the catalog system provides assistance in locating a file, but is in no way an essential part of the access path. Files can be located without use of a catalog by searching the volume table of contents. This will only be done in extreme cases, when the catalog has been damaged, and will probably only be done by those system programs which are responsible for restoring the catalog to a functioning state.

The system does have one central component which provides initial information to assist in locating a file. To prevent failure of this component from tying up the rest of the system, it is n-plexed on different processors, all n components being identical. It is keyed by owner name and for each owner provides the name of a process, a catalog, which references all files for that owner.

Several different catalogs exist on distinct processors. A catalog will serve more than one user, and each user will be served by only one catalog.

For each file, the catalog provides the name of another process, a volume, which is associated with the physical storage on which the file is located. A volume contains files for many users; the files of any one user are distributed over many volumes. A volume process runs in the processor to which the physical media containing the file is attached, and there is one volume process associated with each storage media device (e.g. disc pack or cartridge) which contains the table of contents for that device. The entry for each file in the table of contents includes the fully qualified global name, dates of creation and modification, protection information, the name of the catalog process which referenced that file, and the location (physical address) of the file header on the storage device. The fully qualified global name includes, in addition to the file name, the owners name and the generation number. Note that all this information is accessed through a process, although the data itself will be stored on the media in a predefined, reserved location. Device names and physical addresses are never used except by the volume process itself. The volume process will also manage the storage allocation on its device for creating new files and deleting old ones.

LOCATING A FILE

The following steps are the normal way of locating a file in the Distributed File System. Those steps will most reasonably be carried out by a process which is acting as an agent for the user at his terminal or for the user process which requires access to a file; however, there is no reason why the user or his process cannot carry them out.

Step 1: Broadcast a message to all the central components, specifying the owner name for the file that is desired. Accept a reply from the first one to respond. The reply should specify the catalog name which serves the owner of the file desired.

Step 2: Send a message to the catalog specifying the file name and owner of the file desired. This message is to a specific process; there will be only one process which can provide the information. The reply will specify a volume name which will give the final access to the file.

Step 3: Send a message to the volume (this is still a general message in the communication system, not a hardware address) specifying the file name and owner of the file that is desired.

Step 4: The volume will reply requesting verification for the protection specification in the table of contents. If the user is authorized to access the file in the way he has requested, the volume will request the operating system to create a process with a name which corresponds to the file name. The process name will be communicated to the requestor and all further traffic with the file will take place through this process. Thus files are addressed as general processes in the communication scheme. Files retain the same flexibility and mobility of general processes both in initially locating and in the transfer of data into and out of the file.

The details of the system structure for an example system are provided in FIGURE 3.

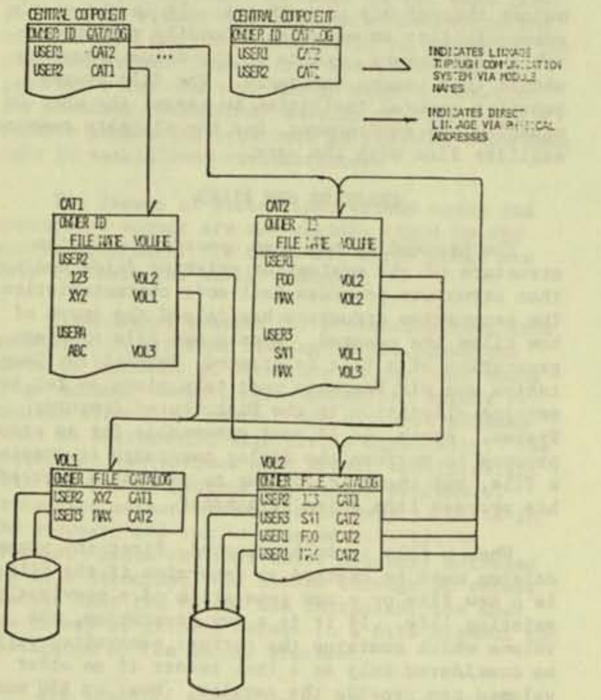


FIGURE 3. DETAIL OF A PORTION OF AN EXAMPLE CATALOG STRUCTURE (ONLY SOME LINKAGE PATHS ARE SHOWN)

FAIL SOFT CHARACTERISTICS - THE TOTAL SYSTEM

How does this structure provide the fail-soft characteristics claimed? Consider the possible failures that could affect the system.

(1) The storage media itself could fail (e.g. head crash on a disc pack).
(2) The device on which the media is mounted or the processor which controls that device may fail.
(3) The volume process may fail (either the software or the hardware of the processor which contains it).
(4) The catalog may fail (either software or hardware).
(5) The central component may fail (either software or hardware).

To recover from a failure of type (1) is nearly impossible. It would require that all volumes have a duplicate; however, we are trying to avoid this type of redundancy. It is possible that archival backup may be able to restore some or all of the lost information into another volume. All the files on the other volumes remain intact, the access to them in no way affected.

Failures of this type (2) are much easier to recover from. Assuming the media itself remains intact, all that is necessary is to move the media to another device or processor. Traffic with the files on that media and access to the information in the volume table of contents is only temporarily interrupted. When the media is on the other processor, the volume process is reactivated in that processor and the access processes are reactivated there also. Then access to the volume can begin again and the traffic with the files can resume. Since all addressing is to processes, the user processes are not affected by the change in location. It appears to them only a slight delay has occured.

This type of failure (3) is also fairly easy to contend with. All that is needed is to reestablish the volume process, possibly on the same processor if only a software error occured. Since the table of contents information is stored on the media in a predetermined location, it should be simple to establish a process to have access to that information. In this case, however, the access process for all the files in that volume that were active will probably be lost, effectively closing the file. Either a hardware or software error in a processor will probably destroy all information about what processes are active. So in this case users who were trying to open a file will gain that access with only a delay; however, the users who already opened files will have to reopen the file and continue where they left off. It is also possible a file that was in use may be damaged beyond the ability of the user to recover and the file will have to be recreated.

Failures of type (4) are less serious in consequences to users, but will probably take more time to recover from. The failure of a catalog will not directly affect any files, however it will make locating a file inconvenient. For all practical purposes those files owned by the users served by the missing catalog will be inaccessable for a short time. Recovery requires reconstructing the catalog information. This is easy to do but may take some time. All that is necessary is to read the table of contents for each volume, extracting the information for those files which were referenced by the missing catalog. A problem can occur if there is a volume missing at the time the catalog is reconstructed; however, when that volume is again made available, all the catalogs which are referenced in that table of contents can be checked to insure that they are up to date.

Failures of type (5) have virtually no effect. Since the central component is n-plexed on different processors, the failure of any one of them does not prevent access to the information contained in the others. It is possible there may be a slight degradation in response if the traffic with the other copies is high, but this is fairly insignificant.

The structure provides fail-soft behavior for the failure of any component of the file system, either software or hardware. Of course multiple failures of more than one component can cause much more chaos, but it is expected that the probability of simultaneous multiple failures is much lower than for single isolated failures.

FAIL-SOFT CHARACTERISTICS- THE INDIVIDUAL USER'S ENVIRONMENT

Providing fail-soft behavior for the individual user's environment implies that the loss of any file, either due to system failure or user error, should cause a minimum of lost information. It should be easy to restore the file to its current state without having to recreate it entirely from scratch.

Standard archival backup (i.e. file system dumps) does provide some measure of backup in this area. However it is often the case that the archival backup is too old to be of any real value. File system dumps are usually not taken more often than once a day, but it is usually the case, especially with program source files, that if any changes were made at all, there were several successive revisions within a short interval. Thus, the archival backup provides either an exact copy of the file needed or a previous version which may be so old as to have minimum value.

To provide backup facility with a finer increment, and thus a better chance of minimizing the work necessary to recreate the file, all files will have a generation structure. When a file is updated or a new version of a file is created it becomes the new current generation of that file. The old current generation still exists and is available as generation 'current minus one'. All previous generations are available, being accessed as generation 'current' minus some increment. It is possible to update or modify a file without creating a new generation, but the user assumes the risk involved in losing the incremental backup. Effort is made by the file system to insure that successive generations are on different volumes to prevent permanent loss of a volume from affecting the incremental backup.

To prevent the generation structure from becoming too extensive, the file system will provide an automatic facility for deleting files older than a certain number of generations, that number being specified for each file. Thus when a new generation is created, the oldest generation of that file will be deleted, providing that at least the minimum number of generations for that file exist. Also, users will be able to explicitly delete any generation.

To provide additional backup, and to guard the user against his own inadvertent errors, files which are deleted will not actually be removed from the file system. Files which are deleted will only be marked as deleteable and will remain in storage and in the catalog until the space is needed. At that time they will be expunged, that is, removed from the file system and the catalog. A file which is deleted but not expunged can be undeleted if the need should arise. Files which have been expunged can be reloaded from archival backup if that should be necessary. Users may expunge files explicitly when it is known that the file is of no value or if it contains information best not left in the system.

Thus the users environment is made fail-soft mainly through his own efforts. It is the user's responsibility to maintain an easily reconstructable environment, and to do the actual recovery should that become necessary. The file system provides several facilities to assist the user in managing his environment, but the ultimate responsibility lies with the user.

CREATING NEW FILES

The preceeding has been concerned with the structure of the system for existing files and how that structure provides fail-soft characteristics. The generation structure has raised the issue of how files are created. When a new file or a new generation of a file is created, Request For Quotation and Bid Response must take place as for any service allocation in the Distributed Computer System. Again, it is most reasonable for an agent process to perform the dialog necessary to create a file, but there is nothing to prevent a user or his process from doing it himself.

When a file is to be created, first the proper catalog must be checked to determine if the file is a new file or a new generation of a previously existing life. If it is a new generation, the volume which contains the current generation will be considered only as a last resort if no other volumes can provide the service. Next an RFQ must go out to all processors containing volumes. After a suitable delay, the bids which have been received are evaluated, the volume containing the current generation is only considered as a last resort. The confirmation and actual request for service then takes place with the chosen volume. The volume performs the necessary storage management, expunging deleted files if necessary, to provide space for the file. Then the volume requests the operating system to create an access process for the file and this access process becomes the channel for all traffic into the file. The volume must also request the appropriate catalog to update itself, reflecting the new file as the current generation. Thus the creation of a new file is carried out in essentially the same manner as any resource allocation in the Distributed Computer System.

SHARING, PROTECTION, AND ACCESS METHODS

Rather than create a new file, a user may wish to add to his catalog a file which already exists in another users catalog. This is not a redundant copy in the second users area, but is true structural sharing; the two file names reference the same file. Changes by one user are apparent to the other user regardless of the name they use, since the files actually refer to the same physical location on the storage media. Since the entries in the volume table of contents contain the fully qualified global name, including the owners name, as well as reference to the catalog which contains that file name, separate entries must be made in the volume table of contents for each file name, each pointing to the same header record on the storage media. This permits different global names to refer to the same file. However, if either user creates a new generation of that file the other will still be accessing the old generation. Thus a file may be the current generation for one user and an older generation for another. Structural sharing of files with a dynamic generation structure is probably best done only in specialized circumstances.

The issues of protection, access modes and concurrent access are not rigidly fixed by the system structure. In fact, all three areas are very flexible and easily changed.

Traditional schemes for protection vary from passwords to explicit listing of all permitted users to reliance on structure in the user names (e.g. account number, user number) to permit or exclude classes of users. Any of these schemes could be implemented with little difficulty. The protection conventions only affect the interface with the volume process and access processes. Protection mechanisms should be easy to build in the process, and can be changed with little trouble. Since the user never actually accesses the file directly, but always through an access process that the volume has initiated, the user has no way of gaining access to a file without the assistance of the volume process.

The types of access modes are also very flexible. The protection for granting initial access based on type of access requested is provided by the volume process as part of the general protection mechanism. To insure all transactions with the file are in accordance with this permission, the access process will check each transaction against the type of access permission granted. Although the system in no way fixes the types of access, the following is a reasonable set for general purpose use. (1) Update-includes read, write, execute (2) Read-includes read and execute (3) Execute-execute only (4) Append-allows only writing on end of existing file.

The access process also provides flexibility for the file's internal structure and the way the data is accessed. Since all access takes place through this process, the structure and access can be sophisticated as desired. It is anticipated that this flexibility will lead to the development of data independent access techniques. When a file is opened, the user will specify how he wants to access that file and regardless of how the file is formatted, the access process will make the data available in the format the user requests, doing any translation necessary. It is hoped that a procedural language for operating on data bases can be developed. The access procedure would then be an interpreter for the language, providing more generalized file operations than the simple read and write.

The access process provides additional flexibility which can be implemented if desired. The process can be used to resolve the problems associated with concurrent access by more than one user to the same file. This is generally no problem if all the users are reading the file; however if one user is modifying the file, the other users might be prevented access until the modification is complete. More sophisticated solutions to the problem exist and can be implemented without difficulty in the access process.

It would also be possible to allow the user to write prologues and/or epilogues for the access process supplied by the system. Different user supplied routines could be associated with each file, and would allow the user to create a more sophisticated environment. The user routine could be used to provide a sophisticated, interactive protection system much more extensive and flexible than the system supplies. Other possibilities include user supplied special access techniques and internal file structure which might not be available in early versions of the system.

SUMMARY

In summary, the Distributed File System is a modular, reliable fail-soft system for use in the Distributed Computer System. The system is made up of independent modules , the failure of any one of these modules has only limited effect, in fact much of the system will remain completely intact. The file system also provides facilities to help the user create a fail-soft environment for his files, protecting him against system failures and from his own errors. The system is flexible in the areas of protection and access methods, allowing for changes and future extensions without modification of the basic file system structure.

ACKNOWLEDGMENT

We wish to thank Julian Feldman, co-principal investigator of the Distributed Computer System, for his support. It is also appropriate at this point to acknowledge other members of the DCS staff, Ken Larson and Larry Rowe for the mutual interaction that has led to the development of the initial DCS design. Special thanks goes to Don Loomis, whose several detailed conversations on the file system suggested many of the ideas incorporated herein.

BIBLIOGRAPHY

1. D. J. Farber and K. Larson

   The Structure of a Distributed Computer System - Communications. Proceedings of the Symposium on Computer - Communications Networks and Teletraffic, Microwave Research Institute of Polytechnic Institute of Brooklyn, 1972.

2. D. J. Farber and K. Larson

   The Structure of a Distributed Computer System - Software. Proceedings of the Symposium on Computer - Communications Networks and Teletraffic, Microwave Research Institute of Polytechnic Institute of Brooklyn, 1972.

3. R. C. Daley and P. G. Neuman

   A General Purpose File System for Secondary Storage. Proceedings SJCC, 1965.

4. PDP-10 Reference Handbook Digital Equipment Corporation, 1970.

5. T. H. Myer and J. R. Barnaby

   Tenex Executive Language - Manual for Users Bolt, Beranek and Newman Incorporated, January 1971.

6. S. E. Madnick and J. W. Alsop

   A Modular Approach to File System Design Proceedings SJCC, 1969.

7. A. J. Collmeyer

   Database Management in Multi-Access Environment. Computer, November/December 1971.

8. R. C. Owens, Jr.

   Primary Access Control in Large Scale Time Shared Decision Systems. Project Mac Rpt Tr-89 July 1971.

THE USE OF DISTRIBUTED DATA BASES IN INFORMATION NETWORKS

Grayce M. Booth
Honeywell Information Systems, Phoenix, Arizona, U.S.A.

ABSTRACT

An overview of theories concerning distributed data base creation and use within a computer network is presented.

Alternative methods of creating a distributed data base are suggested, with notes on some of the problems associated with the more complex methods.

How to match up jobs with the correct files within the data base is discussed. Manual and software controlled methods are proposed.

The problems of allowing a single application to access files at more than one location are described. Several methods of surmounting these problems are given, with evaluations of where each method can best be used.

Finally, the problems of protecting file and system integrity when using a distributed data base are presented.

INTRODUCTION

This paper presents theoretical work concerning creation and use of distributed data bases, with emphasis on potential problems. Before these topics can be discussed, however, it is necessary to define what the term "distributed data base" means. Also, we must define an information network, which is the environment within which a distributed data base exists.

What is an information network? It is an association of elements including all or part of the following:

Information processors (data processing computers) – or IP's

Network processors (data communications computer) – or NP's

Distribution facilities (trunks, lines)

Terminals

Coupling devices

Concentrators, multiplexors, etc.

A distributed data base exists when an information network includes two or more information processors (or IP's), each of which has permanent – as contrasted to temporary or working – files attached. The permanent files attached to all IP's form a distributed data base, as shown in Figure 1.

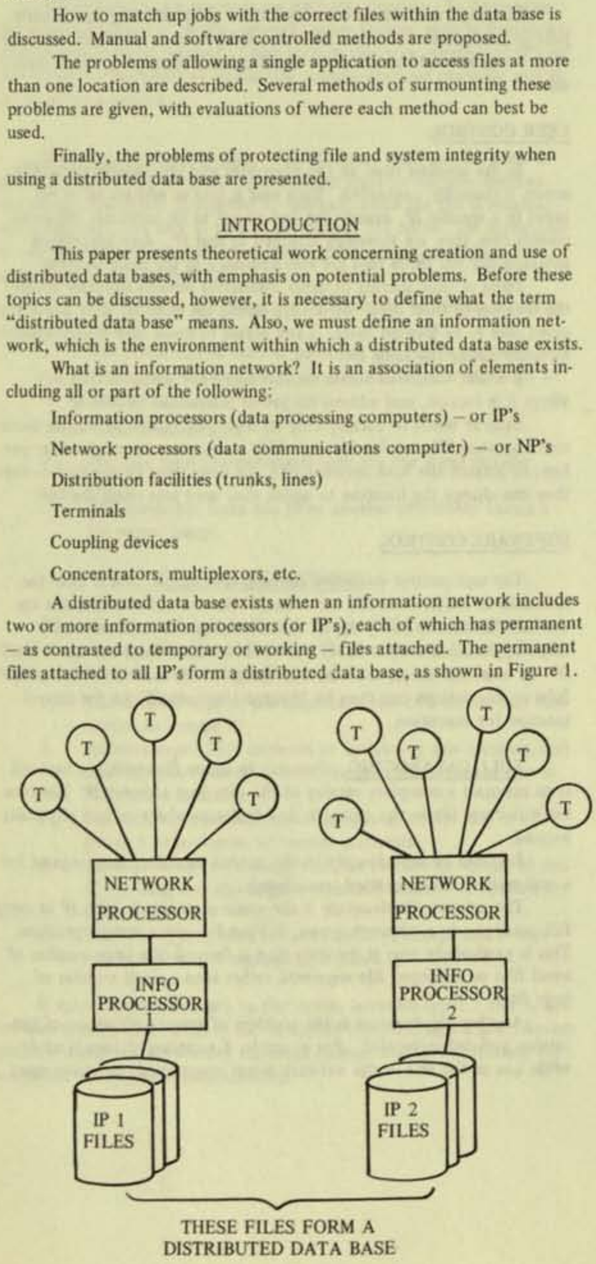


Figure 1. Typical Network

CREATING A DISTRIBUTED DATA BASE

A distributed data base can be created in any of several ways. If the data base consists of files created by individual users, and generally private to those users, its creation is implicit rather than explicit. In contrast, great care is taken when explicitly splitting an integrated file into distributed segments.

INDIVIDUAL FILES

The distributed data base which consists of individual files requires no special effort in creation. Each user creates his files at whichever IP he uses. In practice, most users access the closest IP. Files will therefore usually be located optimally in a geographic sense.

This type of distributed data base in general presents no problems. Methods of associating jobs with files in this type of data base are discussed later in this paper.

COMPANY FILES

A distributed data base which consists of company files is much more interesting to investigate. In practice, almost every information network requires some company files, as well as some users' private files.

Files which fall into this category include payroll records, order files, customer records, inventory records, and other similar data. These files do not belong to an individual user, but to the company, or to a department or division of the company.

Company files are created and controlled by a person or persons who can best be called the Data Base Manager (DBM). (This term is taken from the CODASYL Data Base Task Group Report.) The DBM is responsible for deciding file location, organization, format, and other similar factors. He may also decide who is privileged to access the file, and in what manner – inquiry, update, etc.

A distributed data base is made up of at least one, but more often multiple, files. "File" here is used to signify a logical entity; a collection of records logically related because of the owner's application(s). A data base consists of one or more files, grouped either because they are logically related or because they are all the property of a single owner – a company, government agency, or other owner.

Different methods of creating a distributed data base which includes company files are described in the following paragraphs.

GROUPING FILES. The multiple files which form a distributed data base can be separated into groups, and each group attached to one of the IP's in the information network. The method of grouping is most often geographic. Figure 2 shows an information network which spans the USA. One IP is located on the west coast, the other on the east coast.

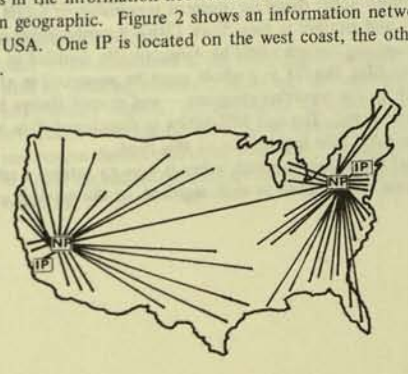


Figure 2. Nationwide Info Network

With this configuration, it is logical to attach files heavily used by east coast offices to the eastern IP, and attach those most used by west coast offices to the western IP. The DBM forecasts where heavy file use will originate, and causes the files to be placed there.

The main reason for geographic grouping is the cost of transmitting data to/from IP centers. In general, the shorter the transmission distance, the lower the cost. As new specialized carriers bring increasing competition into the information transmission field, line cost may become less significant. New technology will also lower communications costs. Because of these changes, processing centralization may become more attractive, since the cost penalties will be less than at present.

In the case of company files, as contrasted to private files, provision must be made for at least some use from outside of the file's area of location. In the example above, the east coast files will no doubt sometimes be accessed from west coast locations, and vice versa. In a true distributed data base such access must be allowed.

Another, less common, method of grouping files is by type. For example, all financial files might be at one location, all inventory files at another, and so on. This assumes some specialization of the IP's within the network. In this example, the financial files would be associated with the IP used for financial work.

These two methods of grouping files – geographically or by type – are the most common. Other methods applicable to a specific company could equally well be used. The important point is that each logical file is kept intact, but groups of files are distributed in the network.

DUPLICATING FILES. Within a distributed data base, it may be useful to duplicate some files at each of the IP's. Duplication may be needed for one of the following reasons:

- To increase access to the file by providing more paths to it, or more memory and/or processors capable of accessing the file.
- To provide rapid backup in case of the failure of a device, channel, or IP accessing the file.
- To decrease communications volume and/or dependencies on the communications facilities between IP's.

SPLITTING FILES. Still another method of creating a distributed data base is to split a logical file or files between IP's. Part of a logical file is then attached to one IP, part to a second IP, and so on.

We can ask why anyone would want to split a file. The most apparent answer is because it cannot be handled by one IP. It may be too big to be connected to a single IP, or it may be so active that its access rate cannot be supported by a single IP.

File size is not often a problem today, as mass storage technology improves. Large IP's today can support storage in the range of hundreds of billions of characters.

File access rate is more often a problem. Mass store designers seem to be building larger and larger capacity devices without increasing the device's access paths correspondingly. Load, therefore, is one reason why a logical file might be split between two or more IP's.

In this case there is no reason for the IP's to be physically remote from each other. If file activity is the only reason for file splitting, the processors can perfectly well be located in the same room.

Another reason for splitting a file is because accesses to different parts of it originate in different locations. For example, a file of all customer records for a company would form a logical entity. But customers on the west coast would originate orders from there, while east coast customers would originate their orders from the east. Communications economies might therefore, make it reasonable to locate one IP in Los Angeles and another in New York City. An even larger number of regional processing centers might be economically justified in some cases.

In split files, the file as a whole must be processed in at least some cases – perhaps for reporting purposes – and cannot always be treated in its separate parts. The real difficulties in distributed data base processing arise from split logical files of this type.

One of the problems which arises is how to catalog a split file. A separate name can be given to each segment, or the catalogs can show the same name for all segments, with some indication of the information included in each.

Although it is easy to postulate splitting a file into segments, there are a number of problems associated with doing so. Some of these problems are:

- Software Complexity

  Very complex software is required to handle a single logical file split across IP's. Consider, for example, an indexed sequential file split across two IP's. Where should the indices be located; in one IP, split between them, or duplicated? How can identical device addresses in both IP's be resolved in the indices? These are only two of many questions difficult to answer.
- Execution Overhead

  The complex software needed to handle situations of this type causes added overhead. Inter-IP messages to access data add overhead. Maintaining duplicate catalogs and/or accessing remote catalogs also increase overhead.

ASSOCIATING FILES WITH JOBS

In using a distributed data base, there must be a way to associate jobs or interactive executions with the file(s) needed. This can either be done under manual control or be handled automatically by the IP software.

USER CONTROL

In the simplest case, all file placement and job addressing for file access is manually controlled. Each user is told to address all of his input to a specific IP, usually the one closest to his terminal. When he creates files, they are therefore attached to that IP. Any jobs which subsequently use those files are also addressed to the same location.

Users of time sharing or other interactive services similarly control the location of their files. By always calling the same IP, they cause any saved files to be located there. Later access to saved files is accomplished by again calling the same IP.

If a user wishes to access another user's file, he must find out where it is located, and address his job or interaction to that IP.

Each IP needs to catalog only its own files. Little or no file movement between IP's can be allowed. If it is necessary to move a file, perhaps to balance the load between IP's, file users must be notified so that they can change the location to which they send jobs using the file.

SOFTWARE CONTROL

The user control described above is very simple, and is ideal for networks in which most file use is local. However, it may become unworkable if there is heavy use of files in multiple IP's accessed by a single application, and/or heavy use of files remote from the user's local IP. In cases like these, software knowledge of file location is necessary. Jobs or interactions can then be directed automatically to the correct location for execution.

FULL CATALOGING. One way to locate files within a network is to maintain a complete catalog of the data base at every IP. The catalog lists every file in the network, and indicates where each is physically located.

Jobs can be sent directly to the correct location. Any request for a non-existent file is spotted immediately.

The primary disadvantage is the space necessary in each IP to carry full catalogs. As a network grows, this can become a serious problem. This is particularly true if the data base is formed of a large number of small files or cataloged file segments, rather than a small number of large files.

Another disadvantage is the problem of keeping all copies of the catalog updated in parallel. For example, if a catalog change is made while one of the IP's in the network is not operational, provision must be made for recording the change later when the IP is again active.

The final disadvantage is that catalog search time for all files, including local files, may be increased by the presence of the catalogs for remote files.

INTER-IP INQUIRY. The alternative to full cataloging is to maintain only a local file catalog at each IP, and take special action if a request is received for a file not cataloged.

For a remote file, the IP software inquires within the network, and if the file exists, determines its location. If it cannot be located, the request is rejected. If the file is found, the requesting job is directed to the appropriate IP for execution.

This inter-IP inquiry may take any of several forms. All IP's can query a central IP which has a full catalog. Each IP can broadcast an inquiry to all other IP's, looking for a positive response. Each IP can pass the query to its neighbor, who can either respond affirmatively or pass the request on. If the query makes a complete circuit of the network and returns to the originating IP, the file is non-existent.

One problem is that a query may be unanswered if a remote IP is not operational. The inquiring processor may, therefore, be unable to determine whether a particular file is attached to the unavailable IP, or does not exist.

All of these query methods are perfectly feasible. To determine the most suitable in a specific network, many factors, such as frequency of non-local file use, number of files used in this way, number of IP's in the network, bandwidth of connecting communications channels, and so on, must be analyzed.

FILE NAME PREFIX USE. The file name itself can be made to indicate file location by prefixing a location code to the user assigned file name. The prefix is supplied by software when the file is created, and returned to the user as part of the official file name. He is required to use the prefixed name on all subsequent file access requests. The advantages of this method are:

- File location is obvious, and requires no processing to determine.
- The prefix can be used to manually determine file location, and so jobs can be sent to the correct IP either manually or automatically.

It has these disadvantages:

- The user must remember not only the file name but the prefix.
- File movement from one IP to another effectively causes a file name change.

FILE ACCESS METHODS

Once the assumption is made that a logical file can be split between two or more IP's, the problems of file access must be considered. Access requirements can be categorized as follows:

1. The simplest type of job requires access to a file or files connected to a single IP.
2. Another type of job retrieves or updates at one location, and also needs to retrieve a limited number of records at one or more other locations.
3. The third type is identical to number 2, except that it requires a large number of remote retrievals.
4. Another category covers jobs which retrieve or update at one IP location, and also require a limited number of updates in other location(s).
5. The final type is identical to number 4, except that a large number of remote updates is required.

It may seem unnecessary to distinguish between types 2 and 3, and between types 4 and 5, but in fact the differences are significant. Access methods which are satisfactory for small volume may be completely unacceptable for large volume situations.

Class number 1 presents no problems, so the remainder of this section will describe possible ways of providing the other types of access.

FILE TRANSMISSION

A complete copy of a file can be transmitted from its home location to another IP, so that it can be used by a job executing there. The decision to do this can be made either automatically by the operating system, or manually by the user with the file access requirement.

There are some obvious restrictions on the use of this technique. First, some limit must be set on the size of files which can be transmitted in this way. Second, if the user wishes to update the file copy which is transmitted to the remote location, care must be taken to prevent interference with any updates in the home location. This can be done by preventing update at the home IP until an updated copy is returned from the other location.

Transmission of a complete file copy has the value of simplicity. Once the file is received by the other IP, it can be cataloged normally, and the same access procedures used for all resident files can be applied to it.

Depending on the number of accesses required, this method may also reduce transmission time and overhead. Sending the entire file as a block is clearly more efficient than sending it record by record in response to individual requests. There is a break-even point at which access volume makes it less costly to send the entire file than to send individual items.

This technique allows access to multiple files, for either inquiry or update, as long as access requirements for the file allow it to be locked while remote update is in process.

SOFTWARE CONTROLLED ACCESS

The IP's operating system can provide automatic access to file data at other locations, on an individual request basis. The sequence of events to handle such an access would be roughly as follows:

- User program issues request for file data.
- The operating system discovers that the file is attached to another IP.
- An I/O request is sent to the appropriate IP, using a very high transmission priority.
- The other IP receives the request, processes it, and returns either the requested data or a denial.

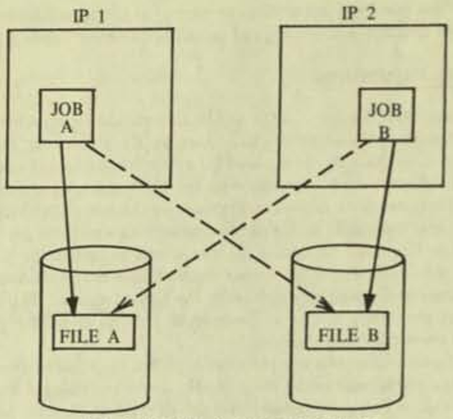
To avoid delay and interference, retrieve only is normally allowed, and the file may be modified by other jobs during retrieval. Data retrieved from multiple records in the file may, therefore, be inconsistent, due to concurrent updates taking place.

This restriction is made because all other access modes (update, exclusive retrieval, etc.) have the following potential problems:

- Interference

  Jobs in each IP must wait for data to be returned from other locations. In the worst case, these jobs (although inactive) might tie up enough system resources to hamper operation. If multiple concurrent update users of the same file are executing at different locations, extensive delays might occur. Deadly embrace conditions could be caused between jobs; each of two jobs could have files locked at a different location, and the two would be mutually hung waiting for access to the other's files. Figure 3 shows this condition. (On next page)
- Software Complexity

  Also to be considered is the complicated software required, with its resulting overhead, to control the multiple update and recovery aspects. An extremely complicated mechanism is required to recognize dependencies between jobs, detect and resolve deadly embraces, and control recovery of an abnormally terminated job which was updating files at multiple locations.



Job A has File A, and needs File B
Job B has File B, and needs File A

Figure 3. Deadly Embrace

To prevent misuse, accidental or intentional, of this mode, the user should specify, at time of file access request, the expected number of accesses. If none is specified, a default limit is applied. If accesses are attempted in excess of the limit, the job is terminated.

The idea that the operating system should provide automatic access to file data at other locations is attractive. Its most obvious advantage is simplicity from the user's point of view. In this mode, he need not worry about file separation or file location, as long as he requires only retrieve mode access.

An advantage as compared to file transmission is that only data actually accessed need be transmitted. If the access volume is below the point where transmission of the entire file becomes economical, or the file is too large to send, this mode saves transmission time and cost.

The major disadvantage is the complexity of the software required to support this mode. All necessary logic must be contained within the software system.

This type of access seems practical only where a limited number of remote inquiries is required. In all other cases the disadvantages outweigh the advantages.

APPLICATION JOB CHAINING

The following method can be used when one application requires access, particularly update access, to files attached to more than one IP.

The job which would otherwise update multiple files at multiple locations can be rearranged as a series of jobs chained together by means of a job activating scheme. Each job in such a chain can update only the files in a specific IP. As each job finishes, it activates the next job in the chain, generally in a different IP. This method requires that the analyst and/or programmer be aware of the exact file distribution among IP's.

As an example, the first job might access File A at IP #1, updating it and extracting data for reporting purposes. This data is included in a new job to be activated. This job is then directed to IP #2 for execution, where it updates Files B and C. It consolidates the information obtained during both updates. This sequence can be continued as many times as necessary to access all required files.

The last job in the chain completes any required reports, and sends them either to the originator, or to any other desired location(s). If the reports are to be printed, the final job can create and activate a reporting job in the originator's local IP, making the report available for printing there.

Using this technique, any type of access to the files at each location can be allowed. Since each job execution at each location is independent of the others, except for the passing of control, the problems of delay and interference noted under Software Controlled Access do not occur.

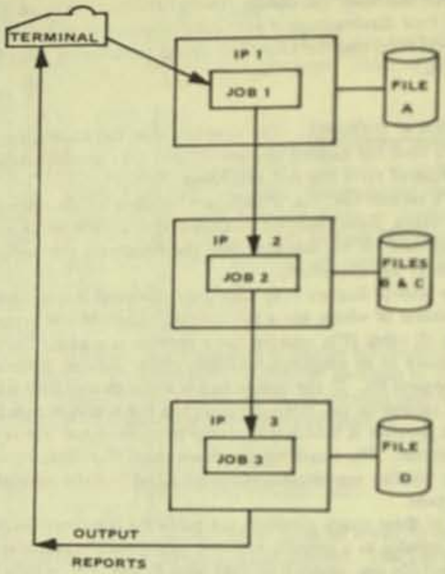


Figure 4. Job Chaining

The advantage of using job chaining to accomplish unlimited access in multiple IP's is that it provides great flexibility without causing any interference with other update users. From the operating system's point of view, it is extremely simple, since all complex logic becomes the user's problem.

There are several disadvantages to this approach. One is that the application designers/programmers must be aware of file distribution and take this into account in system design.

Second, each job in a chain is independent of its predecessors, once it is in execution. If an abnormal termination occurs somewhere in the chain, breaking the chain, it is difficult to remove the effects of preceding jobs. If, for example, a given multiple location update requires a chain of four jobs, and job number 3 terminates abnormally, the updates performed by the first two jobs remain in effect.

The termination notice resulting from a chained job can optionally be sent to the originator of the first job, who can then take appropriate action. This type of job chain requires careful design, so that a chain can be re-initiated at any point if an abnormal termination occurs.

Finally, there is no way to present an unchanging picture of data base status. The data retrieved by each job represents that part of the data base at a different point in time than the data retrieved by the other jobs. In contrast, an exclusive access job running in a single IP can obtain a completely static picture of the entire data base.

In spite of its disadvantages, this seems to be the most promising way to handle jobs which require update access to files in multiple IP's.

DELAYED AND SYNCHRONIZED UPDATING

Another mechanism which can be used to provide access to files at multiple location is delayed updating.

Delayed updating allows multiple jobs to concurrently retrieve from an apparently unchanging file. Changes made by each job are posted, not to the main file, but to a private change file. These private change files are inaccessible to any other jobs. Figure 5 shows two concurrently updating programs in the same IP. Since changes are inaccessible until they are applied to the main file, they cannot affect other jobs in concurrent execution.

Delayed update changes can be applied to the main file when each job finishes execution successfully, or more often if needed.

Multiple delayed update jobs can be run concurrently if there is a program available which can synchronize concurrent changes to the same record within the file. Such a synchronization program must of course

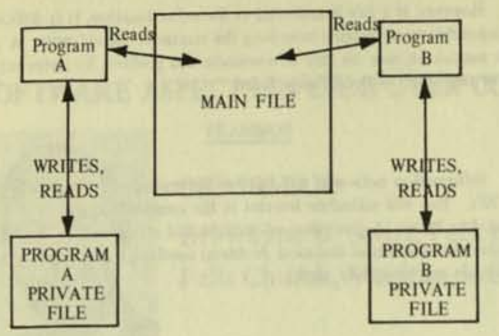


Figure 5. Delayed Update

be tailored to the format and conventions of the file type (indexed sequential, inverted, linked, etc.), and may require access to the definitions for the specific file.

As an example, a synchronization program to handle indexed sequential files would perform the following types of functions:

- Detect and delete duplicate creates or deletes of the same record.
- Detect and resolve changes to the same record. If to different fields, all are applied. If to the same field, differences are reconciled, if possible, and applied. Differences which cannot be reconciled cause one of the jobs involved to be rerun.
- Detect and resolve record creations which conflict. That is, two different records both inserted in the same available location. One of these records must be moved and the indices updated if necessary.

This is quite a complicated procedure, and may be impractical for complex file structures.

Delayed and synchronized updating can be used where duplicate file copies are maintained permanently at different locations. It can also be used when a file copy is temporarily transmitted to another IP and is updated there as well as at the original location.

This mode can also be used to allow updates to be performed in more than one IP simultaneously. Each processor would have assigned one or more record insertion areas, and each would converse with all other IP's before activating any job requiring a current file version, since this would trigger application of delayed updates.

Finally, synchronizing can be used, without delayed update, to allow simultaneous updating of multiple copies of a file, all of which are kept constantly current. Each such update requires a check throughout the network to see if synchronization is required, and a delay in further access until synchronization is accomplished, if needed. Update application can then be performed.

The ability to concurrently update one or multiple copies of the same file has many advantages. It there, at least theoretically, allows great flexibility in assigning file locations, duplicating files, etc. The interference associated with concurrent updates is reduced by confining it to the period of time necessary to actually apply updates, rather than extending it over the time of execution of all jobs involved.

Also, if updates are applied only on successful job completion, no rollback is ever required, since the main file is not changed.

It also has several disadvantages. One is that file accesses are increased, since both a pseudo and actual update access are required for every successful update. Additional overhead is caused by the necessity to search the user's change areas on each new access, to determine whether he should be given his private copy or the common copy of the requested record.

The use of delayed update also means that many file accessors will see a slightly out-of-date version of the file data. Whether or not this is significant depends on the applications involved.

Finally, the synchronization required to allow multiple concurrent delayed update users is an extremely complicated piece of software, particularly for complex file types.

The type of application for which this mode seems suited is transaction processing, where there are often numerous requests for a few accesses to a file, many of which do not involve updating. Using the methods described, rapid, economical, and reliable response can be provided by multiple copies of the file, each of which can be updated concurrently. Retrievals from each copy are delayed only while updates are being applied, but the possibility that a subsequent retrieval will encounter different values has to be tolerated. Multiple copies can be provided at each of several IP's, or the number of copies at each site can be adjusted to the volume of transactions expected there.

ACCESS TO DUPLICATED FILES

Unless controlled, simultaneous access by multiple programs to duplicated copies of the same file can cause differences between the copies. Depending on file content and use, it may be possible to minimize these problems by applying the following restrictions:

- Confine updating to only one location at a time, because:
    - Only personnel there have the authority or the knowledge required to change the file.
    - Certification and coordination of changes makes updating at one place desirable.
    - Volume of changes is so low in relation to retrievals that transmission of changes to a single location, and their posting there, can be tolerated.
- Allow retrievals from a version to which the most recent updates have not been applied.

INTEGRITY ASPECTS

In an information network, file and system integrity are extremely important. File integrity protection ensures that the data content of files, and file structures, are not in any way altered or destroyed inadvertently . System integrity protection similarly protects job data and the operating system. Operating systems within network-resident IP's must provide extensive integrity protection. Some integrity problems were discussed earlier. Two other important aspects of integrity are discussed here.

REMOTE FILE UPDATE

If an application updates files in more than one IP within a network, many integrity problems arise. File integrity protection, whether local or remote, generally takes one of two forms. It may involve journaling file changes, so that the file can be restored to a prior state if needed. Or it may involve delayed updating, in which changes are posted only after known to be correct.

Since file protection involves overhead costs, it is generally optional. Files whose content is not significant, or which can easily be recreated, can be classified as unprotected. Damage to these files must be manually repaired.

Consider the case of a chain of jobs updating protected files in three separate IP's. A job in IP #1 updates File A, then activates a job in IP #2 in order to update File B. This job completes, and activates a job in IP #3 to update File C. However, this third job terminates abnormally, causing the changes to File C to be discarded. File A and File B, however, remain in their updated state.

To properly rollback both File A and File B, other updates should not be allowed until File C is known to be successfully updated. In general, this is impractical, since it restricts file access too greatly.

INFORMATION INTERCHANGE CONTROL

System integrity is involved whenever a job in execution in one IP makes an access request to another location. Because system integrity is concerned with job/system status, it is responsible for keeping track of items such as outstanding remote requests. If any IP or communications link failure causes loss of contact between the IP's involved, action must be taken to re-establish contact and re-synchronize all processes.

Logic is, therefore, required in all IP's to handle remote access requests received just prior to a failure of the receiver, the sender, and/or the intervening communications link. Generally, the method adopted is to ignore received requests assuming that the requesting IP will take all necessary action.

The logic to keep track of outstanding remote requests becomes more complex as the requests take longer to service. In the simplest case, a request for one record, a timer is usually sufficient to check for failure to respond. If an answer is not received in the expected time interval, a failure can be assumed in the other IP or in the communications link.

However, if a job is activated at the other location, it is difficult or impossible to determine how long the requestor should wait. A periodic handshake may be able to overcome this problem by detecting that one or the other of the IP's has failed.

SUMMARY

Information networks will be used increasingly in the decade of the 70's. This will stimulate interest in the creation and use of distributed data bases. A great deal of thought and effort will be required to solve the substantial technical problems involved, before distributed data bases can be widely used.

SOFTWARE ASPECTS IN COMPUTER COMMUNICATION

CHAIRMAN:

Dr. Wayne B. Swift, Policyholder Service Corporation,
Falls Church, Virginia, U.S.A.

PROGRAM COORDINATOR:

Dr. Carl Hammer, Director,
Computer Sciences, UNIVAC,
Washington, D.C., U.S.A.

PAPERS:

Dr. Wayne B. Swift, Falls Church, Virginia, U.S.A., *"Software Aspects in Computer Communications"*

Professor J. S. Sobolewski, Assistant Professor of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington, U.S.A., *"Programmable Communication Processors"*

Monique Somia, Department of Scientific Development, IBM France, Paris-La Defense, France, *"The Approach of Software Problems in the SOC Experimental Computer Network"*

C. R. M. Singer, National Accounts Area Sales Manager, International Computers Limited, Harrow, Middlesex, England, *"The User Department and the Computer"*

Roy N. Freed, Widett & Widett, Boston, Massachusetts, U.S.A., *"Protection of Proprietary Software Programs in the United States"*

DISCUSSANTS:

Dr. Kurt Walk, Manager, Programming Center, IBM Austria, Vienna, Austria

Edward M. Ryan, Central Intelligence Agency, Washington, D.C., U.S.A.

Professor Franklin F. Kuo, Director, The ALOHA System, University of Hawaii, Honolulu, Hawaii, U.S.A.

THEME: Practical networks bring a variety of system-wide problems. Intended primarily for the overall manager of a network, this session discusses communications processor functions, intercomputer interface standards, effective user coupling and the legal protection of proprietary software.

SOFTWARE ASPECTS IN COMPUTER COMMUNICATION

Wayne B. Swift
Policyholder Service Corp., Falls Church, Virginia, U.S.A.

In this session, a series of papers will be presented which address several problems which are critically involved in the effective application of computers and communication to the service of the user.

The first paper deals with the architecture of a processor designed to facilitate the effective harnessing of communications to play its part in a computer/communication system while requiring a minimum of unnecessary struggle by the system implementers.

The other three papers concentrate more directly on software problems arising in connection with such systems. The first of these examines a set of software problems solved in the course of designing a major French computer/communication system.

The second paper considers how the computer can be made to assist in minimizing expensive and time-consuming human activity during the ongoing life of the system. Emphasis throughout is on the problems as seen by the user. The system discussed was developed in Great Britain.

Finally, a very interesting paper is given on the present status of legal protection for the programs and software components of such a system as we are concerned with.

In summary, a series of problems arising when practical computer communication systems are built is discussed. The practical point of view is stressed; the overall manager rather than the data processing technician will probably be the most interested reader.

PROGRAMMABLE COMMUNICATION PROCESSORS

J. S. Sobolewski
Washington State University
Pullman, Washington, U.S.A.

Abstract: Present day computer communication systems are complex and must be able to provide reliable service to a wide variety of users. Communication controllers must therefore be very flexible. In many cases these requirements can be better met by a general purpose controller that is program modifiable rather than by a hardwired controller. As a result, more and more modern communication controllers are built around a minicomputer. Such computers go by various names, perhaps the most general being a programmable communication processor. This paper describes the various applications, advantages and disadvantages of these processors. Desirable features such as tailored instructions, flexible input-output structure and special software for efficient data handling are also discussed.

1. Introduction

Minicomputers have had a remarkable effect on the computer industry in the last four or five years. The main reasons are their decreasing cost and increasing performance, which in some respects is better than that of their bigger brethren. As an example, the IBM 360/65 has a core cycle time of 750 nanoseconds and maximum data transfer rates of 1200 and 110 kilobytes on the selector and multiplexer channels respectively. For an Interdata Model 70 minicomputer, costing about $10,000 with 8 kilobytes of 1 microsecond cycle time memory, the respective transfer rates are 2,000 and 330 kilobytes. Although the 360/65 is far more powerful when the two are compared on the basis of standard operating software, instruction repertoire, range of peripherals supported and core size available, this example does show that at least where high speed input-output is needed, the minicomputer may well be more effective.

One of the areas where the effectiveness of minicomputers can be especially useful is in the area of data communication. Traditionally, most large computers are designed for scientific or business batch processing rather than teleprocessing systems which require real time interaction with a large number of remote terminal users. Such systems require error control, polling and control of communication lines. These tasks are routine but time consuming and can easily burden a large processor if a large number of terminals is involved. For this reason the large processors should be well isolated from these tasks. Some isolation is provided by the hardwired communication controllers made by the main frame manufacturers. In most cases, however, a programmable controller built around a minicomputer with an instruction set oriented towards high speed input-output and character manipulation can provide better and more flexible isolation. This has been recognized by minicomputer manufacturers for some time, but it is only recently that major computer manufacturers have announced such controllers or intentions of providing such controllers.

Programmable communication controllers are but one example of the growing use of minicomputers in data communication networks. There are some others discussed in this paper. First, the most common types of teleprocessing systems currently in use are briefly described. This is followed by a section on various types of minicomputer communication processors and the way in which they can solve problems in a teleprocessing environment. Their desirable hardware and software features are then given. Finally, the advantages and disadvantages of using such processors and some future trends are discussed.

2. Remote Terminal Data Processing Systems

In general, communication processors are used in on-line data processing systems with many remote terminals. These terminals may be located in the same building or thousands of miles away. They may also use various codes and rates for data transmission to and from the main computer. The data processing systems using remote terminals may be subdivided into five groups. These are:

a. Inquiry and response systems

b. Data collection systems

c. Conversational computing systems

d. Remote batch processing systems

e. Message switching systems.

The best examples of inquiry and response systems are the various reservation, automatic document retrieval and inventory control systems. Included in this category are information distribution systems which may broadcast information such as stock quotations or timetables to a select group of customers. In these applications the central computer system is used as a mass storage facility which can be accessed by a large number of remote users via the communication network.

In data collection systems, information from various terminals is entered and stored in the central computer. This data may be subsequently processed or it may be just for updating records used for inquiry and response systems. Examples of such systems are the reporting of transactions on the stock exchange, branch banking, weather reporting and keeping track of the status of factory production.

Conversational computing systems are designed to permit concurrently a dialogue between the central computer and many users. Programmers at remote terminals enter programs, one statement or command at a time, and wait until it is executed before submitting the next statement or command. Such systems are widely used in computer aided instruction, for example, where students at individual terminals study computer assigned material, perform assignments and answer computer posed questions. The answers can be corrected immediately and are used to guide further dialogue between student and computer depending upon the progress made. Included in this category are interactive graphics systems where computer supplied information is displayed on cathode ray tubes in response to inquiries from a keyboard or light pen. These systems are widely used in computer aided design and layout of electric circuits, integrated circuit masks, ship hulls, airplane wings and car bodies.

An essential feature of conversational computing systems is that they give immediate responses to the user. In remote batch processing systems on the other hand, response from the computer is not usually immediate. After receiving a program or data from a remote terminal, the host computer will schedule that job along with other jobs currently in the system. After execution, the output is transmitted back to the originating terminal. This could take a fraction of a second or several minutes. Batch processing systems are more practical than conversational computing systems whenever large programs or large volumes of data must be transmitted. These systems are used widely where the central computer has several remote batch processing stations (consisting of a line printer, card reader and card punch, for example). Sometimes the remote terminals may be other computers. In such cases batch systems may be used for back up or load sharing. In the former, the reliability of the overall computer network is increased while in the latter turn around time on an overloaded computer is decreased by sending jobs to a lightly loaded computer for execution.

Message switching systems are special cases or data collection and distribution systems. Here very little or no processing is done on the data. The main computer acts as a switching center, collecting data from and distributing data to various terminals. In many cases it is possible for several terminals to simultaneously send messages to the same destination. Some messages to that terminal must therefore be temporarily stored until the terminal is free to accept them. For this reason message switching systems are sometimes called store and forward systems.

The above categories only serve to summarize the general characteristics of teleprocessing systems. In many cases the distinction between them is fuzzy. Combination and variations are possible. Thus, monitoring systems are similar to data collection systems but the input is usually from some source other than a computer terminal (e.g. a transducer monitoring heart beat). Process control systems may be thought of as closed-loop monitoring systems which regulate an ongoing process.

3. Role of Minicomputers in Teleprocessing Systems

In all of these teleprocessing systems, a large number of various terminals are connected to the main computer through a communication network. The terminals and the line facilities must be compatible with the types of operation the system supports. The task of planning and designing an efficient data communication network can be exceedingly complex because of the wide variety of network requirements and the large number of solutions [1]. Some of the factors to consider in the design of the network are:

a. The type of teleprocessing systems used
b. Volume of data to be transmitted
c. Response time required
d. Error rate that may be tolerated
e. Number and geographical location of terminals
f. Terminal types and speed
g. Need for future expansion
h. Reliability and maintenance of the network
i. Network geometry to be used
j. Use private lines or dial-up lines

The objective is to satisfy all the requirements, at least cost. This not only includes the cost of the communication lines but also the cost of data sets and other auxiliary equipment such as multiplexers, concentrators and flexible controllers to improve the efficiency of the lines.

Traditionally, the multiplexing and concentrating functions have been performed by Frequency Division (FDM) and Time Division (TDM) multiplexers. However, the low cost of minicomputers, together with simple communication adapters, have made them very powerful tools in this area. The minicomputer, because of its memory which acts as a buffer, can multiplex and concentrate lines more efficiently than either FDM or TDM. Minicomputers used in this manner are usually called concentrators. Similarly, a minicomputer with its operating system and communication adapters may be interfaced to the main computer and perform the functions of a communication controller. Minicomputers used in this manner are usually called programmable communication controllers or simply communication front ends. It is these controllers that are making the greatest inroads. Consequently some main frame manufacturers have been forced to replace their hardwired controllers with the programmable kind.

Because little processing is usually needed in message switching systems, minicomputers also make excellent message switchers or controllers. It should be emphasized, however, that because of their programming flexibility, there is no reason why a minicomputer cannot serve dual purposes. Thus a minicomputer used primarily as a message switcher may also serve as a concentrator for some terminals. Figure 1 illustrates some application of minicomputers in data communications.

4. Concentrators

The cost of a data communication network can be reduced significantly by multiplexing or concentrating many low speed lines onto one or more high speed lines. In the past this was performed almost exclusively by FDM and TDM multiplexers. Because programmable concentrators can do this more efficiently, they have steadily been replacing the hardwired multiplexers in recent years. In practice, terminals do not receive or transmit data at their maximum speed over sustained periods of time. Since FDM and TDM's have no buffering, they must be capable of handling data under worst case conditions, that is at maximum terminal speed. Programmable concentrators on the other hand, because of their program flexibility and buffering capability, can pack the higher speed line to its maximum capacity. The available buffering can absorb peak loads while

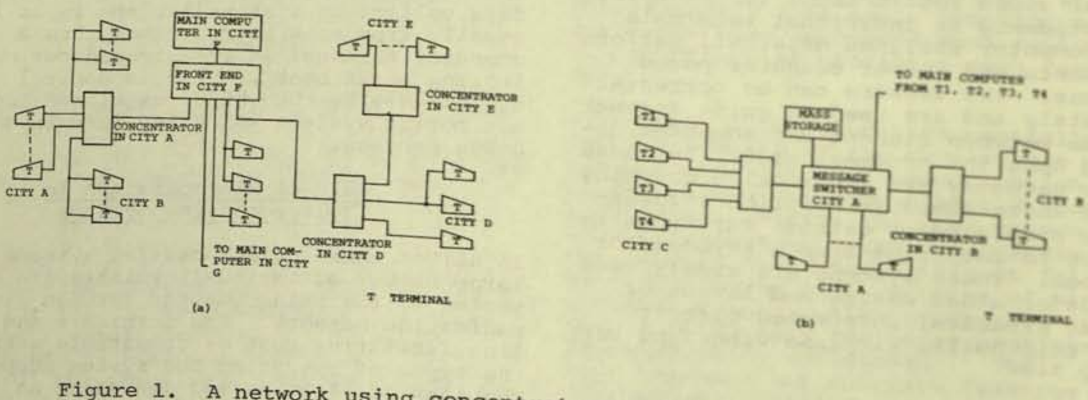


Figure 1. A network using concentrators and a programmable front end is shown in (a). In (b) a message controller also acts as a concentrator for terminals 1 through 4.

occasional sustained peak loads can be accommodated by software. This may be done by temporarily inhibiting transmission to some terminals or by sending a command to a terminal to reduce its transmission.

At present the cost of concentrators is higher than that of FDM or TDM multiplexers. In many cases this can be offset by the reduction in line costs due to the concentrator's greater efficiency. In addition, the concentrator has usually enough power to perform additional tasks such as line control, code conversion, line polling, adaptive speed control, error detection and error correction at no additional hardware cost [1-5]. Another advantage is that the concentrator can accept data from a much wider range of equipment than conventional FDM's or TDM's as shown in Figure 2.

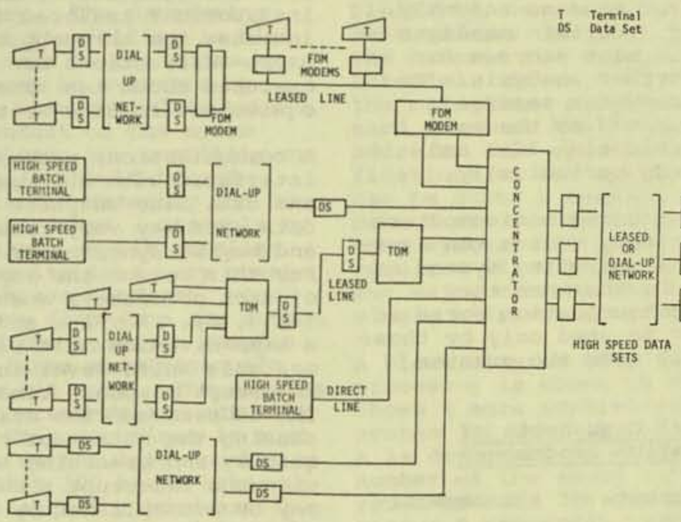


Figure 2. A general network illustrating the versatility of a concentrator. Data from the various terminals is concentrated and sent to the main facility via 3 medium speed lines.

5. Programmable Communication Controllers

Communication controllers provide the interface between the main computer and the data communication network. Programmable controllers are becoming more and more popular because they are usually lower in cost and can provide better performance [1-5] than hardwired controllers. Typical functions that they may provide include:

a. Flexible line and terminal control
b. Code conversion
c. Error control
d. Character and message assembly
e. Message editing
f. Recognition of control characters
g. Data buffering and queuing.

In addition, the more sophisticated programmable controllers may keep a record of all data communication traffic and perform message answering and recording. When hardwired controllers are used, many of the above functions must be performed by the main computer. Programmable front ends therefore may relieve the main computer of many of these routine but time consuming tasks. Furthermore, since most of the above are software functions, they may be easily changed without changing the operating system on the host computer. Changes on hardwired controllers on the other hand cannot be readily made.

Like the concentrator, the programmable front end must be able to accept data from a large variety of equipment. The basic difference between the two is that the front end has a direct interface to the main computer.

6. Message Switchers

A message switching system must have all the input-output and line control capabilities of a concentrator. In addition, it usually has some mass storage in the form of disk and/or magnetic tape in order to keep track of all transactions as shown in Figure 1(b). Otherwise, the essential difference between it and a concentrator or programmable front end is in the software because of the different function the system must perform. The processing on the incoming data is usually trivial, consisting of message assembly, analysis and routing. Certain messages may also be stored in mass storage for record keeping or further analysis. In such cases the minicomputer message switcher may be connected to the main computer via a communication line and transmit these records periodically.

The main advantage of using minicomputers in this area is that they reduce the overhead and simplify the operating system on the main computer. Furthermore they allow the existing communication ports on the host computer to be used only by those terminals that really need the greater processing power.

7. Essential Components of Communication Processors

The essential components of all communication processors are:

a. One or more processors
b. Flexible input-output channels
c. Line interface units
d. Interface to central computer in the case of programmable front ends.

The processor is a stored program computer with memory and data channels. The memory should be large enough to store the required software and provide adequate buffering for all lines. The memory cycle time and processor instruction execution time should be small to allow many lines to be serviced without the possibility of overruns. The processor's instruction set should be geared towards use in a communications environment where the main aim is the movement and manipulation of data [3, 6-8] rather than computation. There should be a powerful set of logical, bit manipulative, character moving, list processing and interrupt handling instructions. Also desirable are special instructions for code conversion and cyclic redundency check calculation. These may be provided by firmware [9, 10] or hardware [11]. As an example, the generator polynomial $x^{16}+x^{15}+x^{2}+1$ is often used for error detection in binary synchronous data communication channels [12]. Evaluation of the cyclic redundency check for this code by software on an Interdata 4 minicomputer requires about 180μsecs per character. Doing it by firmware (microcode) requires about 30μsecs while a suitable hardware implementation requires only 6μsecs and is limited by the time it takes to perform a write data instruction. Because a large number of these specialized instructions may be required to satisfy all potential customers, a dynamic control store is highly desirable. This would enable the customer to microprogram special instructions tailored for his own use and increase the throughput since microcoded macro-instructions can in practice be executed about ten times faster than an equivalent routine written in machine code.

A communications processor must be easily interfaced with a large number of terminals and data line adapters or modems. The data rate may vary from device to device and may be synchronous or asynchronous. For this reason the input-output structure of such processors must be very flexible [3, 7, 8]. For low and medium speed lines, a time division multiplexer channel with maskable multi-level interrupts and rapid interrupt response times is desirable. The address and the status of the device causing the interrupt should be available quickly and branching to a routine to service the interrupt should be rapid. This may be accomplished by automatic swapping of current and new Program Status Words (PSW). A PSW reflects the core location of the instruction to be executed, the conditions codes and the state of the interrupt masks. Several PSW's should be provided, one for each type of interrupt that may occur. For high speed lines (48,000 baud and above say) a special channel with direct access to memory is desirable. Such a channel should access memory on a cycle or instruction stealing basis and provide no interference to the processor once a transfer is initiated. For both the multiplexer and the high speed channels, the handshaking occurring over the bus should be as simple as possible so that the various line and terminal interfaces may be easily designed. A timeout should also be provided to prevent the channels from locking-up should an interface fail to respond. This is to ensure that terminals on the interfaces that are still functioning will continue to be serviced properly. The timeout should set a special interrupt to notify the programmer of the malfunction.

The line interface units link the channels with the terminals or the modems that terminate the communication lines. General purpose line interface units should be speed independent and handle synchronous or asynchronous transmission. This is not always possible in practice because of the wide variety of speeds (45-2400 baud for asynchronous and 600-50,000 baud for synchronous transmission). For this reason these interfaces are sometimes optimized for a given speed range to simplify hardware. The control of modems is usually done by the minicomputer software. The hardware involved inputs modem status and outputs control signals. The program senses the modem status, interprets it and issues an appropriate control command to the modem if necessary. The processing load is small and the flexibility is high because any changes made to the modem by the common carriers may be overcome by appropriate software changes. This makes the interface independent of the modem manufacturer and prolongs the usability or life of the system. It should be mentioned that the character assembly can be done by software or hardware. Assembly by software [4], also known as bit sampling or bit "banging," results in a very cheap interface as it minimizes hardware. One gate for input and one flip-flop for output are usually used and the computer samples groups of 8 or more lines simultaneously at a multiple of the bit rate. The samples are stored and the characters are actually assembled in core. This technique is most suitable for systems with a large number of low speed (less than 300 baud) asynchronous lines. Although the hardware is inexpensive, the software overhead is very high. This overhead is greatly reduced when hardware sampling is used. When a character has been assembled, an interrupt is generated. The computer services the interrupt and fetches the character. The hardware is more expensive but the software overhead is reduced significantly. This technique is used almost exclusively for the higher speed lines but with cheap integrated circuits readily available, this technique is becoming economical even for large numbers of low speed lines.

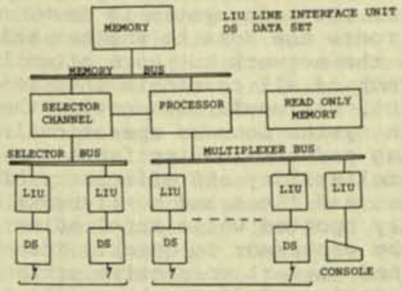


Figure 3. Block diagram of a simple communication processor.

Programmable front ends require an interface to the host computer. Here two approaches are possible; a multiple address or a single address interface. The multiple address interface is much more complex but has the advantage that little or no modification to the main operating system software is required because each address on the interface corresponds to a terminal. The single address interface is much simpler to design but has some serious drawbacks. On input, the front end must send not only the message, but some identification to determine where it came from. Similarly, on output the main processor must address the interface and include in the message the destination for which it is intended. Since most systems provided by the main frame manufacturers assume that each terminal has its own specified device address, this approach requires some modifications to the operating systems. This can be quite a complex task. Should the main frame manufacturer alter his operating system, little or no changes will be required if multiple addresses are used but major changes may be required on the single address interface.

A block diagram of a simple communication processor is shown in Figure 3. Figure 4 shows a more sophisticated dual processor system [4] used as a front end. Processor A is dedicated to controlling a large number of low speed lines and storing the received data in the shared memory. Processor B processes this data and passes it to the main computer. Similarly data from the main computer is placed in core and output by processor A. Core interference is small because processor A is microprogrammed to control the lines and hence accesses memory only to assemble or output characters. Medium speed lines are handled by processor B. The figure also shows special hardware for cyclic redundancy check calculation (CRC) and a communication oriented instruction set implemented in the read only memory of processor B.

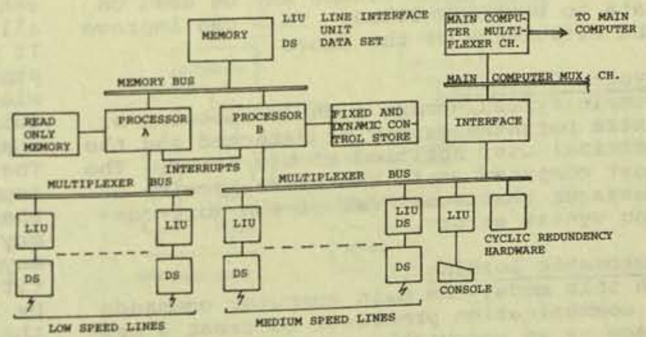


Figure 4. Block diagram of a more complex communication processor used as a programmable front end.

8. Advantages of Communication Processors

The major advantage of communication processors is their flexibility. They may be programmed to perform a wide range of functions some of which are not available on even the most sophisticated hardware controllers. Some of these functions include:

Line Polling and Control

This usually involves automatic dial-up, automatic answering and periodic polling of terminals to determine their status.

Adaptive Line Speed Control

Greatest network efficiency results when any line is adaptable to any speed. Here, the initial dialogue between the terminal and the computer allows the speed of the terminal and user identity to be determined. The software then sets the necessary parameters into the line interface and then allows data transfers.

Code Conversion

The front end or the concentrator may perform all code conversion between the various data transmission codes and the internal code of the main computer. This is a necessity in message switching systems where terminals with several different codes are used.

Message Assembly and Editing

Instead of sending each character to the main processor, whole messages may be assembled before being passed on. The messages may also be edited first. An example of this is when the user corrects his mistake by the use of the delete character. The communication processor corrects the character in error and ensures that only correct messages are passed on. The assembly and editing improve the efficiency of the channels and the main processor.

Data Compression

A data compression scheme may be used on data to be transmitted. This can improve the efficiency of the lines.

Syntax Checking

Simple syntax errors such as omission or extra parentheses may be detected and the terminal user notified of his error. The host computer would therefore receive messages that were free of the more common syntax errors.

Automatic Loading

In this mode, the main computer commands a communication processor to treat a message as an executable program. This allows the main computer to have absolute control over its entire communication network. Thus when software changes are needed on one or more communication processors in a network (to improve or add a new class of services for example), the main computer may load the new program via the network. At present this is used quite extensively only for the front ends. However automatic loading holds considerable promise for diagnosis of faults at remote locations [2].

Line Monitoring

An efficient network should be dynamic in nature. It should be constantly monitored and altered to ensure all portions of it are used efficiently. Monitoring can also isolate consistently bad lines and thus be an aid in fault location. Communication processors may keep a running record of all communication traffic including a count of errors. Analysis of these statistics can show up any inefficiencies and bottlenecks within the network.

Many other functions may be programmed, including buffering, recording, answering and routing of messages. All of these functions relieve the main computer and hence allow it more time for the tasks it was primarily intended for, data processing. It should be remembered however that there is a tradeoff. The more of these functions that are performed by the communication processor, the less lines it can handle.

The programming capability has some other far reaching advantages besides relief of the main computer. One of the more important ones is terminal independency. A terminal that is not supported by the main computer may be made to appear like a terminal that is supported by having appropriate software routines in the communication processor. This is usually used to interface a large variety of foreign peripherals to a large processor at low cost. Another not so obvious advantage is fail-soft capability. Here, the front end may send a message to the terminal saying that all or part of the main system is down. It may even route the data to another main processor in the network but more often it will keep track of all terminals that required service. It could then notify them when the main system becomes operational. The monitoring and autoloading functions improve the reliability and efficiency of the networks. Bad lines and bottlenecks may be quickly spotted while autoloading may enable the processor to quickly identify faults and suggest corrective action by running a series of diagnostic tests of the entire network.

Communication processors tend to be more economical than their hardwired equivalents on a cost-performance basis. They are flexible and can be adapted to support a large variety of local or remote terminals irrespective of speed, code, or control conventions. They allow easy growth of the network. Addition of a new terminal requires a line interface unit and a special module within the control program to service the terminal. The number of terminals that can be handled and the total data throughput is usually higher than for a hardwired controller. This expandability and flexibility increases the usable life of these processors and the network as a whole. The relief that they provide to the control processor enables the latter to also have a longer useful life before it needs to be upgraded to a more powerful model.

9. Disadvantages of Communication Processors

It should be obvious by now that communication processors have certain potential characteristics that make them very attractive. Unfortunately these systems have not been around long enough for their capabilities to be fully assessed or exploited. The teleprocessing software on the central processors is geared towards hardwired rather than programmable controllers. This prevents the full potential of the programmable controller to be realized. Even today most front ends simply emulate hardwired controllers because users are reluctant to change the operating system provided by the main frame manufacturer.

Another drawback is the lack of good software support from the manufacturers of communication processors. When these first came out several years ago, only an assembler was usually provided. It was up to the user to write his own software. Things have improved greatly but they still have a long way to go. The software offered today by the independent manufacturers is still quite basic. In most cases it is adequate enough to get the user off the ground but not much more. Manufacturers realize that their systems are capable of performing a wide variety of functions and that each customer wants something different. They are therefore unwilling to invest large sums of money into any software other than a basic scheduler, interrupt handler, various line control modules and emulators for some standard communication controllers like the IBM 2701, 2702 or 2703. The customer therefore still has a lot of programming to do to realize the full potential of his communication processors.

10. Other Uses of Communication Processors

Two other important uses of minicomputers that include data communication are in small time sharing systems and in systems where the minicomputer and main computer share large files. An increasing number of small users do not require a sophisticated time sharing system. Their needs may be served by a minicomputer supporting BASIC on eight to sixteen terminals. The organization of such systems varies greatly but one approach uses a system similar to that shown in Figure 4. Processor A handles all input-output and controls all terminals. Processor B has a disk or drum on its selector channel and does the compiling and executing [13]. Compute bound jobs requiring a larger computer may have access to one by providing a suitable link between it and the minicomputer. This may be via a data set or computer to computer interface.

Another method in which a minicomputer may reduce the overhead in a large machine is to allow it to share access to common disk files [14, 15]. Such a configuration is sometimes known as a delta connection [2] and is shown in Figure 5. The front end assembles messages and transfers them directly to disk. The connection between the multiplexer channels is used for passing control information such as the location and type of message stored on disk. With the delta connection, the central

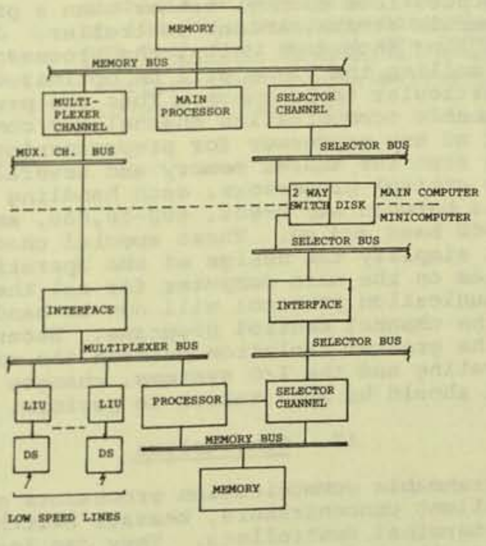


Figure 5. Example of a delta configuration.

computer is used to process the assembled messages only when required to do so by the front end. Another advantage is that messages can still be received and stored in the event of a malfunction of the central processor.

11. Future Trends

Communication processors of the future will be very similar to those today but will have a more efficient instruction set and a more flexible input-output structure. Dual processor configurations similar to that shown in Figure 4 will be quite common. The memory may be shared or there may be three memories, one for each of the two processors and one shared. This would minimize memory conflicts and greatly improve the throughput. The control programs for each processor would be in their own memory. As before, processor A could be dedicated for terminal and line control while processor B would process the data. The shared memory, which could be quite small, would be used to pass data between the two processors. The use of dynamic control storage would greatly increase the efficiency of such systems by allowing programmers to microprogram instructions tailored for their own use.

Such systems will be especially popular as front ends. The incoming messages will be assembled, edited, syntax checked and code converted before being passed to the main computer. The data path will be direct to the main memory rather than through a channel as is commonly done at present. In this respect we will have a programmable communications channel rather than a programmable communications controller. Just like line interface units, the processors controlling the lines will be optimized to a particular speed range. Thus the programmable communication channel may consist of one processor for preprocessing data from the shared memory and several line control processors, each handling all lines between say 0-600, 600-50,000, and 50,000 baud and up. These special channels will simplify the design of the operating system on the main computer for all the communication protocol will now be handled by the channel control programs. Because of the greater isolation between the main operating and the I/O systems, changes to both should be much easier to perform.

12. Conclusions

Programmable communication processors make excellent concentrators, message switchers and terminal controllers. They can increase the efficiency of the communication network and the central computer. Their flexibility and ever increasing performance/cost ratio ensures them a big future for years to come in the data communications field. Unfortunately they have not been around long enough for their full capabilities to be realized and assessed. This is partly due to the fact that the main frame manufacturers have in the past geared their machines and operating systems to hardwired controllers. The independent manufacturers have seized this opportunity to provide emulators for these controllers at great cost savings. This competition has already forced the main frame suppliers into rethinking their philosophy on hardwired controllers. The future will see major announcements on new programmable front ends neatly integrated with the main computer. The teleprocessing systems will reflect this change. They will be simpler and greatly improve the efficiency of the main frame.

References

[1] J. S. Sobolewski, L.T. Krebs, "Design of a Data Communication Network For a Regional Computing Center," Proceedings of the Fifth Australian Computer Conference, May 1972, Brisbane, Australia.

[2] C. J. Ball, "Communications and the Minicomputer," Computer, September, 1971.

[3] R. B. Hibbs, "Features of an Advanced Front-end CPU," Spring Joint Computer Conference, AFIPS Press, Vol. 38, 1971, pp. 15-21.

[4] H. B. Burner, R. P. Million, O. W. Rechard, J. S. Sobolewski, "A Programmable Data Concentrator for a Large Computing System," IEEE Transactions on Computers, Vol. C-18, Nov. 1969, pp. 1030-1038.

[5] F. E. Heart, R. E. Kahn, S. M. Ornstein, W. R. Crowther, D. C. Walden, "The Interface Message Processor for the ARPA Computer Network," Spring Joint Computer Conference, AFIPS Press, Vol. 36, 1970, pp. 551-567.

[6] W. F. Sherman, "Utilizing a Microprogrammed Processor in a Communications Environment," Proceedings of the 1971 Computer Designer's Conference, Jan., 1971, pp. 28-39.

[7] R. P. Holpern, "Daddy, What's a Communications Processor?" Proceedings of the 1971 Computer Designer's Conference, Jan. 1971, pp. 40-44.

[8] D. Wall, "A New Approach to Message Switching - The Sandac Communication Processor," Proceedings of the 1971 Computer Designer's Conference, Jan. 1971, pp. 45-47.

[9] Interdata Reference Manual, "Interdata 270X Telecommunication Control Unit for IBM 360/370."

[10] J. S. Sobolewski, "Error Control for Programmable Communication Controllers," to be published.

[11] M. A. Liccardo, "Polynomial Error Detecting Codes and their Implementation," Computer Design, Sept. 1971, pp. 53-59.

[12] J. L. Eisenkies, "Conventions for Digital Data Communication Design," IBM Systems Journal, Vol. 6, No. 4, 1967, pp. 267-302.

[13] H. B. Burner, "The Application of Automata to Multiple Level Computer System Design," Ph.D. Thesis, Washington State University, Pullman, Washington, 1972.

[14] F. Ives, "A Minicomputer Network for Interactive Data Management in a large Scale Batch Processing Environment," Ph.D. Thesis, Washington State University, Pullman, Washington, 1972.

[15] J. S. Sobolewski, "An Interface to Enable a Minicomputer to Simulate an IBM 2860 Selector Channel," ONR Technical Report WSU 1970-1, Contract N00014-68-A-0410-0001, Project NR 049-259.

THE APPROACH OF SOFTWARE PROBLEMS IN THE SOC EXPERIMENTAL COMPUTER NETWORK

Monique Somia
IBM-France
Paris-La Défense, France

Acknowledgment

The author wishes to acknowledge the contribution of the following people to the design and the implementation of the system described in this paper :
J.N. Guillon, C. Moireau from the CEA ; J. du Masle, Miss R. de Caluwe from IMAG ; J.P. Pouchès from CIRCE ; P. Ullmann from Ecole des Mines ; J.C. Coez, S. Girardi, J.P. Leclaire, H. de Wouters and Mrs M. Somia from IBM-France.

Summary

The SOC project is a research study, carried out by IBM-France and French Universities, specially oriented to developing and implementing a software system able to solve some of the problems inherent to Computer Networks. This project aims specially to simplify the use of Computer Networks facilities.
This paper presents :

- a brief overview of the general aspects of the SOC experimental network.
- the user oriented Network Command Language developed for the SOC System.
- the basic principles of the implemented SOC, System : the objectives of its first version, its general architecture, its functions, its operating mode...
- the Network Internal Language, developed for internal use by the SOC System.
- a brief description of the basic general functions : for example the Communication Interface.
- the foreseen extensions : for example, introduction of interactive mode in the Network System.

The implemented Network System and the description of the defined Language are particularly developed in this presentation.

General Purpose Computer Networks, as they are described for example in Mr. Elie's thesis(1), should allow users to access all the resources located on any computer of the network, and therefore to share them.
Thus with a network, the facilities available to a user are increased;with this increase, there is an increase in the complexity of accessing these facilities. Therefore a General Purpose Network must provide user aids of a practical and efficient nature to ameliorate these complexities just as operating systems provide the aids for user access to input/output devices.
The SOC (Système d'Ordinateurs Connectés,i.e. Connected Computers System) project is a software oriented research study headed by IBM-France and carried out with French Universities or Government Agencies.
This project aims to study and to implement a software particularly oriented to the use of Computer Networks and to test it through an experimental Computer Network, built with Participants'Computers, linked by telephone lines, and running only during test periods.
In this report, we will describe mainly the software facilities provided to the user of the SOC System, but for a more complete understanding, the solutions chosen for other aspects of the problem (like configuration telecommunication problems), will be quickly described in the next paragraph. More details about them can be found in the reports (2) and (3).

I - Environment

For Computer Networks designers, one of the main difficulties is to choose scientifically, the optimal configuration of the Network. In any case, a lot of economic, geographic, political or technical factors must be considered. So, for a general study, it is difficult to determine the most appropriate configuration ; as the general configuration supported by the SOC System, we chose the most flexible one, similar to the one chosen in the ARPA project (4) : the distributed configuration with an undtermined number of possible paths between any pair of computers, for example see figure 1. Such a configuration allows not only flexibility adaptable to specific needs, but also equivalence and complete reciprocity between each autonomous and independent computing System.
The particular configuration chosen for this experimental implementation is shown in figure 2. It must be emphasized, however, that the SOC software is not specialized to this configuration; rather it adapts itself to this configuration when it is initialized.
This configuration is made up of Standard PTT leased Lines (4800 bauds), Modems (IBM/3978),

and Transmission Control Units (IBM/2701). All of which are well understood and have been used for Remote Batch processing for several years.

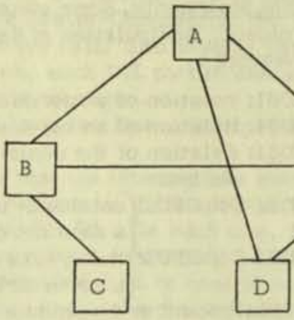


Fig. 1 - Example of general distributed configuration

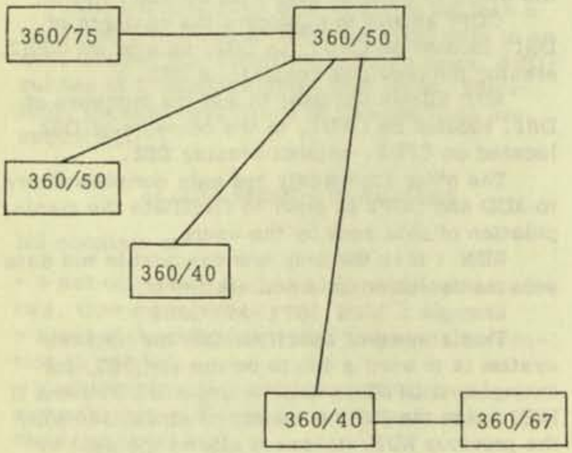


Fig. 2 - Experimental Configuration

The computers involved in the experimental network are :

360/75
2 x 360/40
1 x 360/67
2 x 360/50

In the first step of the project, all will run under OS-360 (MVT). In the second step, the 360/67 will run under CP-67 (10). More heterogeneity will be introduced if the first experiments are successful.

II - System Characteristics

Designers built the SOC System having in mind that it should be :

- first of all, easy to use by non-specialists.
- easy to implement under different local operating systems.
- independent in its conception, on a particular Operating System.
- dependent in its implementation on each Computing System.
- able to manage all the transactions in the network.
- able to use particular facilities.

So, the implemented system called NS (Network System) has the following characteristics :

- it is considered as an entity : the operating system for the network.
- a part of this entity will run on every computer in the Network.
- each part is implemented outside of the local Operating System, but uses the local facilities : for example, they are jobs under OS/360, virtual machines under CP/67...
- it supports a Network Command Language particularly adapted to Computer Networks, independent of any specific environment, and oriented to the requirements of the network users.

The main tasks of NS is to manage communications with :

- local users, by translating and carrying out Network Commands and returning answers.
- other computers, by sending and receiving messages for telecommunication and message switching system.
- local operating systems in order to facilitate the use of local resources.

III - The Network Command Language (NCL)

The Network Command Language, called NCL, which is derived from the language developed in the INTENET Project (6), has been studied and developed in the SOC project in order to ease the use of the Network facilities. A complete description of this language can be found in Miss de Caluwe's report (5).

The first version of NCL was designed so that the users can :

- transmit data sets from one computer to one a nother in the system.
- send a job to any remote computer in the network to be processed as if this job had been entered directly into the remote operating system.

The ambition of this version is very limited ; for example, it does not permit remote access to conversational subsystems. We plan to develop this extension and some others in next steps of the project. However, the two possibilities, implemented in the first step, already allow users to access most of the resources in the system.

NCL Description

NCL is an high-level language which includes two kinds of statements : declarative statements and executable statements.

Declarative Statements. Every variable must be declared and initialized before being used in any executable statement.

Variables may have one of the following types : CPU - UNIT - VOLUME - DATA SET - MEMBER (of a Partitioned Data set) - INPUT STREAM - and characteristics of data sets (like SPACE, organization, format of records...) .

So the users will be able to manipulate more easily data set or input stream through the network in executable statements.

Declarations relative to data sets manipulation :

examples :

```
CPU  A : = IMAG67 , B : = CIRCE75;
UNIT  U1 : = 2314/A , U2 : = 2314/CIRCE75;
VOLUME V1 : = MFT111/U1,
       V2 : = MFT222/2314/CIRCE75;
DSN  D1 : = SOC.OBLIB/V1,
     D2 : = SOC.OBLIB/MFT222/U2;
MEMBER  M1 : = MAIN/D1;
```

The types of declared variables are pointed out by the keyword of the declarative statements. The identificators not declared in a previous statement, and used on the right-side of : =, are considered like constants of the system.

Declaration of input stream :

example :

```
   INP      ALPHA : = '///*'
            Data cards
   ///*
```

This declaration allows to assign a name (ALPHA) to the data, written on the cards located between the INP statement and the mark of 'end of data'.

Any data may be written on the cards ; only a non-ambiguity with the mark of 'end of data' is necessary ; this mark is fixed by the user himself and indicated between quotes in the INP statement in order to avoid any confusion with particular data.

The INPUT STREAM declaration is useful to enter data which can be either stored into a data set, somewhere in the network, or used as the Control Language necessary for a job on a particular local operating system.

The example (Annexe 1) shows the declaration of the necessary job stream to compile and execute a PL/1 program on an OS/360 machine.

Executable Statements. Some statements have a main object, manipulation of data sets through the network :

```
CREATE     DS1; creation of a new data set.
DELETE     DS1; deletion of an old one.
EMPTY      DS1; deletion of the contents of an
                old one.
CATALG     DS1 ON CPU1;⎧catalog or uncatalog
                       ⎪a data set on
UNCATALG   DS1 ON CPU1;⎪the system catalog of
                       ⎩a given CPU
MOUNT      VOL1;⎫mount or dismount a
DISMOUNT   VOL1;⎭private volume.
COPY       DS1 TO DS2;
ADD        DS1 TO DS2;
```

The most important statements in this group are COPY and ADD that give the possibilities to the users to transfer data sets between CPUS.

COPY allows to duplicate the contents of DS1, located on CPU1, in DS2, located on CPU2, erasing the previous contents of DS2.

ADD allows the user to add the contents of DS1, located on CPU1, to the contents of DS2, located on CPU2, without erasing DS2.

The other statements are only complementary to ADD and COPY in order to facilitate the manipulation of data sets by the users.

RUN : it is the only one executable not data sets manipulation-oriented statement :

Example : RUN INP1 AT CPU2 ;

This statement specifies that the network system is to send a job to be run at CPU2; for example, if CPU2 is running under OS/360 and if INP1 holds the JCL necessary to an OS/360 job, the previous RUN statement allows the user to ask for the execution of this OS job at CPU2.

NCL General Aspects

We have seen above that NCL allows users to transfer data sets and to send jobs from a CPU to an other one, through simple statements.

A set of commands, required by a user can be joined together to form a Network Job.

A Network Job has the following characteristics :

- it belongs to a particular user
- it begins by a NETIN statement which includes the user's name and eventually accounting informations.
- it ends by a NETOUT statement.

Network Job Characteristics. The Network Job is a very important concept because it is the only way by which the NS can describe the user's needs to each of the different computing systems.

The life of a Network Job includes several steps :

- first, the Network Job is read and compiled by the Reader-Compiler to the Network Internal Language (NIL) into several parts.
- then, each NIL part is executed on the appropriate computer.
- at last, the Network Job enters a step of print.

Note that the intermediate step requires as many executions as there are computers involved in the Network Job. On each one, NIL commands require only local processing, except when data must be received from or sent to another computer ; in this case, a reciprocal command is generated in the NIL part corresponding to the associated computer ; the interpretation of this command in synchronization with the interpretation of the first one, allows to execute easily the complete function; the main problem is the synchronization.

The interpretation of a Network Job take a very long time because there are many data to be sent through the telecommunication lines, difficulties of synchronization, and so on. Therefore, parallel executions of Network Jobs are essential.

IV - NS General Architecture

NS consists of :

- a set of service routines, for storage, resources, time management
- a set of functions in which sequential execution is needed
- a multitasking supervisor which allows to execute several functions at the same time, on the same computer.

A task executed on one computer under NS is called a process, in order to distinguish it from an OS/MVT task, for example.

Network Process Management. A process can ask for the creation of an other process ; but a process will disappear only when it asks itself "to die".

The creation and the death of a process is executed by service routines.

The Dispatcher gives the CPU control to the different processes by the use of the Process Control Block (PCB) associated with each one.

A process can enter a Wait State (not dispatchable) if it asks itself for it ; but it cannot be interrupted by an other process of the system.

Two main classes of process exist :

1° - System processes necessary for the running of the NS, which are always resident in main storage. Some of these processes are, for example :

- the communication interface which handles communications between computing systems involved in the network.
- the Network console services which handles communications with the local operator.
- the Network Initiator-Terminator which handles communications with the local operating system.
- the Network Job Initiator which starts processes needed by the Networks jobs.
- the Logger which handles particular commands received from other CPU.

2° - User Processes which are started when they are needed, to handle users'Network Jobs. There are four kinds of User's processes, each of them performing a specific function :

- the Reader-Compiler which reads Network Jobs from a local card-reader (or from RJE Terminal), creates the blocks which describe the Network Jobs for the NS and translates them into Network Internal Language.
- the Internal Reader which reads the Network Internal Language relative to a Network Job sent from an other computer.
- the Interpreter which executes the NIL commands relative to a particular Network Job, on a particular computer.
- the Writer which prints the results.

One and only one system process of each kind must always exist on each computer when NS is running. Any number of user processes can exist on each computer, at the same time, provided that enough storage is available.(Fig. 3).

Network Job Management. The main functions that the NS performs for users are :

- to translate network jobs.
- to transmit commands included in a network job to the specified computing systems.
- to ensure the execution of these commands
- to put out the results of a network job on on the CPU required by the user.

When the Reader-Compiler finds a new Network Job, in the input stream, it creates an entry in the Permanent Network Job Control Table (PNJCT). This entry will be destroyed only when the execution of the job has ended and the output printed or punched.

After the creation of a new entry in the PNJCT of the 'local' CPU, the NCL commands are compiled into sets of NIL commands which will effect the job. When compilation is completed the Network Job Initiator causes an Interpretor on the local CPU to be activitated and to start processing the set of NIL commands for the local CPU.

One of the first NIL commands is a DJC command (request for a correspondent job) ; its interpretation will create a PNJCT entry for this job on a remote computer. The remote Network Job Initiator will start an Internal Reader for this job which, after synchronization with the interpreter on the source computer, will receive from it its NIL commands for this job.

Such operations must be repeated as many times as there are computers involved in this particular Network Job. When the sending of each part of the NIL has been done by the use of an interpreter on the source computer and by the use of an Internal Reader on each of the remote computers, then, on each of these, the Network Job Initiator will be able to start an Interpreter. The communications between computers will be negotiated through the interpretation of two complementary functions, generated in the same compilation, and interpreted separately but synchronously, on the two computers.

V - The Network Internal Language

The Network Internal Language, generated by the Compiler for a Network Job, is divided into three parts :

1 - General Information . The general information area includes :

- the total length of the generated NIL program.
- a pointer to the next step to be executed ; the interpretation of a job can be interrupted for some time ; therefore the interpreter must know at which step it must restart .
- a pointer to the step to be executed if an abnormal end occurs in the interpretation of any NIL command.
- a byte giving the level of the errors which will produce an abnormal end.

2 - Data. The data part includes the blocks associated with variables declared by the user in his Network Job.

The type of the variable described by a block is identified by the first byte of the block which contains a type identificator. The format of the block depends of the type described, but for a given type, it is determined and contains a complete description of a variable of this type.

3 - NIL Executable Commands. A NIL command has the following format :

- an action code (1 byte)
- the instruction length (1 byte)
- parameters for this command which can be:

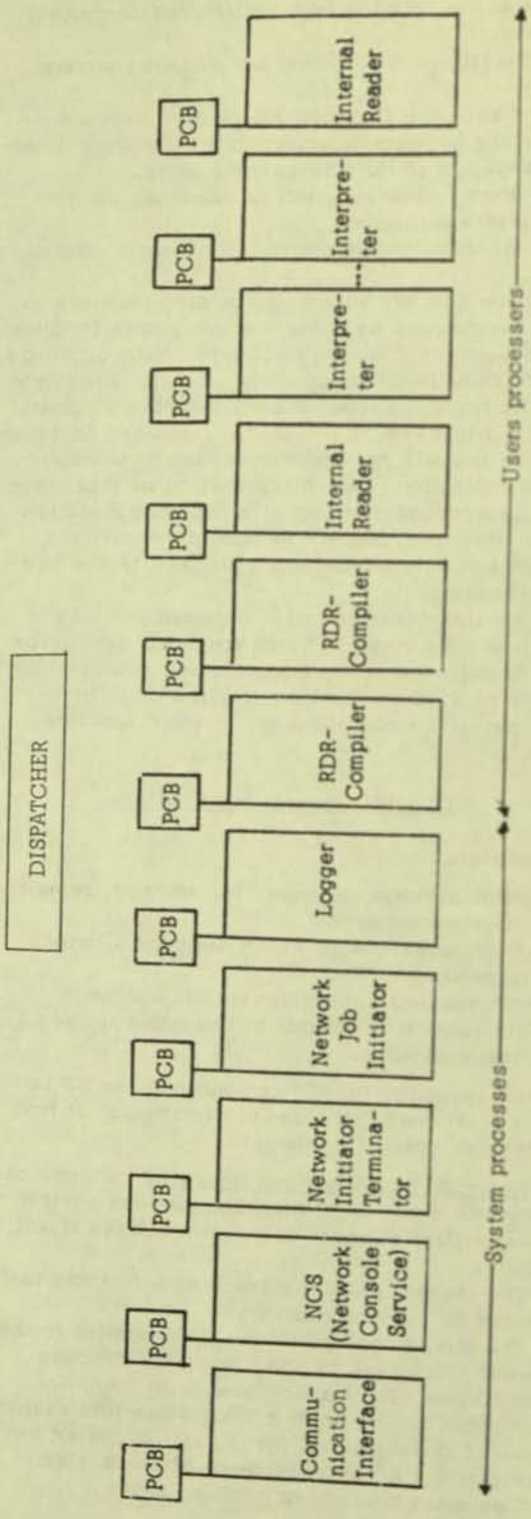


Fig. 3 NS General Architecture

. values directly stored in the command area

. pointers to blocks holding information about the necessary parameters ; these pointers are relative addresses in the data area ; so, two bytes will be large enough.

Action code	Instruct. length	Parameter 1 (or point to)	Parameter 2 (or point to)

NIL Command Format

The instruction length allows us to use variable length statements or statements with optional parameters.

More than thirty NIL commands has been already defined. Most of them are used for the local manipulation of data or for local facilities. The most interesting ones are the following :

Communication commands :

- SINFO and RINFO which allows to send and to receive system control information of any kind between two computers.
- TRMS and TRMR which allows transmission and receipt of data from a data set on a sender computer to a data set on a receiver computer.

Initialization and synchronization commands :

- DJC : request for a corresponding job
- DPC : request for synchronization with a process associated to the same job on the computer, named in the parameter area of the command ; the synchronization will be established if the interpreter on this computer for this Network job, starts and interprets a complementary DPC towards the first computer, during the time when the first interpreter can wait. The synchronization mechanims is complicated enough ; in this paper, we will say only that the interpreter asks for help each logger to handle the DPC Command.
- SPC (Stop Corresponding Process) stops the conversation with the corresponding process. It is the inverse function of DPC.

Execution command :

- INSERT : allows the introduction of a job in the local operating system.

See an example in Annexe 2.

VI - Additional Main Functions

A lot of service functions are included in NS. The most important of them are :

- the communication interface. The used transmission technics are : message switching, store and forward, multiplexing technics. More details about them and the use of these technics in the NS environment are given in the presentation of Mr. Girardi (3).
- the Network Console Service (= NCS) : the NS is the Operating System managing the Network ; so, it must communicate with operators of each machine. NCS allows the operator to introduce commands to NS, relative to NS itself (e.g. number of active computers) or to a particular Network Job (e.g. the state of a job).

VII - Foreseen Extensions

The first extension which we plan to introduce into the first version is the conversational use of network.
This one allows the users :

- to generate Network Jobs from an interactive terminal
- or to use a Conversational System located at a remote computer via terminals connected to a local computer.

The activity of the terminals local to a CPU will be controlled by a special process called NETS (Network Terminal System). More details about it can be found also in (3).
Other extensions are foreseen which will optimize or develop the facilities given by the NS. They will be defined according to the results of the first version.

Conclusion

Now, near the end of the implementation of the first version of SOC, two points seem important to us :

- first, the users have tools to use this network which allow them :
  . to access all the data located on every computer
  . to use any program or special hardware facilities available somewhere in the network
  . to run their job on the computer of their choice.
- second, the chosen environment is restrictive, but this is justified by :
  . the means we had at our disposal
  . the difficulties which the use of a heterogeneous network offers, to the non-specialist users.
  Further, we thought that a simple environment could, at least in a first step, present better increase in the use of General Purpose Computer Networks.

It seems to us that the main task which the people involved in the realization of the networks, must do , is to improve the use of the Networks in many applications problems ; this, in order to persuade the potential users that the advantages they can obtain are attractive enough to face the difficulties of use.

Annexe 1

The example of a Network Job

```
NETIN  SMITH ;
CPU  B : = IMAG67, C : = CIRCE75 ;

DSN  DS1 : = SOC.OBLIB/111111/2314/B,
     DS2 : = SOC.OBLIB/MVT192/2314/C ;
COPY  DS1  to  DS2 ;

INP  PL1JOB : = '. ./'
//PL1  JOB ......
//  EXEC  PL1FCLG
//PL1.SYSIN  DD  *
      MAIN : PROC  OPTIONS (MAIN) ;
           :
      END  MAIN ;
/*
//GO.DS  DD  DSN = SOC.OBLIB, UNIT = 2314,
              VOL = SER = MVT192, DISP = OLD
//GO.SYSIN  DD  *
                 :
             Data
                 :

/
. ./
RUN  PL1JOB  AT  C ;
COPY  DS2  TO  DS1 ;

NETOUT ;
```

Annexe 2

Example : part of the NIL program for the Network Job given in annexe 1.

A (origin of the job)	B (IMAG67)	C (CIRCE75)
DJC B DPC B TRMS LI(B) SPC DJC C DPC C TRMS LI(C) SPC	Corresponding functions on B and C are included in the Internal Reader	
DR B	Reservation commands for DS1 DPC C TRMS DS1 SPC	Reservation commands for DS2 DPC B TRMR DS2 SPC
DPC C TRMS PL1 JOB		DPC A TRMR PL1 JOB INSERT PL1 JOB SPC
SPC		
	Reservation commands for DS1 DPC C TRMR DS1 SPC	Reservation commands for DS2 DPC B TRMS DS2 SPC
JEND	JEND	JEND

References

1 - M. Elie, "General Purpose Networks of Computers", University of California, Los Angeles 1970.

2 - "SOC Report 1", IBM-France, Scientific Development, June 1971.

3 - S. Girardi, SOC Project, "an Experimental Computer Network", Proceedings of the ACM International Computing Symposium, Venice 12-14 April 1972.

4 - L.G. Roberts, "Resource Sharing Computer Networks", ARPA Project, Spring Joint Computer Conference, Atlantic City, May 1970.

5 - R. de Caluwe, "Etude d'un Langage de Commande pour un Système d'Ordinateurs Connectés", Rapport de D.E.A., Université de Grenoble, 1971.

6 - "INTENET Report 2", IBM Research, Yorktown Heights, 1970.

7 - R.M. Rutledge, A.L. Vareha, (Carnegie Mellon University), L.C. Varian, (Princeton University), A.H. Weis, S.F. Seroussi, J.W. Meyer, J.F. Jaffe, M.K. Angell, (IBM Research Yorktown Heights), "an Interactive Network of Time-Sharing Computers", 1969.

8 - D. Fredericsen, L. Loveless, J. Rooney, R. Ryniker, S. Seroussi, A. Weis, "OS/360 Network Interface Users Guide", IBM Research, July 1971.

9 - D. Fredericsen, R. Ryniker, "a Computer Network Interface for OS/MVT", IBM Research, April 1971.

10- Control Program-67/Cambridge Monitor System (CP67/CMS), Version 3 - System Description Manual, (IBM-GH20-0802-2) .

THE USER DEPARTMENT AND THE COMPUTER

C.R.M. Singer
International Computers, Ltd.
Harrow, Middlesex, England

Summary

For many years systems which computerise administrative procedures have been implemented with reasonable success albeit with some sacrifice of timeliness and accessibility. Relatively little success has been achieved in using data from these systems to aid day to day decision making. A major difficulty throughout has been the main files being out of date and the remoteness of the computer from the user department (defined as those departments that originate the computer input and/or use the output.) The lengthy procedure required to get valid data into the computer is identified as the main cause of the problem.

This paper then proposes a solution based on new generalised data capture software from ICL called 'Datastream'. It uses large screen visual display units that can be sited locally by a main frame computer or remotely over telephone lines. The system is described from the user's point of view showing how it can be used as a data preparation system, a form filling system, the basis of any VDU system or a combination of all three. It is shown how the system can be cost justified from the start, how the user becomes responsible for getting his own data into his own systems, how valid data gets into the computer very much faster; thus how master files are updated very much earlier, how the user can economically access his main files through VDU's and thus how the potential to use the computer to assist in day to day decision making becomes practical.

Ever since computers were first introduced to the commercial world, there has been a distressingly frequent "gap" between early expectation and final realisation of systems benefits. At the outset seemingly reasonable claims were made about, for instance, computerising administrative procedures to cut overhead costs, with the argument that a natural development would be to use the data so collected to help in day to day decision making - to help forecast sales, control stock, plan production. Figures would also be produced to show that whilst the system would be cost justified in the first instance in the saving of administrative costs, the real pay-off would be the later developments, where one would be essentially implementing systems to aid the control of cost of goods. Furthermore if this computer aided decision making could produce a 10% reduction in cost of goods, the pay-off would probably be as high as four times that of a 10% reduction in administrative costs.

Why were the initial aims not fully met? It would appear that the problems were not so much with the technical aspects of computerisation but with the general modus operandi. And one finds that there is a frequent repetition of the same problem areas amongst many different installations. It is the purpose of this paper to identify those problem areas, to analyse their basic cause and to propose solutions.

Problem Areas

In conducting an analysis of existing systems one frequently shows that administrative costs have been cut as planned or that the business has grown without a similar increase in administrative overheads. And to that extent the initial objective of the system will have been largely achieved. However, one almost equally frequently finds that this objective has been achieved at the expense of less accessible data and less timely data, the result of which is the retention of the old manual files by the user departments. Not only of course does this retention weaken the initial success but far more important it leads one to question the chances of being able to use that computer data to aid decision making - "How can we use the computer to run the business when its files are out of date?" the user will say, "We must retain our manual systems to do that". And, of course, so long as that situation prevails there is little chance of progress. However, all this so far is an effect, not a cause. What is then the fundamental cause? To study this one must look closely at the cycle of events required to get valid data into the computer from a user department using conventional punching and verifying.

The cycle is typically:-

1. user department fills in a form
2. sends it to the punch room
3. where it is checked,
4. punched,
5. verified,
6. input to computer
7. which prints invalid list
8. which is returned to the punch room and/or
9. sent back to the user department

This cycle will seldom take less than a day, frequently 3; or more if the user is remote from the computer; and the cycle gets repeated until the data is valid. And the cycle is often even longer. And frequently the whole batch of data is held up pending corrections.

Furthermore many different people in many different departments are concerned with handling the flow of data. The effect of this cycle is, of course, that valid data takes a long time to get into the computer whose files then become out of date; this may be tolerable in many administrative systems, even though cash flow may suffer but quite intolerable in the day to day decision making. A further factor is expense. That cycle of events is very expensive.

Solutions

What can be done about that fundamental cause: that lengthy cycle of getting valid data into the computer.

One can install key to disc data preparation systems. The principle here is that the 'punch' operator has a keyboard connected on line to a mini computer with validation performed as she keys in. Thus there is no need to do so much verification by repetition. So fewer girls are needed; and because records are checked whilst the source document is still with the girl one cuts out part of the cycle (6, 7 and 8 above) and saves time. As the device is all electronic the girls key faster. Result - even fewer girls. Overall: data into the computer more quickly; probably at less expense; and the master files are more up to date.

Key-to-disc systems have been successful in reducing excessive data input costs and delays. But they do ignore one major cause of that lengthy input cycle; and that is the time and cost of getting the data corrected at source in the user department; i.e. outside the confines of the tightly controlled computer Data Preparation Department.

So we need to concern ourselves with the problems of computer data input as felt by the user department. ICL have produced some generalised software to help which uses a large screen visual display unit linked into the main frame computer. The software, called 'Datastream' provides all the normal facilities of the key to disc systems and thus can be used as a pure data preparation system providing all the benefits described above.

But it can go a stage further. The hardware and software are such that one can site the VDU either locally or, over telephone lines, remote from the computer. What is more, as a large screen VDU with 2000 characters is used the software is structured such that the operator in effect fills up a form on the screen. An example will be given later. Thus one can site the VDU in the user department and get the user to fill up a form, perhaps not as now, on paper, but on the VDU screen. Thus data is captured and validated at point of origin; the user is responsible for putting in and validating his data; in so doing of course we cut out numbers 2, 3 and 9 of the cycle so data gets into the computer more quickly. And what is more, if the clerk who used to write out the old punching documents now types into the VDU, we've cut out the data preparation function altogether; so there's no need for the data preparation operators at all.

Overall result: data into the computer even more quickly; at even less expense and master files even more up to date.

But the system can go a stage further still. The software uses the standard main frame communication handling utilities. So it is very easy to add onto the software a purpose built communications system like order processing or sales ledger etc. That of course cuts out yet another part of the cycle - no more form filling even, as the system is now an integral part of the business. But more important in the context of the purpose built system; you can base the purpose built system on this package. Compared to starting from scratch you cut development costs, you add the flexibility of the package's formatting routines to your purpose built systems, you can add the benefits of the data capture system to your own system without necessarily adding to your cost; perhaps the most comforting from the planning point of view, if the purpose built system doesn't quite cover conveniently all the input required, or if the systems designer has forgotten some, it can frequently be left to the package, even if only as a temporary measure.

So what is this software? It is a data preparation system, a form filling system, the basis of any VDU system, or a combination of all three. But above all it is a system which allows the user departments to cut the time and cost of getting valid data into the computer; allows them to get close to their computer systems; and unlock the potential for developing systems to assist in decision making. The responsibility and control of data input is in the user's own hands.

The Datastream System

The software consists of three major facilities, namely:-

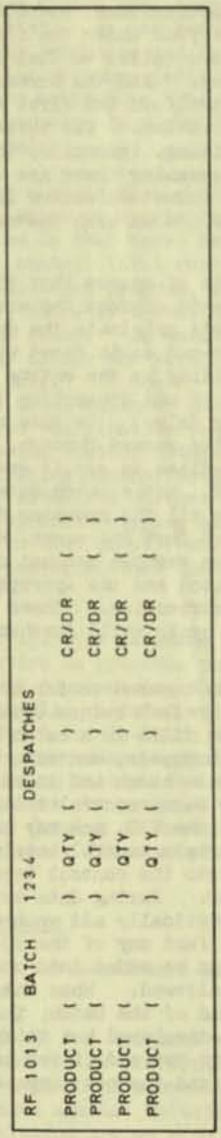
Record formatting which allows the user to define via VDU's the formats he wants for any one job. Each field is suitably labelled just as he wants and layed out as he wants. This is a once off exercise. At the same time validation checks are assigned to appropriate fields.

The second facility, Data Entry, is the main facility and allows the input, validation and verification of data via the VDU's and the dumping of complete batches of input to user files.

The third, and optional facility, provides a printed analysis of the VDU operator's performance during previous Data Entry runs.

Let us consider the system working in practice.

The start of most such systems is a punching document. The example is a typical one using part of an invoice but the system is relevant to all companies concerned with getting data into computers. You will see that various fields are labelled with gaps between brackets all ready for the operator to fill in with data. Fields can be headed vertically just as easily as horizontally as shown.



```
RF 1013    BATCH    1234    DESPATCHES
PRODUCT (    ) QTY (    )   CR/DR (    )
PRODUCT (    ) QTY (    )   CR/DR (    )
PRODUCT (    ) QTY (    )   CR/DR (    )
PRODUCT (    ) QTY (    )   CR/DR (    )
```

Fig. i. A typical Datastream format

The operator types in, left or right justified whatever is convenient for the job. What she types appears in brackets and when she gets to the end of a field she tabs and automatically moves over the headings into the next field. She knows where she is by the position of the cursor which moves across the screen as she types and tabs. If she consciously makes a mistake she can easily move the cursor back and overwrite what she's just typed, or insert some more characters automatically moving the others across. What she cannot do is to upset the headings. They are protected from her.

When one form is complete she presses the send button and the data is transmitted - only the data, not the headings. The computer performs its validation checks. The sort of checks you can make are all the usual ones like numeric or alphabetic, check digits and so on. Also the program contains a User Entry Point through which the user can include any special checks of his own, including file look-ups.

Normally the girl will key in data rapidly at about 10,000 to 18,000 key depressions per hour depending on job complexity, barely glancing at the screen.

As soon as she's satisfactorily filled the screen and transmitted the data a fresh formatted screen is presented to her.

The above illustration was part of a despatch note. What was illustrated was the body of the despatch note. A real despatch note would probably consist of two header sections containing customer details (addresses, order number, date etc.) and then the body section containing itemised despatch details. Those who are familiar with the ergonomic problems of preparing data for computers will know that this presents the great dilemma. In a conventional system one has to compromise between the single bit of paper that is easy to fill up in the user department and the card punching which requires each part of the form to be on a different card format, yet initially in the same computer input record. In Datastream there is no problem; either you lay out the screen to look like one complete despatch note or you have the option to have up to 9 pages per format. Fig. ii shows four pages.

There are two essential points here, one is that you can set up the system to represent existing forms, or a whole series of existing forms, if necessary. This means minimum change when the VDU's are located in the user's own office thus making the system easy and acceptable to non specialist operators. Secondly by dispensing with punching documents one reduces transcription

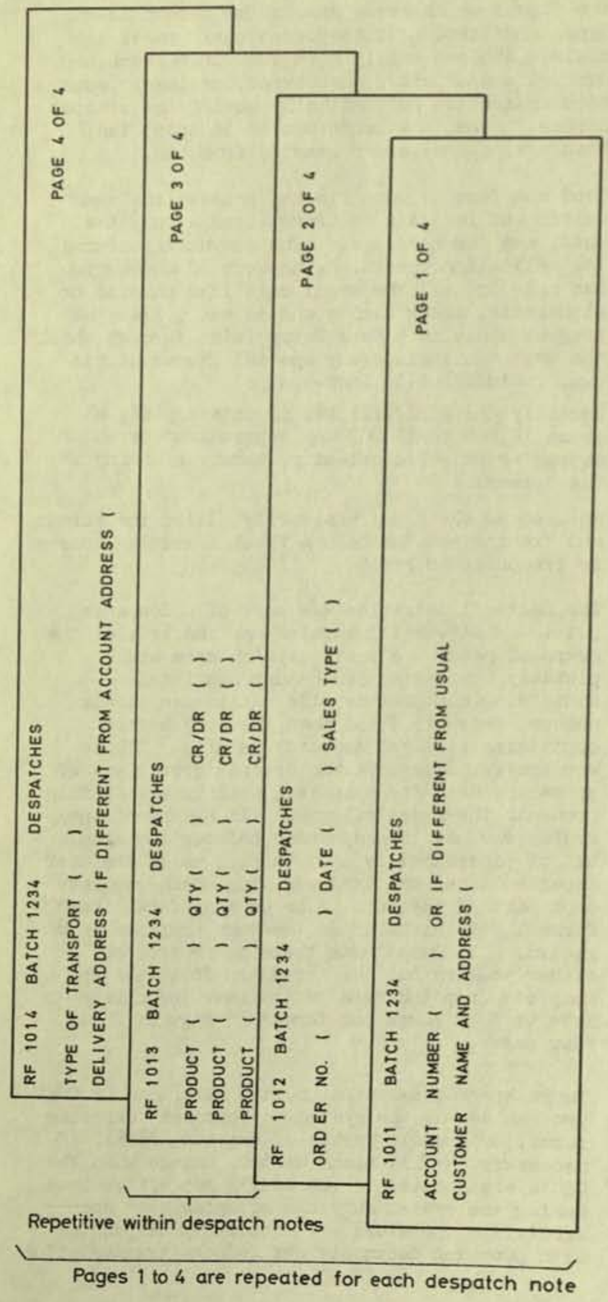


Fig. ii. A typical multi page format illustrating a despatch note

errors. Remember formats are set up as a once-off exercise using the record formatting program. It normally takes about an hour to design a page and about 15 minutes on-line time to set up.

Let us now consider what happens when the VDU operator has completed a screen, transmitted and the computer finds errors. As she will be keying in fast and barely looking at the screen, a buzzer sounds and the keyboard locks. And as she may not be a professional operator, special error codes will appear under the fields or characters that have failed so that she knows what has gone wrong. And the cursor automatically positions itself on the first error field all ready for correction. She corrects the mistake by overwriting, transmits, the computer re-validates and assuming there are no further mistakes, a fresh formatted screen is presented to her and she carries on with the next document.

It is possible of course that the VDU operator is unable to correct the error. If the operator doesn't originate the data she may have to refer back and so is faced with the possibility of holding up the entire batch or removing that record and disturbing the control totals, so causing delay. In this system one can opt to force the record through, using a special key. Needless to say it gets flagged as an error record. At a later date when the correction is made all the operator has to do is to type in the fact that she wants, say the ninth record or for example account number 123 of a particular batch and the appropriate record is presented for correction. These facilities reduce that input cycle still further.

One aspect not covered so far is the batch control totals. In fact before data entry starts the operator fills in a batch header. This, as the name suggests, contains data relevant to the whole batch and in addition contains space for batch control totals. Before one even starts on the VDU, one may have add-listed off-line certain control totals and these totals are typed into the control total fields in the batch header. During data entry the computer will automatically add up the appropriate fields; in fact any of the 90 fields allowed per page can be added into one or other of the two totals allowed. When the operator signs off at the end of the batch, the original batch header is re-displayed but this time showing a comparison between the add listed totals originally entered and the new computer calculated total.

Verification. Once the data is entered into the computer it is ready for verification although this is entirely optional. If the VDU is located in the user department then the operator may well be filling up a VDU form rather than a paper form in which case a visual check will be adequate. Alternatively if the VDU is being used as a pure data preparation device one may wish to key in all the data again as in conventional punching techniques, comparing against the original data. When miscompares occur the buzzer sounds, the keyboard locks and error codes (M's for miscompare) appear under the appropriate characters all ready for correction. However, the most common form of verification is called selective verification. It is the same as full verification except that only certain fields are verified. The rest are automatically skipped over. One will frequently omit verification on those fields that have been comprehensively validated. As has been mentioned above one has the option to add in one's own special validation routines including file look-up. Remember the more validation, the less verification, less equipment, less staff, lower costs and less input delay.

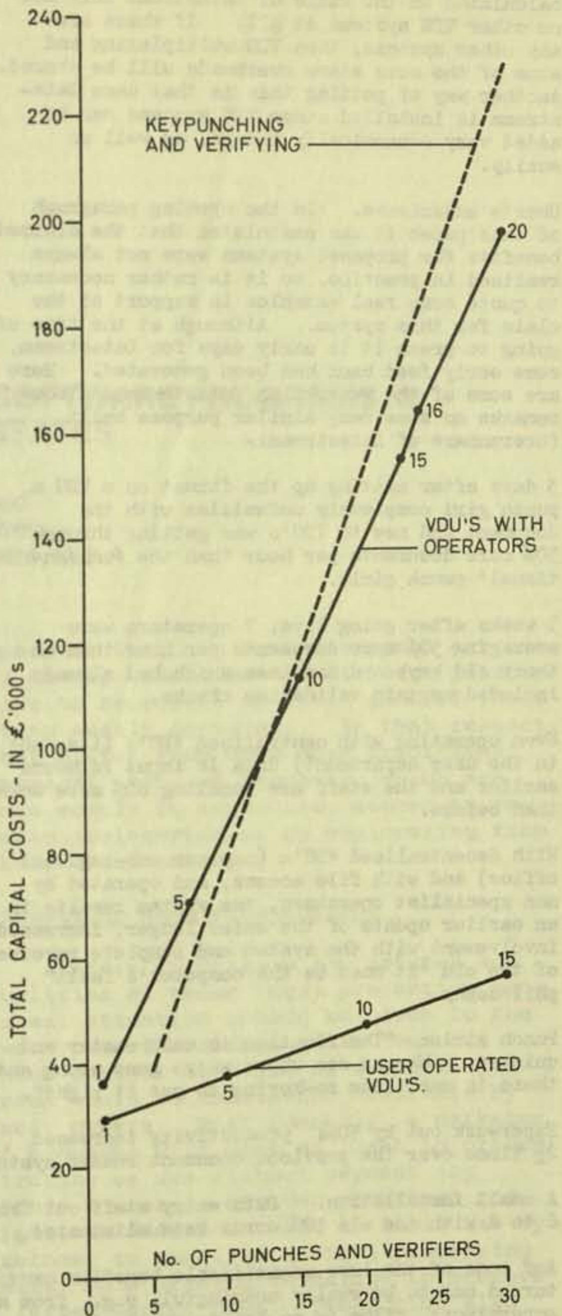
There is in fact a third more specialised option in verification which is specially designed to cut down the time in that basic data flow. It is called batch control total verification. In a normal batch validation program, when control totals disagree, there is a considerable delay in establishing the cause. In Datastream this special facility allows the verification of control totals only. It can be used immediately after a control total failure, possibly before full or selective verification begins, to establish where the control totals have gone wrong. By verifying control total fields only, one can quickly establish whether a record has been left out (without retyping all fields of all the records up to that point), whether a control field was mistyped or whether there was an error in add-listing.

Finally, whilst all this is going on the system compiles statistics on operator performance (i.e. the number of key depressions, the rate of keying, the number of invalid records, etc.). These are available for print out in batch mode and/or for input into one's own bonus payment scheme. Also whilst data is being entered the supervisor can use a VDU to monitor the progress of each batch. She controls the dumping of data from a work file into main files ready for user programs when convenient.

Practical experience

Two aspects are considered here: cost and user experience.

Cost is a difficult one to cover as each installation tends to be different. However, fig. iii shows a graph of conventional punching equipment compared against Datastream used as a pure data preparation system with specialist operators, and Datastream used as a form filling system operated by non specialist operators in between normal clerical duties. In practice any one situation is likely to be between the two Datastream curves.



Note:

(1) Nos. of VDU's are adjusted for normal increased throughput.

(2) Allowance is made for the whole cost of the VDU's and half the cost of core and disc where will generally be used in one shift only.

Fig. iii. Typical cost comparison

Another point on cost is that these graphs are calculated on the basis of Datastream only and no other VDU systems at all. If there are any other systems, then VDU multiplexing and some of the core store overheads will be shared. Another way of putting this is that once Datastream is installed other VDU systems can be added very economically indeed, as well as easily.

User's experience. In the opening paragraph of this paper it was postulated that the claimed benefits for proposed systems were not always realised in practice, so it is rather necessary to quote some real examples in support of the claim for this system. Although at the time of going to press it is early days for Datastream, some early feed back has been generated. Here are some of the remarks on Datastream and some remarks on some very similar purpose built forerunners of Datastream:-

3 days after setting up the format on a VDU a punch girl completely unfamiliar with the document and new to VDU's was getting through 30% more documents per hour than the 'conventional' punch girls.

3 weeks after going live, 7 operators were averaging 50% more documents per hour than on their old keyboard machines which had already included certain validation checks.

Even operating with centralised VDU's (i.e. not in the user department) data is input 24 hours earlier and the staff are handling 60% more work than before.

With decentralised VDU's (one per sub-regional office) and with file access, and operated by non specialist operators, the system results in an earlier update of the sales ledger, increased involvement with the system and complete reversal of the old "it must be the computer's fault" philosophy.

Punch girls: "Verification is much easier and quicker. We can see where we've gone wrong and there is much less re-keying to put it right".

Paperwork cut by 50%; productivity increased 2½ times over the previous document reader system.

A small installation. Data entry staff cut from 6 to 2 with the old 10% error rate eliminated.

And some of the less quantifiable benefits have turned out to be really meaningful, e.g. from a supervisor: "Time is no longer required to reload paper tape and ribbons and far less time is taken in controlling 'batch control' slips. And it is so much quieter. So the whole operation is much more efficient".

Conclusion

The fundamental objective of the Datastream system is to get the user closer to the computer to allow him to unlock the full potential of his system. We identified a common complaint of out of date and remote files which are expensive to maintain and no use for aiding day to day decision making. We identified the cause of complaint as a cycle of at least 9 separate events which have to take place to get valid data into the computer. We showed that these 9 could be reduced down to 3, one of which (verification) could itself be substantially reduced by using Datastream. So the effect of Datastream is that:-

1) the user can be responsible for getting his own data into his own system,

2) valid data gets into the computer very much faster, so

3) the vital master files are as up to date as necessary and

4) the user gets close to his own systems (particularly as it is easy and economic to add coding to the package to access main files) and so

5) one can start developing meaningful systems for aiding decision making and so contribute to the reduction in cost of goods sold.

And in addition because Datastream can cut the cost of getting data into the computer, it is possible to install the system, cost justified, in its own right with very limited systems and programming work.

Overall, this adds up to a greater return on computer investment, quickly.

PROTECTION OF PROPRIETARY SOFTWARE PROGRAMS IN THE UNITED STATES

Roy N. Freed
Widett & Widett
Boston, Massachusetts, U.S.A.

Summary

Despite skepticism about their effectiveness, legal techniques for protecting proprietary software programs from unauthorized use have reached a fairly high level of development in the United States in practice. However, few of those techniques have yet been validated in the courts and the matter generally remains in flux. This review is intended to help persons in the United States appreciate the significance and effectiveness of those protective measures and to help persons in other countries assess, in light of their own legal systems, the adequacy of protective measures used or suggested for use in their countries.

If software program properties are valuable enough to warrant legal efforts to prevent their use without authorization, a sound legal foundation should be created. Although most customers will observe restrictions imposed on them by marketing contracts, there are outsiders to the transactions who have an incentive to seek to avoid the limitations contained in legally vulnerable approaches. Since ideal legal measures usually will be obnoxious in the marketplace, compromises with legal requirements are inevitable, but they should be made consciously and carefully.

Software Especially Vulnerable

Legal measures are considered to be more necessary to protect proprietary programs than other properties largely because software programs are valuable items of property (being expensive to create or producing extremely desirable results) that can be transferred extremely easily (their materials being inexpensive to reproduce or their general ideas being easily conveyed). In that respect, they differ from most more traditional physical items of property, which are more costly to reproduce, either by reverse engineering or by engineering from their basic concepts.

Situations Requiring Control

Before discussing the natures and qualities of those legal protective measures, attention should be given to the two types of business situations for which control over use of software programs would be considered desirable by their owners. Most commonly, a marketer of programs will want at least to prevent pirating or use without payment (by either a user or another marketer) and might, in addition, want to limit use by a customer to specific central processing units, as a pricing technique to achieve a larger market base. Protection also will be sought by owners of software programs who want to enjoy their exclusive use in their own operations, possibly because a program embodies a secret manufacturing process that had been in use or because a newly computerized method itself is considered to give them a competitive advantage. Those latter owners want to prevent all other use. In view of this diversity of goals, the types of legal protective measures have to be selected for the particular role they are to play.

Legal Measures Last Resort

Since most legal protective measures introduce complications into the marketing of software programs and the use of those programs by customers, they should be utilized only where genuinely required and then only to supplement feasible technical methods for controlling program use or practical limitations that might obviate the need for some of them. The latter include delivery only of object code, keying program tapes to particular CPU's, and making use of programs available only through time-sharing systems. Sometimes, the need for substantial maintenance might make pirating impractical.

Meaning of "Program"

The word "program" means here the general nature of the set of instructions given to a computer system to carry out certain information processing operations. In that sense, a program is essentially a concept of a method or of a group of circuits. The program is embodied or represented in specific types of physical materials that make up a software package, such as a magnetic or punched paper tape, a magnetic disk, a deck of punched cards, a listing print-out, a narrative description, a flow diagram, or a manual. Some protective measures actually involve the program; others relate to the program materials. It is important to make that distinction lest the desired protection not be secured because of a misapplication of legal measures.

In using any protective approach, it is essential to know the legal basis relied upon in each case. For that purpose, it is important to recognize exactly what types of property are protected by patents, copyrights, or trade secrets. It might be information or ideas stated in writings, on the one hand, or physical items constituting either writings or operative articles, on the other. The distinction, however subtle, can be important in securing desired legal protection.

Many marketers of software programs who do not secure patents are properly reluctant to rely solely on copyright protection, particularly where they seek to charge customers based on the number of CPU's on which a program is used. They resort, in addition, to various contractual restraints on disclosure and transfer. The need to rely on that combined approach, which includes contractual commitments, is the reason why outright sales of marketed software programs are so rare and why continuing relationships of relatively limited duration, such as licenses and leases, are used instead. Fairly general resort to those types of relationships is a reflection, often unwittingly, of the general principle of law that a seller of tangible personal property may not impose restraints on the buyer's alienation of it.

It should be recognized that the various legal protective measures discussed below have received relatively meager scrutiny, and hence validation, by courts, even in the case of patents. Despite their high level of development in practice, software protective measures still are in a considerable state of flux legally.

Patenting

Patenting normally is regarded by owners of technical ideas as the highest form of protection for patentable items that are marketed. The owner of a valid patent may institute legal action to bar others from, or to charge them for, practicing the patented invention, regardless of the fact that those others might have arrived at the same idea entirely independently. That invention may be either a process or an article of manufacture.

However, attitudes in the United States toward the desirability, and hence the propriety, of securing patent protection for inventions reflected by software programs have been ambivalent. Many large manufacturers of computers and the United States Patent Office adhered to a negative position for a long time after the possibility was first broached. In contrast, a number of companies engaged either primarily in the marketing of proprietary software or as users of software developed in house favored patenting. The Patent Office took the position that software programs are not "statutory material" and hence are not patentable. It undertook to bolster its adverse position by describing the apparent difficulty it foresaw of making novelty searches of the prior art, in view of the alleged inadequate documentation of software programs in technical literature, and by expressing concern lest it grant an unreasonable number of patents that are invalid because of anticipation.

Essentially, the struggle between the opposing forces continues. As time went on, a number of applicants appealed rejections by the Patent Office and secured favorable rulings in the Court of Customs and Patent Appeals.[1] Those decisions indicated that software programs might reflect inventions patentable either as processes or as articles of manufacture. Despite the rulings of that court, considerable uncertainty (and hostility) remained. Since patent validity may be challenged in the United States district courts and the judges of those courts are not bound by the decisions of the Court of Customs and Patent Appeals, conflicting judicial views might arise. In an effort to secure a binding determination for all courts, steps were initiated in late 1971 to get a ruling on the subject by the United States Supreme Court and that court has agreed to consider the matter in the case entitled Benson v. Tabbot.

Proponents of the patenting of inventions reflected by software programs and in program materials argue that patentable inventions (assuming the requisite novelty and inventiveness) might be found in the processes carried out, in the tapes or disks as articles containing unique arrangements of signalling means, or in the specially programmed general purpose computers, again as articles of manufacture.

Many opponents have been insisting that software programs are unpatentable because they reflect essentially series of steps that can be carried out in a human mind, an activity not readily subject to legal restraint.

Copyrighting

Although most attention seems to have been devoted to the possible use of statutory copyright to protect proprietary software programs, actually common-law copyright appears to be useful and could be superior in some situations.

The Copyright Office of the Library of Congress is willing to accept computer tapes and similar program embodiments for registration of statutory copyrights on them.[2] It requires that those things be accompanied by counterpart items readable by people, such as print-outs. Although many suppliers profess to copyright their software program materials, few have registered them, probably because statutory copyrighting is accomplished simply by affixing a proper notice and registration usually is necessary only as a condition for an infringement suit.

It is not certain that program tapes and punched card decks are statutorily copyrightable items. Federal courts, rather than the Copyright Office, are the authoritative bodies to interpret the copyright statutes and none has had occasion to rule on the subject.

Statutory copyright is applicable primarily only to published works. It might be questioned whether many program materials have been "published," in the statutory sense. If published, the content information or ideas of the works are available to the world for use without restriction and without charge (unless they describe a patented invention). The copyright owner merely has the power to prevent the physical copying or translation of the work without permission. Traditionally, he has been supposed to secure his economic reward by publishing and by using that power to exclude competition with that publishing activity.

Moreover, if the copyright owner sells a copyrighted work outright, the purchaser may use that work in any way he desires and may transfer it to any other person, except that none of them may make copies without permission. It is worthy of reiteration that the seller of a statutorily copyrighted work may not control the buyer's use of it.

Where, as is frequently the case, a software supplier couples a claimed statutory copyright with a contractual commitment that the customer not disclose the work to others, a fundamental antithesis is created. It seems to be a derogation of that type of copyright to try to bar dissemination of the contents of the work.

In contrast, common law copyright attaches automatically, by virtue of common law of the various States, to writings that are not published. Conceivably, software program materials distributed to a relatively small group with restrictions on further transfer and disclosure might qualify for that protection. The use of express non-disclosure commitments would seem to bolster that form of protection, in contrast to the reverse effect in the statutory copyright situation. Although no form of notice of common law copyright is prescribed, use of a notice on software program materials is advisable since the circumstances of distribution usually would not impart notice, as would be the case with a manuscript of a book.

Trade Secrets

Rules of law applicable to trade secrets, found both in the rulings of judges, under the common law, and in statutes, also have been suggested as means for protecting software programs. They make it illegal to use the trade secrets of another secured by bribery, theft, or other improper means.

Trade secrets consist of valuable business or industrial information acquired at large expense or by much effort or skill and screened by particular efforts from general disclosure. That information need not be unique to the person claiming it. A number of companies may simultaneously have identical trade secrets that originated from different sources.

Trade secrets may represent manufacturing process data, business information, marketing strategies, and, it is said, even physical articles. It is argued that the information reflected in software programs may constitute trade secrets. Information included in a software program for carrying out a secret manufacturing process probably would so qualify. The tapes and other materials embodying that program very likely would contain more complete data respecting the secret process than would be recorded in a non-computerized operation. Their use internally probably would satisfy the requirements for trade secret protection.

Many trade secrets are marketed. The traditional arrangement is to license the use of the technical know-how. In the past, that know-how was made available to instruct people in carrying out a process and was communicated in writing. Now, the information frequently guides machines by means of computers. Although it is stated that an article may constitute a trade secret, technical information revealed by a marketed article cannot be a trade secret. More accurately, the technical data respecting a special article useful in a manufacturing process, rather than the article itself, may constitute trade secrets.

Licensing

Many software marketing arrangements labeled as licenses probably are not fully enforceable because of legal deficiencies. If a truly effective arrangement is desired, it is essential to understand the nature of licenses in the law and to fulfill legal requirements.

A license is a permission to do some act that would not be permissible without it. To state merely that the use of a software program is licensed is not enough. It is necessary to go further to establish some legal basis that bars the unrestricted use of the programs involved and then to license the practicing of a patented invention, the copying of a copyrighted work, or the use of secret information, depending upon the particular legal basis established for protection.

For example, some forms purport to "license" customers to use, but not to copy, statutorily copyrighted manuals only in connection with particular CPU's.[3] If those manuals are sold outright, that limitation probably is unenforceable legally. Normally, only the lessor of such a copyrighted work may restrict how it is used during the lease period.

Leasing

In some cases, software program materials are leased. That is proper for materials that are tangible personal property. Such leasing has many legal advantages because it involves a well-recognized relationship in which the lessor enjoys many powers sought to be exercised by the software program supplier.

A lessor normally may limit the use to which the leased property may be put and may forbid its transfer to others. These specific powers provide a firm basis for efforts by software program suppliers to limit use of their products to particular CPU's and to prevent transfer of the products to others.

It is permissible to lease items covered by statutory copyright. Motion picture films have been distributed that way for years.[4] The same approach would seem to be consistent with common law copyright.

Since both suppliers and customers usually prefer an outright sale of software programs to a continuing but limited-term relationship of lease or license, many suppliers are tempted to try the expedient of a so-called single payment lease and other tactics to reduce the nature of the continuing relationship. To the extent that those measures depart from the traditional qualities of a lease, they tend to undermine the legality of the characterization. For example, if it is required that only a single payment or relatively few payments be made and the supplier fails to exercise significant dominion over the property involved and to compel its return at the end of the lease, a court well might find that the transaction really was a sale.

Securing Written Non-Disclosure Commitments

Since many software programs can be created from the concepts and approaches of other programs, marketers frequently want to bar their customers from disclosing the information content of their products. As indicated, statutory copyright does not prevent use of ideas contained in published copyrighted works. On the contrary, those ideas are available freely to the world. Hence, as already suggested, non-disclosure commitments are incompatible with statutory copyrights. They are entirely consistent with common law copyrights and probably should be used to remind customers of their legal obligations.

Non-disclosure agreements not only are compatible with the claim of trade secrets rights but should be used in order to bolster them legally. If owners of trade secrets do not make reasonable efforts to preserve secrecy, they jeopardize their property rights.[5]

For legal purposes, software program marketers relying on trade secrets probably should impose more stringent non-disclosure duties than they can secure as a matter of business practice. However, burdensome duties cannot be forced on customers unless a product is truly unusual.

Proposed Software Program Registration

IBM has proposed that a new form of software program registration be authorized by Federal statute.[1] Under it, a program owner would file program listings in secret with a Government agency and would publish a general description of the program. If an unauthorized use of the filed materials were encountered during the relatively few years of protection accorded, suit could be brought and the secret materials could be used to show unauthorized copying. No novelty would be required of the registered program. The protection would bar copying but permit use of ideas by those prepared to implement them at their own expense. The approach is similar to existing protection of trade secrets.

Combination of Measures Appropriate

Each protective measure has limitations and, for software programs that are marketed, none is a panacea. It is essential, thus far, to create combinations of measures in designing software program marketing arrangements, but the elements should be selected with a clear understanding of their qualities and compatibilities.

It appears that the best combination, in principle, in many marketing situations includes common law copyright (because no publication is involved and it is compatible with non-disclosure and use limitations), leasing of the program materials (to validate limitation of use to a particular CPU and to bar transfer to others), licensing of the making of copies (to restrict the quantity available) unless the supplier will furnish all copies required, and imposition of non-disclosure commitments (to bar disclosure of the ideas to unauthorized persons).

In the many situations in which it is advisable, for valid marketing reasons, to depart from the strict legal requirements of the various protective measures, it is important that the marketing reasons truly be sound. All too often, marketing personnel seek to degrade those measures in order to make it easy for them to secure customers. Where compromises are firmly based, their existence should be acknowledged within the supplier's organization explicitly to avoid later recriminations when deficiencies are taken advantage of by customers or others successfully.

Where it is sought to achieve exclusive use of a software program in-house, it seems to be advisable to treat it as a trade secret, fencing it in with non-disclosure agreements and marking the materials with notices that common-law copyrights are claimed.

Legalities Are Important

It is stated frequently, usually by non-lawyers but occasionally by lawyers as well, that the type of concern expressed above for observing legal niceties in structuring marketing arrangements is unnecessary because customers and suppliers are stating their agreements in writing and those agreements will be observed and, if necessary, enforced. That attitude can be somewhat cavalier in many situations and conceal weaknesses in the supplier's property rights. There are numerous circumstances in which the legal soundness of the supplier's protective structure might be challenged. In many transactions, third persons with interests adverse to both parties will seek to overturn the legal characterization adopted by the parties. They would include a trustee for a bankrupt customer and a creditor seeking to collect a judgment from a customer. Both would seek to assert that a lease or license actually was an outright sale to the customer. Even the customer itself might renege and take such a position. Licensees under patents occasionally seek to challenge the validity of those patents, and the Supreme Court recently permitted that to be done while the license is in effect.[6]

References

1. E. Galbi, "Software and patents: a status report." 14 Comm. of ACM No. 4, April 1971, p. 274.

2. Copyright Office, The Library of Congress. "Computer Programs" Circular 31D.

3. International Business Machines Corporation. "License agreement for IBM program products."

4. Warner Bros. Distributing Corporation. "License agreement." Reprinted in part in Freed, "Materials and Cases on Computers and Law," Third Edition 1971, at p. 302.

5. Healy v. Murphy, 1970 Mass. Adv. Sh. 1051, 260 N.E. 2d 723 (1970).

6. Lear, Inc. v. Adkins, 395 U.S. 653 (1969).

TOOLS IN RESPONSE TO A SOCIAL NEED

COMPUTER COMMUNICATIONS–AN EMERGING TOOL IN RESPONSE TO SOCIETY'S NEEDS

CHAIRMAN:

Reg A. Kaenel, AMF Incorporated
Stamford, Connecticut, U.S.A.

PROGRAM COORDINATOR:

Dr. Raymond L. Pickholtz
George Washington University
Washington, D.C., U.S.A.

PAPERS:

Dr. Reg A. Kaenel, *"Computer Communications–An Emerging Tool in Response to Society's Needs?"*

Dr. Takeshi Utsumi, Assistant Department Manager, Information Processing Department, Mitsubishi Research Institute, Tokyo, Japan, *"Global Gaming–Simulation with Computer Communication for International Cooperation"* (abstract only)

DISCUSSANTS:

Dr. W. J. Bray, Director of Research, British Post Office, London, England

Dr. N. Seshagiri, Director, Information Planning Analysis Group, Electronics Commission, New Delhi, India

Professor I. Paz, Government of Israel, Armanent Development Authority, Tel-Aviv, Israel

Professor F. Reza, Ambassador, Permanent Delegate of Iran to Unesco, Paris, France

Dr. Addison Richmond, Department of State, Washington, D.C., U.S.A.

THEME: Combination of the established technologies of computers and communication is giving rise to an accelerating emergence of information networks. Effective uses of these networks depend on the requirements of the communities they are to serve. This session aims at identifying requirements that can be exceptionally well met by such information networks. In particular, the discussion will explore regulative requirements of different nations and attempt to project the growth profile of the requirements of developing nations.

COMPUTER COMMUNICATIONS
AN EMERGING TOOL IN RESPONSE TO SOCIETY'S NEEDS?

Reg A. Kaenel
AMF Incorporated, Stamford, Connecticut, U.S.A.

Combination of today's established technologies of computers and communication is giving rise to an accelerating emergence of information networks. The communication traffic between computers and terminals of these networks has been predicted to grow by at least an order of magnitude within the next decade, increasing the common carrier revenues from communications-oriented computer installations from about \$1.25 billion in 1970 to about \$17.5 billion in 1980.[1] In the late 1960's, there were some 75,000 terminals served by 2,500 computer systems.[1] It has been forecast that by 1980, 50,000 - 75,000 systems will serve 10 - 30 terminals each.[1]

Computer systems are currently used in some 3,000 different applications.[2] Of these, about 300 are concerned with statistical analysis and 175 with operations research; in addition, about 100 general-business programs are used for billing, payroll, sales analysis, and the like.[2] Long lists of distinct applications of communications-oriented computer networks have been projected which are expected to impact such areas as economic, social, environmental, medical, quality of life, safety, decision making, management, politics, leisure/recreational, etc., etc., etc.[2]

But each social structure has its own requirements for information and for the processing of such information. For example, these requirements are higher for an industrialized nation than for a developing one. Thus, while an application of an information system may be highly effective in an industrial environment, it may be of little use to a developing society and may even be detrimental to such a society.

Information networks represent a formidable new technological resource for society. Today's challenge, now, is the definition and execution of plans by which these networks are usefully and nondisruptively integrated into existing social structures, and by which they are made to follow the dynamically changing requirements of these structures. This Session, then, will explore where information networks can uniquely contribute to solving needs of different societies and what strategies are suitable for attaining useful applications of such networks. Also, practical shortcuts for establishing information networks in developing countries will be examined.

---

1. GE Submission to the FCC in February 1969.

2. Computer-Communications Networks by H.V. O'Neill, The Mitre Corporation (MTR 6009, Vol. 3)

GLOBAL GAMING—SIMULATION WITH COMPUTER COMMUNICATION FOR INTERNATIONAL COOPERATION

Takeshi Utsumi
Mitsubishi Research Institute, Tokyo, Japan

ABSTRACT

Thanks to the advancement of computer hardware and simulation technologies, interactive gaming-simulation has become an effective means of decision-making and personnel training in management, socio-economics and political sciences.

Up to the present, however, simulations have been performed on a real-time digital computer located at a particular university or research institute. Their usage is currently being extended to other computers at various locations through computer communication networks.

On the other hand, the computer simulation of socio-economic-political systems has been progressing rapidly in social dynamics and econometrics for business, local, national and international affairs. As the boundaries of their simulation expand making them more realistic, it is increasingly evident that the simulation models require computer communication links for the sake of resource sharing of computer hardware, data banks, simulation software and especially of research resources of know-how and man-power. These requirements are due also to the fact that the socio-economic model building, either with social dynamics or econometrics approach, requires enormous effort even for a single nation and yet, the model builder knows well the need for interaction with other nations in natural resource allocation, environmental control, foreign trade, and monetary policies.

In the present state of chaos and instability, it is a vital necessity for scientists and simulationists to cooperate not only nationally but also internationally, in order to plan ahead for the establishment of interactive gaming-simulation models on a global scale, taking advantage of social dynamics for long-range planning, of econometrics for verification and updating of models, and of political science for decision-making. The global interactive gaming-simulation proposed here will be performed with communication satellites which interface computer networks across the Pacific and/or Atlantic Oceans.

The purposes of this project are therefore:

(1) International resource sharing for the development of computer hardware, data banks and simulation software; and when the simulation model is verified,

(2) Global cooperation on the planning of natural resource allocations and environmental control,

(3) Harmonization and stability of foreign trade and monetary systems.

THE WIRED CITY

CHAIRMAN:

Weston E. Vivian, Vicom Manufacturing Company, Inc.
Ann Arbor, Michigan, U.S.A.

PROGRAM COORDINATOR:

Dr. Phillip H. Enslow, Jr.
Office of Telecommunications Policy,
Executive Office of the President
Washington, D.C., U.S.A.

PAPERS:

R. M. Alden, United Telecommunications, Inc., Kansas City, Missouri, U.S.A. *"The Wired City: The Role of an Independent Telephone Company"*

W. F. Mason and R. K. Lay, The MITRE Corporation, McLean, Virginia, U.S.A., *"The Wired City: Services for Home Delivery via Interactive Cable TV"*

John P. Thompson, Arthur D. Little, Inc., Cambridge, Massachusetts, U.S.A. *"The Wired City: Commercial Services to be Provided by Broadband Telecommunications Systems"*

DISCUSSANTS:

David Foster, President, National Cable TV Association, Washington, D.C., U.S.A.

Paul Visher, Assistant Group Executive, Hughes Aircraft Company, Culver City, California, U.S.A.

William Butters, President, Transworld Communications, Inc., New York, New York, U.S.A.

Dr. Nathaniel Feldman, Senior Engineer, Rand Corporation, Santa Monica, California, U.S.A.

THEME: Technical standards for community cable television networks recently have been published by the FCC which will require all future cable television networks in metropolitan areas to provide to and from every home, office and store, multi-megahertz-bandwidth duplex transmission capacity. In several cities today, messages, data, and private television already are being transmitted under computer control over portions of the city's cable television network.

Within this decade, most United States metropolitan areas can be expected to have such service. Over half the homes in Canadian metropolitan areas are served by cable now. As broadband two-way, city-wide transmission capacity comes into existence, new uses, new services, new communications, computer and peripheral configurations, and new computer programs will proliferate. Existing telephone installations also will provide competing services. By 1990, an investment of 30 to 50 billion dollars can be expected.

During this Session, the types of uses foreseen will be summarized and experimental tests to date described. Expected revenues, costs and capital requirements will be projected. Social and legal implications will be discussed.

THE WIRED CITY: THE ROLE OF AN INDEPENDENT TELEPHONE COMPANY

R. M. Alden
United Telecommunications, Inc.
Kansas City, Missouri, U.S.A.

The attention which is now being given to the concept of the wired city does not in the least exaggerate its importance. The long-term impact of communications on such a scale truly challenges the imagination, and provokes the farsighted into predictions of wonderful -- or terrible -- things to come. To this technical audience, however, the idea of a wired city presents a familiar dilemma -- an engineering possibility held in abeyance by financial and sociological circumstances.

Although the state-of-the-art, as they say, changes rapidly, it is already well known that the means exist to connect every residence and business in a city to sources of entertainment, knowledge, and computer power without apparent limit. It remains to be demonstrated that there are financial sources ready to pay the bill, or policy decisions ready-made to establish priorities and responsibilities.

The problem, therefore, is not so much "What do we do?", but "What comes first?", and "Who is to do it, and on what terms?" It is the purpose of this paper to explore these problems in a preliminary way, and to suggest how the telephone industry might serve the needs of the computer industry in the environment of the wired city.

The techniques for providing the necessary communications capacity will not be considered in detail, but will be mentioned briefly in order that we may draw from them some tentative conclusions which will be helpful for our principal purpose. Similarly, the uses to which communications channels are likely to be put is a subject being explored adequately by others. These uses will be mentioned here only that we may observe that they are subject to classification, leading to some pertinent observations. This paper, then, will address the organization of the cable industry and its relationship to the telephone industry.

The technical characteristics of the wired city will be derived from several considerations, among which are geography, the availability of space, electrical interference, multiplex techniques, and market demand.

We will begin by observing the obvious. The purpose of a wired city is to make accessible to each home and business an enormous reservoir of communications resources. These homes and businesses are scattered over the entire city. Conversely, the sources of the program material which they will wish to receive, and the locations of the computers or stores with which they will wish to communicate, are comparatively few in number, tending to be centralized or grouped in several suburban business centers. The physical layout, in short, resembles that of an urban telephone system, where lines from many scattered locations are brought together at one or more central points, and then the central points are interconnected by trunk lines.

In any wired city, the wires must be put somewhere, and space is not unlimited. Perhaps there is no compelling reason why two, three, or four organizations cannot share pole lines and underground conduits. Obviously, however, the limited space under city streets already shared by water mains, sewers, storm drains, gas lines, electric lines, telephone lines, traffic control cables, signposts, and structural foundations cannot accommodate a large number of additional parties without chaos.

To some visionaries, the answer lies in microwave, infrared, or laser beam transmissions, but, as a practical matter, the use of such techniques for general distribution purposes is at least questionable in the short run, and possibly beyond the horizon. For the near future, physical constraints are limiting.

The electrical characteristics of urban communications systems also deserve some comment. For safety reasons, a distinction must be made between communications and power facilities, yet most modern communications systems suitable for broadband service require access to power at regular intervals. Even if physically there is "room for one more," considerations of electrical interference impose some limitations on the design and installation of new systems.

Communications systems are conveniently classified according to their information-carrying capacity. The terms "voice frequency," "video channel," and "broadband" are too vague to be very useful in the field of computer communications. Nevertheless, because of the large quantities of these services, "voice channels" and "video channels" are convenient units with which to deal. Closely related to the capacity of communications systems is the multiplex technique employed. Are channels separated from each other by space, frequency or time? It is also useful to distinguish between systems providing one-way and two-way capability.

Among broadband services, the need for privacy and the means employed to obtain it help us to distinguish among different transmission techniques. Some services are directed to a widespread target audience, located throughout the city. Other services are of interest to small groups of customers who may be geographically scattered or may be clustered in "neighborhoods." A few scattered, individual customers form a target requiring particular techniques if privacy is to be assured. The means employed to reach audiences which are dispersed in a variety of ways has an important bearing on the efficiency of spectrum management which can be achieved.

Finally, at the risk of belaboring the obvious, all of the above considerations change with time. A picture taken at any given moment of the geographical distribution of sources and terminals; the limitations of aerial and underground space; the characteristics of the signals transmitted as to level, band width, and multiplex technology; and the grouping of customers or audience having a common need will be a valid picture for only a brief interval. When these changes take place rapidly, the obvious answer is the introduction of a switching matrix into the system design. When switching is not appropriate, then successful operation will be dependent upon a management structure and technical design which are both flexible enough to accommodate to changing conditions.

The salient technical considerations which have been identified are, of course, all applicable to yesterday's technology, as well as tomorrow's. All of these considerations have received attention from telephone companies everywhere, which leads us to our first point: We already have our cities wired. The principal issues which are raised by video and computer communications involve questions of space management, spectrum management, the flexibility required to serve a changing market, and, of course, a host of sociological and political questions relating to control, competition, and definitions of the public interest. The physical considerations all argue for assigning the task of system manager either to the telephone company or to some other entity which is organized and regulated in the same way as is a telephone company. The sociological questions -- control and "the public interest" -- have been faced before in simpler times, and the common carrier principle has been employed as an appropriate solution.

Since the conditions of the marketplace inevitably control communications system design, we should make a few observations about the services to be provided to a wired city. Without detailing them to any extent, some broad classifications can be identified.

Video entertainment is and surely will continue to be a major user of spectrum space. This will include at least three kinds of program material: (1) the distribution of off-the-air broadcast programming, (2) the distribution of prepackaged, classified program material from films and tapes, and (3) local program origination. These three types of entertainment programming place different requirements on the business organization which provides them. The first is largely a technical requirement, while the second requires an exercise of judgment about the entertainment market. Program origination is, of course, a task of production, direction, and staging, more than it is of engineering.

Video information is a second useful classification, intended to identify (1) school programs, (2) adult education, both formal and informal, and (3) advertising-oriented information. School programs are built around classroom requirements, and are an exercise in education techniques. Adult education may be similar, but often takes the form of an accumulation of useful information without formal schooling structure. For example, a course in fashion design can be distinguished from a program illustrating the latest fashion trends at Paris couturiers. Finally, the presentation of the latest fashions stocked by Jones Department Store becomes an advertising-oriented program. (From this discussion, we exclude entirely as purely commercial a program extolling the virtues of the bargains to be found at Jones Department Store.)

Another group of programs and services need only be listed without comment: shopping services, news delivery services, information retrieval, computation services. The variations on these are innumerable and need not be discussed for the purposes of this paper.

As a final classification, we must include pure communications, either person-to-person, computer-to-computer, or combinations of the two.

These classifications of communications content are admittedly arbitrary. They serve, however, a useful purpose to support our second point: For each of these services to be provided in our wired city, there is a set of identifiable technical requirements, and these differ from one service to another over a wide range. Finally, for each of these programs or services there is also a different source requirement, i.e., the techniques of entertainment program origination have little in common with those of computation services. They start from different places, they require different transmission standards, and they are offered by different people or organizations to different audiences. It is because of a coincidence of time that we lump them together as examples of the benefits to be achieved through this marvel we call "the wired city." Entertainment programming and widely dispersed computer services have, in fact, nothing else in common, and each of them has little in common with the other services to be served by this new medium of communications.

We may conclude, therefore, that we are dealing with a wide variety of physical and electrical needs to serve a diverse market within constraints which are quite similar to those faced, in our experience, by the telephone industry.

If common carriers did not exist, it would be necessary to invent them in order to serve this market efficiently. Space resources are limited, so the physical plant needed to augment that which is already in place to serve the wired city must be shared among a diversity of users and suppliers of services. The diverse talents required of the business organizations providing these programs and services are unlikely to be found in any one organization. Spectrum resources, while greatly enlarged by the wiring technique as distinct from microwave transmission, for example, can still be conserved to economic advantage by the sharing of facilities.

The use of limited, shared communications resources by many people having diverse and sometimes conflicting interests requires that the management of those resources be separated from the control of communications content. This requirement is met only by the use of common carrier procedures.

Many organizational solutions may be possible to bring about the shared use of wired city facilities, including the use of the existing common carriers. All of the technical considerations mentioned earlier have been dealt with in the experience of the telephone industry. Similarly the CATV industry has had to deal with some of these same common carriage factors, and in recent years has demonstrated, in some locations, a capacity to develop into a common carrier.

The advent of satellite communications, and of Comsat, the imminent authorization of several domestic satellite common carriers, and the recognition of the specialized common carriers by the FCC, demonstrate that the common carrier field is now open to others. At the same time the FCC, by its rulings affecting the CATV activities of telephone companies, has placed serious economic handicaps on these companies who would be willing and able to serve the needs of a wired city, as we have described those needs.

Our effort by the United Telephone System to find a marketable offering of channels for wired city services is representative of what might be done with encouragement instead of harassment by regulators. Most of the United Telephone System companies have filed Wide Spectrum Service tariffs with the FCC and with the state regulatory agencies, offering channels of various bandwidths for the transmission of information in a community-wide network. This offering reserves the right for the telephone company to assign the center frequency of each "Spectrum Unit" so as to make maximum use of the spectrum between 30 Kilohertz and 300 Megahertz which is available in coaxial cable facilities.

The United Wide Spectrum Service Tariff filed with the FCC became effective May 1, 1968. It is neither the first effort of its kind nor necessarily the best which can be devised to meet the public's need. It does, however, have some interesting and innovative characteristics.

The Tariff does not identify any specific use, or type of customer, or business. It was not developed exclusively for CATV service with use for other services being incidental thereto. United's Wide Spectrum tariff was designed to provide channels of different bandwidths from different sources to different audiences, as may be required to serve the needs of widely differing business enterprises.

The relative weighting of the costs assigned to the cable and to service connections has been designed to encourage the development of new services, and to give both the telephone company and its customer a stake in the success of the customer's business. Further, the base costs to the telephone company's customer are lower during the start-up of his business when his revenues are lowest.

Recognition has been given to the added investment in filters which will be required to furnish leased channels to several customers over the same facility. Further, the rates for Spectrum Units, or channels, decline as the number of adjacent channels ordered by a single customer increases. This declining rate structure is designed to encourage customers to use contiguous frequency bands, which conserves the available spectrum and holds down the telephone company's investment in equipment necessary to separate the "Spectrum Units" of several customers.

This offering of the United telephone companies has received little attention, and may therefore be presumed to be deficient as a full answer to the needs of the wired city. It has been described here in the expectation that it will be recognized as having certain virtues which should be incorporated into whatever means are eventually chosen to wire our cities for broadband services. Chief among these is the fact that our Wide Spectrum Tariff excludes the carrier from control of the information content of the transmissions. It is a "common carrier" offering, in every sense of the term, appropriate to the diverse needs of the market as previously described.

The choice of who should provide wired city facilities is dictated to a great extent by the need to conserve frequency and pole space and to prevent costly duplication of facilities. The consolidation of these two factors and the need to ensure equal access to these facilities by everyone argues in favor of a common carrier providing the wired city facilities. The management of the physical space would be simplified if the existing telephone company were to be chosen as the common carrier. However, this does not preclude the presence of two or more carriers, or whatever number of carriers might be determined to be in the public interest.

Telephone companies are doing other things to remain competitive as the communications market becomes larger and more diverse. They should not be ignored even if other carriers enter the scene of the wired city. Switching systems are becoming quieter and faster, and the bandwidth of available channels -- switched and unswitched -- is increasing. Four-wire switches and digital transmission systems are becoming more common, and digital switching systems will make their appearance in due course. Telephone service as we know it today can be provided on broadband coaxial distribution systems, and long range trends favor the economy of this approach.

Those who provide facilities for the wired city will necessarily be common carriers completely divorced from the control of the information content on this new communications medium. The existing telephone companies, handicapped at times by restrictive regulations, have done and will do much to be ready to serve this new market as it develops.

THE WIRED CITY: SERVICES FOR HOME DELIVERY VIA INTERACTIVE CABLE TV

W. F. Mason, R. K. Lay
The MITRE Corporation
McLean, Virginia, U.S.A.

ABSTRACT

There is every reason to believe that the use of interactive television will provide the nation with a whole new means of communication between people and between the government and the people. A great number of service ideas have been generated by the interactive TV concept and the NSF has asked MITRE to evaluate the possibilities. This paper discusses the types of services that have been selected for implementation in a demonstration system, and the rationale for their choice. These demonstrations will provide the opportunity to study the mechanisms involved in making new services effective, the community response and some of the social impacts that may result from this new technology.

MITRE has been operating an interactive TV system in Reston, Virginia, since June, 1971. This system demonstrates the delivery of several hundred services by cable on a single channel of the local cable system in Reston. A market survey has been part of this demonstration. Results and cost predictions presented in this article will be qualified and extended during the period March – October, 1972.

INTRODUCTION

A wide variety of studies have been made of the types of services that may be offered by cable TV systems. MITRE has been involved in several such studies ranging from the design of a specific state-of-the-art cable system for a large city, to experimenting with a variety of services in a specific community. This paper will summarize our experience, observations and conclusions relative to what can be done, what should be done, and what it would cost.

Television, until recently thought of as a one-way form of communications, may well become an interactive communications medium by the end of the 1970's. New types of two-way services will enter American homes as part of the implementation of urban cable systems. This will introduce a fundamental change in the pattern of national and personal communications. Before the end of the decade, these capabilities will probably be a major factor in how we structure our entertainment, social, cultural, educational, and civic processes. If its potential is fully developed, interactive TV could become a vital, if not an enabling, part of the processes that will emerge to deal with the problems and crisis of our cities and our society.

INTERACTIVE TELEVISION – WHAT IS IT?

Telephone, radio and television each brought a revolutionary change in communications. History shows that the impact and dissemination speed of each was greater than its predecessor. Interactive television will be the next, and in some ways the most dramatic, revolutionary change in media technology. Its potential for public services probably transcends those of its predecessors, and it will, no doubt, have an even greater social impact and spread even more quickly than any of these earlier forms of communications.

MITRE has been demonstrating an interactive television system called TICCIT* for the past year in Reston, Virginia. As shown in Figure 1 and 2, TICCIT uses a keyboard with each TV receiver as means to communicate with a central computer. The computer is programmed to interpret the viewer's key pushes in the context of what the viewer is watching. A key push may be interpreted in the context of public meeting straw vote as a yea or nay; in a search for a new apartment as the number of bedrooms desired; as a part of the answer to a question in a course on repairs; as a regular typewriter as part of an interactive educational program; as the keys in a desk calculator; etc. The keyboard is the viewer's means of interacting with the system in a personal way, and he gets a personalized response, as will be explained below.

The potential uses of such a system, and their importance, stem from certain characteristics:

1. Interactive television is individualized. It responds instantly to the demands of each viewer, permitting him to receive a level of detailed information economically and privately.

2. It is computerized, offering retrieval of information from a tremendous variety of sources, manipulation and computation capabilities and the use of various other computer capabilities that might otherwise be difficult or expensive to obtain in an average home.

3. It provides an unlimited number of entry and delivery points for information transfer (similar to the numbers for telephone and mail) but; in addition, it offers the above mentioned control, storage and access capabilities, plus speed and convenience of retrieval.

4. It is multimedia, providing sight, sound, and personalized computer-service all within one system, giving it the potential of becoming a new kind of common carrier.

BENEFITS AND PROFITS

Which of the many possible applications of this new communications capability should be developed first? Services that might be beneficial to the public in some social sense may not coincide with those that make a profit. It is probably safe to say that if the development of interactive TV is left solely to the forces of the marketplace, the ensuing patterns of use will not reflect the medium's full potential, and will neglect or omit some applications that have promise for producing large social, cultural, educational, and civic benefits.

For these reasons, federal agencies and private, civic and community organizations at all levels, especially at the city and community level where cable systems operate, must know and understand their options, and must get in on the ground floor in establishing requirements, setting priorities, and making decisions about fostering and encouraging uses that might not be provided by normal market forces. So on the one hand, planning and decision-making cannot be deferred, but on the other hand the kinds of technical information, economic data, and empirical information that are needed for intelligent planning and decision-making, do not now exist.

Recognizing this, the National Science Foundation has funded MITRE to select a set of services and to develop specifications and recommendations for terminal hardware, the configuration of systems, and criteria for the selection of a site or sites for conducting meaningful tests and demonstrations. The study stage has set forth a range of possible services to be tested. The next phase will produce economic data that will illuminate the trade-offs involved; the costs of various combinations of services and functions, the impact of various levels of service and subscriber fees on system configuration requirements, and the relative profitability of these combinations.

---

*Time Shared, Interactive Computer Controlled Information Television

WHICH SERVICES AND WHY?

Discussions of the future of cable television often present glowing pictures of potential two-way communications services on cable. Extensive lists of new services have been compiled. Commercial services such as security alarm monitoring, remote shopping, and computer-to-computer data transmission have been cited as important to the economic success of cable television in cities. In addition, it is widely held that such noncommercial services as interactive educational television in the home and direct citizen feedback on local political issues, would be of great public benefit.

As is often the case with a newly emerging technical possibility, however, the promise of two-way services on cable has at times been oversold. Although most proposed new services are technically feasible, only a very few look economically attractive at the present time. Neither are there a set of services that operators think would sell well. But while there are remarkably few new kinds of cable services that are recognized as universally attractive: we have a situation that may not call for the kind of unanimous endorsement that was needed for color television. Cable provides such an abundance of bandwidth that we can now consider the market for channels somewhat analogous to the market for magazines, in that every possible interest can be served and the summing of many special interest markets may be the secret to economic success. Not only does cable provide a large number of channels into each home, but techniques are being demonstrated to time share within a channel, to allow a single channel to be effectively "expanded" to serve hundreds of users with special interest programming on an individualized basis—thus providing means to deliver an even wider range of subject material than the many channels can now provide. We call this latter technique "jerky video" at MITRE where we are developing the idea to allow a selection of hundreds of educational courses.

Another major dimension that should be considered in any reasonable analysis of cable's potential involves the use of the broadband communication system to provide services that do not involve home TV sets, but which are nevertheless enabled by the existence of the cable network throughout the urban environment. In the urban system design developed by MITRE, the cable provides the vital low cost communications needed for computer controlled traffic density throughout the city, an ability to monitor the location of police, fire, transit and other public vehicles for various management purposes, plus the ability to provide improved radio communications to these vehicles with far less frequency spectrum than is required using present techniques.

A separate grid-like network of cable, overlaying the usual tree-like net, can provide point-to-point connections between banks, educational institutions, police and other municipal facilities and commercial establishments. The need for these kinds of services is evident from the 140 license applications for dedicated communications capabilities within special organizations from the companies in 89 cities.

CLASSIFICATION OF SERVICE POSSIBILITIES

There have been two general methods exercised in the classification of service options. The more common is to summarize types of services in such functional areas as education services, health, law enforcement, and community advisory services.

The classification MITRE uses is based on the technical characteristics required to provide the services. For example, a wide variety of services can be provided via conventional one-way cable. Another whole class of services comes with the introduction of two-way and polling capabilities. Still more services can be provided with completely interactive computer capabilities and of course there is a group that simply makes use of a transponder type terminal in the home connected to a variety of sensors, e.g., fire alarms, meter readers.

This type of classification is presented in Figure 3. Here we see that the basic capability that cable provides is more channels and that a variety of home terminals can be used to improve the utilization of these channels. Figure 4 gives the approximate costs of the specific devices associated with each class of system.

Because of the lack of any market tests of these services, not only is the investment community reluctant to move decisively to support expansion in this area, but government planners from municipal to federal, from education planners to telecommunication regulators, are also stymied by a total lack of information on how citizens might respond to this new media. It seems likely that the commercial tests expected in the next one to three years will probably give initial information on the market for the so-called commercial services. We believe the same should be done for the non-commercial services, such as interactive educational television for the home.

Since it will take so much time before mass test and social feedback occur, NSF, using the MITRE Corporation, is exploring the feasibility of a quick thrust implementation in one or a few communities in order to significantly decrease the knowledge gap in this area. The Department of Housing and Urban Development is also using MITRE to help plan some large sale demonstrations of some of the public interest possibilities of cable TV. An Appendix lists the services being considered for these demonstrations.

CONCLUSIONS

The types of services being explored for delivery via interactive cable TV discussed in this paper represent the summing of many relatively small special interest markets. Preliminary surveys indicate the potential subscriber interest shown in Figure 5. Obviously, the development of demand and impact that occurs during the next year will be very much a function of the trial community. It will be of little surprise if the survey of those services people imagine they will employ differs considerably from those actually used. Hence, the important products of the demonstration will be measures of how services become popular and have social impact as well as which ones are preferred by that particular community. It is our goal to understand the mechanisms of this socio-communication revolution: to draw specifics from what has previously been imaginative conjecture.

APPENDIX

Examples of services representing an initial offering from which the community will choose candidates for trial.

(1) Sit in on a meeting of the city council and not only watch the proceedings, but also directly participate while the meeting is in session.

(2) Participate in a preference pool providing feedback not only regarding government policies and services, but also TV programming.

(3) Express views on any topic and have access to the opinions and expressions of other individuals or groups.

(4) Review candidates' positions and records immediately before going to the polls at election time.

(5) Receive information on available social services.

(6) Prepare income tax forms interactively with programs written by experts, or learn how to fill out an application for welfare.

(7) Feel a greater sense of security and protection because of new crime prevention and detection services.

(8) Distribute employment information that is clear and well-defined including job descriptions, locations, accessibility by public transportation, as well as how to apply for the job using interactive, personalized information retrieval techniques.

(9) Raise the level of education in order to improve chances for employment or make leisure time more stimulating. Skill training, literacy courses, high school equivalency programs, and college courses can all be available.

(10) Establish a sense of local, regional and national identity through a sharing of local community information, meetings of civic groups, new books in the library, adult education programs, recreational facilities will be available to outlying areas. The pooling of local recipes, folk crafts, language courses and social activities can encourage more intermingling between areas while also preserving a sense of community identity and pride.

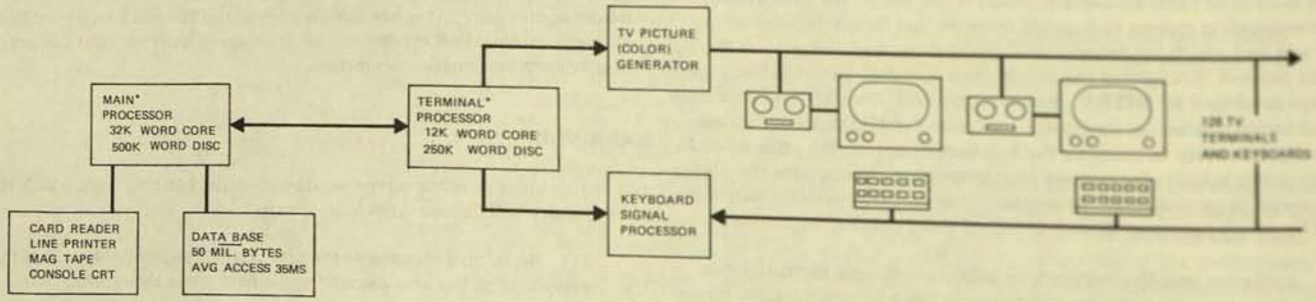
(11) Establish closer relationships between neighbors, and shut-ins using games via the computer, closed circuit TV, visits via the video screen.

(12) Create ties within special interest groups. Share hobby tips on gardening, fishing, knitting.

(13) Have available up-to-date information on the public transportation system, including public transportation schedules, automatically calculated routes and fares, and current traffic conditions.

(14) Use of secretaries in homes to support offices via cable.

FIGURE 1
TOUCH TONE PHONE LINK IN USE AS A SIMPLE INPUT KEYBOARD TO RESTON, VIRGINIA, HOME DELIVERED INTERACTIVE TV DEMONSTRATION



*TWO 16 BIT MINICOMPUTERS
800 NS MEMORY CYCLE
EACH WITH TWO INDEPENDENT FIXED HEAD DISCS
FAST ACCESS (2MS AVG LATENCY, 100,000 BYTES/SEC TRANSFER RATE)

FIGURE 2
DEMONSTRATION SYSTEM DIAGRAM

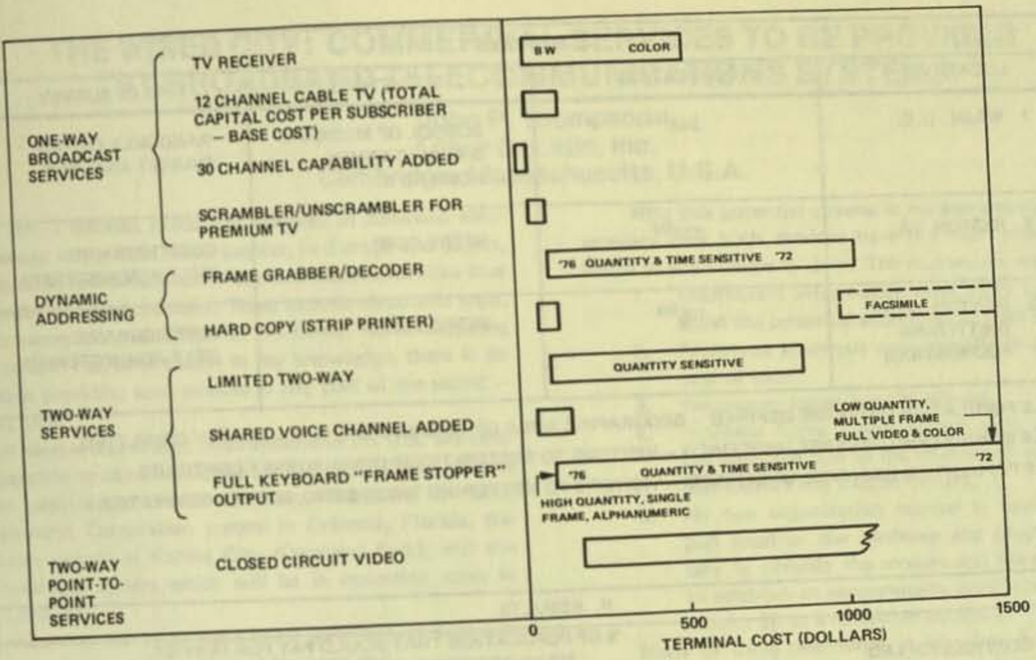
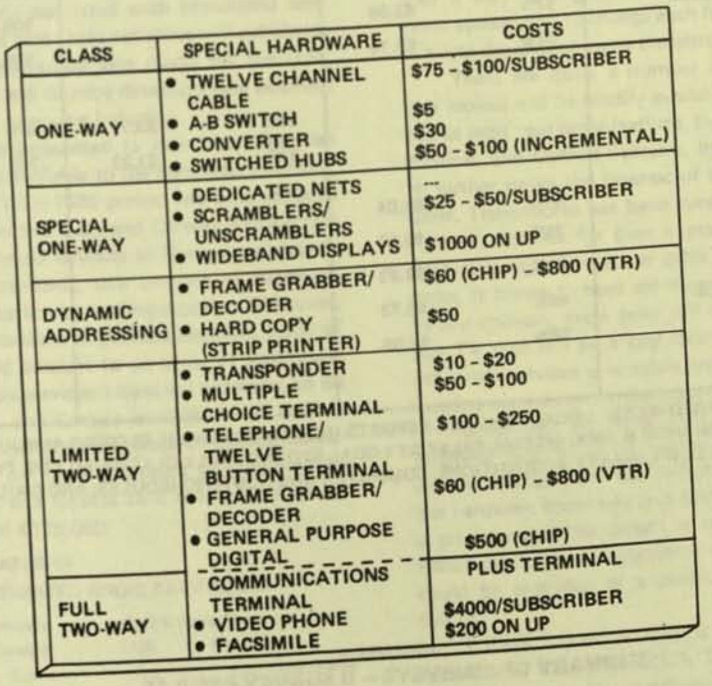


FIGURE 3
PROJECTED PRODUCTION COSTS FOR SUBSCRIBER TERMINAL AND NEW CABLE COMMUNICATIONS SERVICES



CLASS	SPECIAL HARDWARE	COSTS
ONE-WAY	• TWELVE CHANNEL CABLE • A-B SWITCH • CONVERTER • SWITCHED HUBS	$75 - $100/SUBSCRIBER $5 $30 $50 - $100 (INCREMENTAL)
SPECIAL ONE-WAY	• DEDICATED NETS • SCRAMBLERS/ UNSCRAMBLERS • WIDEBAND DISPLAYS	— $25 - $50/SUBSCRIBER $1000 ON UP
DYNAMIC ADDRESSING	• FRAME GRABBER/ DECODER • HARD COPY (STRIP PRINTER)	$60 (CHIP) - $800 (VTR) $50
LIMITED TWO-WAY	• TRANSPONDER • MULTIPLE CHOICE TERMINAL • TELEPHONE/ TWELVE BUTTON TERMINAL • FRAME GRABBER/ DECODER • GENERAL PURPOSE DIGITAL	$10 - $20 $50 - $100 $100 - $250 $60 (CHIP) - $800 (VTR) $500 (CHIP)
FULL TWO-WAY	COMMUNICATIONS TERMINAL • VIDEO PHONE • FACSIMILE	PLUS TERMINAL $4000/SUBSCRIBER $200 ON UP

FIGURE 4
APPROXIMATE COSTS OF DEVICES ASSOCIATED WITH EACH CLASS OF SYSTEM SERVICE

I. SURVEYS			
LOCATION	POPULATION	SURVEYORS	TYPE OF SURVEY
1. WASH., D. C.	248(a)	SCHOOL OF MEDICINE & PUBLIC ADMIN. HOWARD U.	RANDOMLY DRAWN MARKET ANALYSIS
2. RESTON, VA.	279(b)	MITRE CORP.	QUESTIONNAIRE (SELF-ADMINISTERED)
3. DAYTON, OHIO (KETTERING FOUNDATION)	105(c)	MITRE CORP.	QUESTIONNAIRE (SELF-ADMINISTERED)

(a)SAMPLE POPULATION UNIVERSE DEFINED – GEOGRAPHIC AREA OF WASHINGTON, D. C. – SURVEY COMPLETED

(b)SAMPLE POPULATION UNIVERSE UNDEFINED – VISITORS TO RESTON TICCIT DEMO; SURVEY CONTINUES

(c)SAMPLE POPULATION UNIVERSE UNDEFINED – VISITORS TO KETTERING TICCIT DEMO; SURVEY COMPLETED

II. RESULTS						
SERVICES POLLED	% OF POPULATION THAT WOULD PAY FOR SERVICE, MEAN PROJECTED PAYMENT FOR SERVICE					
	D. C.		RESTON		KETTERING	
INFORMATIONAL* ONE-WAY SERVICES	63%	$2.18	75%	$2.08	51%	$1.98
CALENDAR OF EVENTS			72%	$ .79	53%	$ .70
ADDRESS AND PHONE BOOK			33%	$ .80	26%	$ .55
PERSONAL STOCK PROFILE			41%	$1.60	30%	$1.81
BURGLAR AND FIRE ALARM	67%	$2.86			90%	$1.09
REMOTE SHOPPING	52%	$2.10			50%	$1.04
LIBRARY SERVICES					55%	$ .78
CALCULATOR			64%	$3.77	67%	$2.90
CAI – RESTON P/HR. KETTERING P/MO.			94%	$1.22	76%	$6.09
IMPROVED PICTURE	68%	$2.84				
NEW MOVIES	75%	$2.67				
SPORTS	66%	$2.80				
EDUCATIONAL CHANNELS	66%	$2.73				
POLLING	58%	$2.00				

*ONE-WAY SERVICES OFFERED WERE: LOCAL WEATHER REPORTS INCLUDING SCHOOL CLOSING ANNOUNCEMENT; MOST ACTIVE STOCKS/PRICES ON NYSE & AMX; SPECIAL SALES AT LOCAL STORES; DAILY CLASSIFIED ADS; TV GUIDE LISTINGS; SKI CONDITION REPORTS; NEW LIBRARY ACQUISITIONS; COMMUTER BUS ROUTE SCHEDULES; AND DAILY SCHOOL MUNCH MENUS.

FIGURE 5
SUMMARY OF I SURVEYS – II SURVEY RESULTS

THE WIRED CITY: COMMERCIAL SERVICES TO BE PROVIDED BY BROADBAND TELECOMMUNICATIONS SYSTEMS

John P. Thompson
Arthur D. Little, Inc.
Cambridge, Massachusetts, U.S.A.

For the last 3 decades executives involved in advanced electronics technology in the U.S., and I suspect, in Europe and Japan, have been discussing new broadband telecommunication services that could be provided to the household. These include electronic mail, the electronic newspaper, education in the home, home shopping services and others. Yet, as of today, to my knowledge, there is no full-scale system providing such services in any part of the world – That's the bad news.

The good news is that at least three systems in the U.S. are now actively in operation to demonstrate the commercial viability of such services when provided to the homeowner. These are the EIE/American Television Corporation system in Orlando, Florida, the Telecable/Vicom system in Kansas City (Overland Park), and the Thetacom/Thetacable system which will be in operation soon in El Segundo, California.

Each system now has 15 or more home terminals, at least one minicomputer, and disc-storage system at the head end.

If successful, each system plans expansion of terminals to provide for viable economic operation. In addition it has been recommended to the Federal Government that it install 1,000 relatively sophisticated terminals in households in one geographical area. It is also recommended that a comparable number of terminals be installed in public institutions such as education, law enforcement, and others and in business organizations and to allow these organizations to carry out tests over the hardware network.

Many of us who are closely associated with broadband telecommunications hope and expect that these activities will culminate in the development of a new communications media for the U.S., Canada, Europe, and Japan, and will thereby develop a new business for the worldwide electronic and computer industry.

You will undoubtedly be disappointed to hear that, based on my analysis of the opportunities available to the computer industry, the market for hardware in the 1975-1985 period will only amount to an estimated 36-54 million in the U.S., and Canada. During this period, a comparable market should develop in Europe and Japan. By most computer industry standards, this amount of hardware volume is hardly worth consideration by leading suppliers. However, the software required for the provision of potential services, some of which I will discuss later, could amount to an order of magnitude more than the potential hardware services. I base my estimates on 83 million TV homes in the U.S., and Canada in 1985, 40-65 million CATV subscribers, each system serving an average of 25,000 households. (See Table 1A.) Based on these estimates there would be 1200-1800 systems in the U.S. and Canada each having a computer and peripheral head end value of $125,000.

TABLE 1A

U.S. AND CANADA ESTIMATED WIRED CATV HOMES

	No. of TV Homes		CATV Subscribers		
	U.S.	Canada	U.S.	Canada	
January 1, 1972	62.2	5.5	6.2	1.5	(millions)
January 1, 1975	66.1	6.4	10-13	2.0-2.3	
January 1, 1980	71.0	7.2	20-30	3.9-4.2	
January 1, 1985	75.0	8.2	35-60	5.0-5.5	

Sources: TV Fact Book, ADL estimates, and Canadian Bureau of Statistics.

In addition, I estimate that there may be an average of $200-400 worth of hardware required per household to provide these services. If 30-45 million homes were furnished with broadband terminals, this would add another $6 billion to $18 billion on a cumulative basis in hardware to the North American computer and electronics industry over this 10-15 year period.

With this potential volume in hardware businesses available, one wonders why such services have not been provided to the homeowner in any volume to date. The reasons are many but include:

1. Insufficient information until 1971 to better understand the potential economics of such services.
2. Prototype terminals were unavailable up until the last year or two.
3. Regulatory issues were ill defined.
4. Potential home entertainment such as movies were not available due to the attitude of movie producers, distributors and theater owners.
5. No one organization seemed to have the ability to pull together the hardware and programming necessary to provide the services and the money required to establish an economically viable system which may require $5 to $15 million in equity.

Many of these deterrents, in my view, have either been overcome or should be overcome in the next year or two.

A highly encouraging factor in the development, or toward the development of these new communication networks is that the movie industry's attitude toward providing films to the homeowner over a cable in addition to in the theater seems to have changed almost 180 degrees since the referendum in Los Angeles in 1964 that declared subscription television to be "unconstitutional." The movie industry is now coming to realize (and many involved are convinced) that it may be highly desirable to add thousands of viewers via a cable system and to charge each household for each movie presented, at a rate depending upon the attractiveness of the movie.

There are quite a number of problems to be overcome before new movies will be readily available to be shown over CATV systems, but at least two organizations, the Trans-World Division of Columbia Pictures, and Optical Systems, Inc., are now developing test systems to further prove the interest of the public in movies provided over a cable. Trans-World has been running a system in the Regency Hyatt Hotel in Atlanta for over a year, and by October 1972 should be operating experiments on cable systems in more than six different cities. It hopes to have six more on stream at a rate of one a month. In my opinion, these tests will prove the interest of homeowners in pay TV and will be a key factor in stimulating the development of other new services over cable systems as well as improve saturation of existing systems and installation of new systems.

Work that we have been carrying out at ADL in analyzing potential services to be carried over cable networks has been based on the premise that the fastest and most efficient way to establish the hardware terminals and cable networks to provide these services is private enterprise, based on investing for profit. Further, if we can install the hardware necessary to provide initial basic services that could be provided at a profit, then other profitable services will follow.

I think a good analogy is the installation of the first electrical networks in the early 1900's. The objective here was to light homes and cities, and the installation of the network was justified on lighting homes and cities, alone. Few seers, if any, at that time would have foreseen the air conditioning, disposal, dishwasher, heating, and many other applications which we now have to serve us based on the installation of this initial network for lighting. Therefore, our initial objective on new broadband communication services has been, and still is, to get the "lighting network in."

This network could initially provide the following services and be highly profitable. I want to make it clear, however, that the detailed information developed for sponsors of the Arthur D. Little, Inc., broadband communications network project are not these specific series of services nor were the specific estimates of revenue developed by the study. The key objective of all sponsors was to assist in the development of these networks in North America for the public and the development of a new industry. The numbers are my own estimates.

I suggest that the services to be provided initially consist of:

1. A 60 channel audio service to provide uninterrupted stereo music and other audio programming in the home, such as the service being tested by Cablenet International Corporation in Columbus, Ohio.

2. An alphanumeric readout on the home broadcast TV receiver such as now being operated on a trial basis in Reston, Virginia.[1]

Specific services to be provided include:

- a. Purchasing
- b. Rapidly updated news services.
- c. Basic rote-type education services and others.

3. Moving pictures, for which the viewer would be charged based on the specific picture he viewed and the amount of time he viewed it.

4. A one channel shopping[2] service such as that operated by Cox and Young and Rubican in Warner Robins, Georgia, but with a response capability such as that now being tested by Sterling Manhattan in New York City.

Now let us assume that subscribers will pay $5 per month or the cost of one stereo album for the Cablenet sound service, $8 per month for the viewing of movies and $5 per month for the Reston type services. The shopping service would be totally funded by advertisers.

These numbers appear reasonable to me. If they don't to you try your own. However, in a survey carried out in early 1971 by a major west coast consulting organization which had interviewed 1200 household owners picked on a random basis in California, it was found that individuals interviewed would spend in excess of $20 per month per household on such services. I suspect, however, that interest in services will vary in different geographical areas in North America and the amount that people will pay for such services will also vary.

Figure 1 shows the estimated income from subscribers and from advertisers. Advertising income is based on an average of $320 spent per year per household in the United States.

The income figures assume that 5,000, 10,000 or 20,000 subscribers will be "on line" 3 years after installation of these new services over a system and that an average of 50% would be "on line" the first year and 75% the second year of operation.

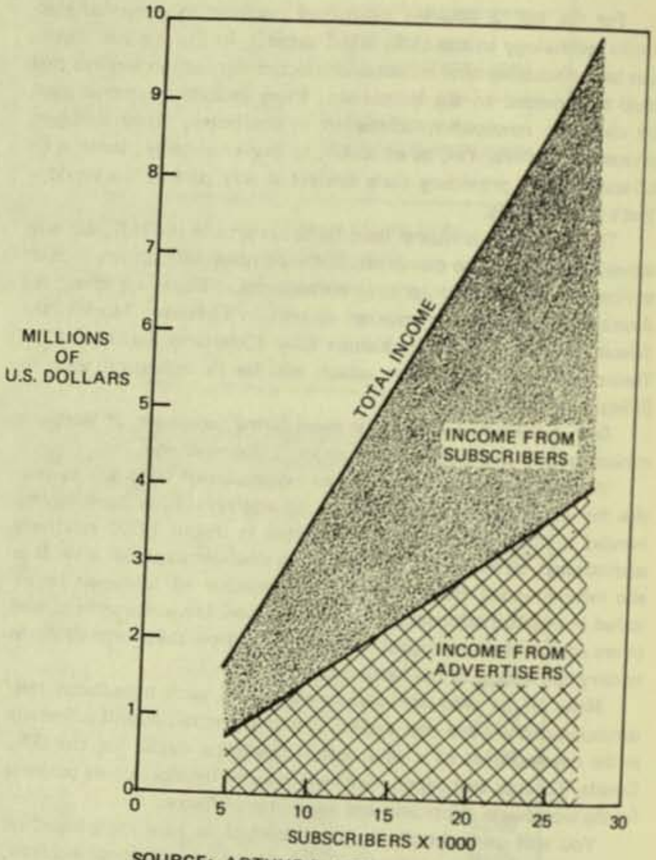
The hardware proposed consists of head end equipment, including two minicomputers and two disc packs, having a capacity of about 75 million characters to provide alphanumeric information to the viewer at his request through his own television set. Storage would be required at the home terminal which would consist of electronic storage of 1000 characters, 500 of which would be viewed at the request of the viewer.

In addition, at the head end there is an audio storage system which would store up to 1,000 selections of stereo music or other audio programs which would be presented through 60 channels in the FM band of the cable and played to the customer through his own stereo hardware equipment.

A control panel in the home would allow the viewer to:

1. Vote
2. Ask for further information
3. Call up information from the head end or interact with his TV set

Four video bandwidth channels would be required to provide the described services. $22,000 is included in the estimated operating and programming costs shown in Table 1 to lease each channel annually.



SOURCE: ARTHUR D. LITTLE, INC.

FIGURE 1

TABLE 1

ESTIMATED OPERATING & PROGRAMMING COSTS BY YEAR
($000)

Year	1	2	3	4
5,000 Subscribers				
Cablenet	60	60	60	60
Shopping	250	250	250	250
Reston	–	260	260	260
	310	570	570	570
Movie	–	275	325	375
Total	310	845	895	945
10,000 Subscribers				
1st 3 Services	310	570	570	570
Movie	–	350	450	550
Total	310	920	1,020	1,120
20,000 Subscribers				
1st 3 Services	310	570	570	570
Movie	–	550	750	950
Total	310	1,120	1,320	1,520

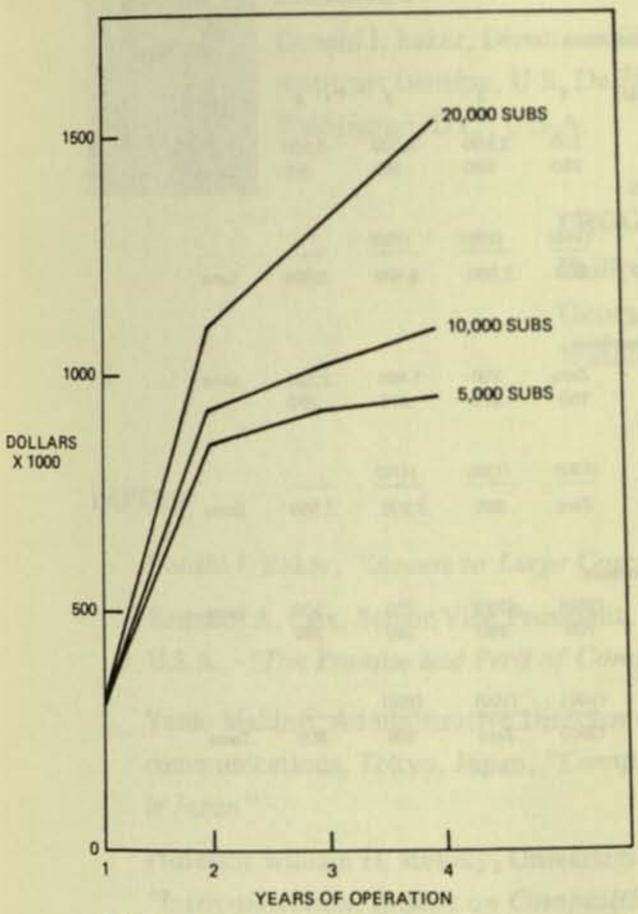
Source: Arthur D. Little, Inc.

---

[1] See Electronics September 27th, 1971, p. 45.

[2] This is being tried on a limited basis (12 terminals) by Sterling Manhattan Cable. See "Media Decisions," August 1971, page 51.

The table shows costs for the first 4 years of operation for systems of 3 different sizes, 5,000, 10,000 and 20,000 subscribers. Also see Figure 2.



SOURCE: ARTHUR D. LITTLE, INC.

FIGURE 2 ESTIMATED OPERATING AND PROGRAMMING COSTS

General administration and selling costs are $400,000 annually regardless of the number of subscribers.

Table 2 shows the estimated capital required for each of 4 proposed services. Provision of movies is as low as $25,000 only because the Reston type hardware is considered to be installed. Capital costs to provide movies so that one could bill the subscriber for what he is watching and for how long would be about $20 per subscriber. In addition, more channels could be provided for movies as demand developed.

Two services, the Reston and movie, would not start until the second year of operation.

TABLE 2

ESTIMATED CAPITAL OUTLAY
($000)

Year	1	2	3	4
5,000 Subscribers				
Cablenet	15	–	–	–
Shopping	1,000[1]	–	–	–
Reston		1,700	–	–
Movie		25	–	–
Total (rounded)	1,100	1,800	–	–
10,000 Subscribers				
Cablenet	15	–	–	–
Shopping	1,500[1]			
Reston		2,100	–	–
Movie		25	–	–
Total (rounded)	1,600	2,200	–	–
20,000 Subscribers				
Cablenet	15	–	–	–
Shopping	2,500	–	–	–
Reston	–	3,000	–	–
Movie	–	25	–	–
Total (rounded)	2,600	3,100	–	–

[1] Includes retrofitting cable to two way system.

Source: Arthur D. Little, Inc.

A projected cash flow for 5-10 and 20,000 subscribers systems is shown in Table 3. The 20,000 subscriber system returns its full investment in a little over 2-1/2 years and then pays a 100% return each year for the next seven years. Taxes before income have not been deducted, however, so don't react too enthusiastically. You'll note that the 5,000 and 10,000 subscriber systems are less attractive. On the other hand just the Cablenet audio and shopping service could be attractive on a 5,000 subscriber system above.

Each of the four services suggested will, in my opinion, require automatic billing and verification systems. The Reston service will require an 80-120 million character data bank with access times varying from one second to three to four seconds and with interrogation frequency ranging from once a second to once a year. Therefore, various types of storage will undoubtedly be required ranging from film to magnetic.

In conclusion, in my opinion, the electronics and computer industries have a great opportunity throughout the work to provide a major new communication facility for the public and in turn, generate a major new market for their products and services. There are many obstacles to be overcome before these markets develop, which will require the dedication and extra effort of many. The rewards are commensurate with such effort and I urge you to do whatever you can to see that the good news continues to far outstrip the bad news.

TABLE 3

PROJECTED CASH FLOW
($000)

Year	Prior to Year 1	Year 1 Capital	1	2	3	4	5-10
20,000 Subscribers							
Net Income Before Taxes			700	2,500	5,000	5,500	
Depreciation & Amortization			250	560	560	560	Same
Hardware & Software Investment	2,515	3,025					
Pre-Operating Expense	40						
Working Capital			(155)	(155)	(190)		
Total (rounded)		5,600	800	2,900	5,400	6,000	Same
10,000 Subscribers							
Net Income Before Taxes			Zero	700	1,900	2,200	Same
Depreciation & Amortization			150	370	370	370	
Hardware & Software Investment	1,515	2,125					
Pre-Operating Expense	40						
Working Capital			(130)	(130)	(170)		
Total (rounded)		3,700	Zero	900	2,200	2,500	Same
5,000 Subscribers							
Net Income Before Taxes			(350)	(200)	250	500	Same
Depreciation & Amortization			100	280	280	280	
Hardware & Software Investment	1,015	1,725					
Pre-Operating Expense	40						
Working Capital			(100)	(100)	(150)		
Total (rounded)		2,800	(350)	Zero	400	800	Same

Source: Arthur D. Little, Inc.

THE ROLE OF COMPETITION

CHAIRMAN:

Donald I. Baker, Director of Policy Planning,
Antitrust Division, U.S. Department of Justice,
Washington, D.C., U.S.A.

PROGRAM COORDINATOR:

Philip M. Walker
Georgetown University, Law Center
Washington, D.C., U.S.A.

PAPERS:

Donald I. Baker, *"Access to Large Computer Systems"*

Kenneth A. Cox, Senior Vice President, MCI Communications Corporation, Washington, D.C., U.S.A., *"The Promise and Peril of Competition in Intercity Communications"*

Yasuo Makino, Administrative Director of Telecommunications, Ministry of Posts and Telecommunications, Tokyo, Japan, *"Competition in the Fields of Computers and Communications in Japan"*

Professor William H. Melody, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A., *"Interconnection: Impact on Competition-Carriers and Regulation"*

Charles R. Cutler, Kirkland, Ellis & Rowe, Washington, D.C., U.S.A., *"Beyond the Computer Inquiry: Who Should Be Regulated in Computer/Communications"*

DISCUSSANTS:

Philip M. Walker, Georgetown University, Law Center, Washington, D.C., U.S.A.

Lt. Col. Sebastian Lasher, Office of Telecommunications Policy, Executive Office of the President, Washington, D.C., U.S.A.

THEME: This session will explore the role and implications of competition in the computer-communications industry, focusing on the following aspects of the subject:

Access to large computer-communications systems—efficiency versus competition.

Competition among common carriers—specialized carriers and domestic satellites.

Competition in the remote-access data processing services industry—comparison of the U.S. experience with that of other countries.

Competition and carrier performance—a case study in interconnection.

THE ROLE OF COMPETITION

CHAIRMAN

Donald I. Baker, Director of Policy Planning,
Antitrust Division, U.S. Department of Justice,
Washington, D.C., U.S.A.

PROGRAM COORDINATOR

Philip M. Walker
Georgetown University Law Center
Washington, D.C., U.S.A.

PAPERS:

Donald I. Baker, "*Access to Large Computer Systems*"

Kenneth A. Cox, Senior Vice President, MCI Communications Corporation, Washington, D.C., U.S.A., "*The Promise and Peril of Competition in Intercity Communications*"

Yasuo Makino, Administrative Director of Telecommunications, Ministry of Posts and Telecommunications, Tokyo, Japan, "*Competition in the Fields of Computers and Communications in Japan*"

Professor William H. Melody, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A., "*Interconnection: Impact on Competition, Carriers and Regulation*"

Charles R. [illegible], Kirkland, Ellis & Rowe, Washington, D.C., U.S.A., "*Beyond the Computer Inquiry: How Should the Boundary be Redrawn for Computers/Communications*"

DISCUSSANTS:

Philip M. Walker, Georgetown University Law Center, Washington, D.C., U.S.A.

Lt. Col. Sebastian Lasher, Office of Telecommunications Policy, Executive Office of the President, Washington, D.C., U.S.A.

THEME: This session will explore the role and implications of competition in the computer-communications industry, focusing on the following aspects of the subject:

Access to large computer-communications systems—efficiency versus competition

Competition among common carriers—specialized carriers and domestic satellites.

Competition in the remote access data processing services industry—comparison of the U.S. experience with that of other countries

Competition and carrier performance—a case study of interconnection

ACCESS TO LARGE COMPUTER SYSTEMS

Donald I. Baker
U.S. Department of Justice, Washington, D.C., U.S.A.

We all live in the shadow of the computer. This is just a basic fact of life for citizen, businessman and consumer. The computer, with its growing capability and widening range of uses presents all kinds of interesting new practical and legal issues. I propose to discuss one of these.

One can see a clear trend toward larger and larger computer systems, since there are important economies of scale in computation. As part of this, we see increasing use of what has come to be called remote access data processing. This has been combined with a movement towards ever more complicated and specialized programs. Thus many computer services are no longer locally based -- but rather involve a local communications link to some distant computer or data bank. The stock market quotation systems and airline reservation systems are among the more familiar to the general public, but the same thing is going on with other less visible computer systems. Thus we increasingly are seeing the creation -- often on a national scale -- of unique data processing systems based on remote access.

The Department of Justice has actively encouraged this course in a general way, by seeking to have the Federal Communications Commission eliminate many of the ancient common carrier tariff restrictions which rendered the telephone system less useful for these new modes.[1] The FCC's decision in this field was broadly along the lines urged by the Department.[2]

From the standpoint of antitrust, it is desirable that these remote access data processing services be offered on a competitive basis. Competition is the cornerstone of our national economic policy. This is because competition offers the most effective way of assuring that suppliers respond quickly to actual demands of customers, with prices based on the cost of providing service. The broad presumption in favor of competition has often been reiterated by Congress:

> "The essence of the American economic system of private enterprise is free competition. Only through full and free competition can free markets, free entry into business, and opportunities for the expression and growth of personal initiative and individual judgment be assured."[3]

and the Supreme Court:

> "Subject to narrow qualifications, it is surely the case that competition is our fundamental national economic policy, offering as it does the only alternative to the cartelization or governmental regimentation of large portions of the economy."[4]

In many areas, it will probably be possible to offer remote access and other data processing services on a competitive basis; and, so far as these services are concerned, normal antitrust rules will apply. Thus agreements among competitors in a market with regard to the price of services or the territories or customers each will serve would be treated as *per se* violations of Section 1 of the Sherman Act which bars restraints on competition.[5]

On the other hand, as we go increasingly to larger and more specialized computer systems, combined with the advantages of remote access, we can expect to see an increasing number of systems which are effective monopolies for customers requiring the highly specialized services that each system offers. This development raises the whole antitrust question of fair access.

Where a monopoly position is secured by improper means, this is subject to prosecution under Section 2 of the Sherman Act.[6] Thus the courts have punished monopolies which have been achieved or maintained with the assistance of mergers, tie-ins, predatory practices, and the like.[7] Illegal monopolies are subject to a range of equitable remedies, including divestiture.

That is not the situation I am talking about. I am talking about the situation where a computer-based data processing or information service has a legal monopoly in a specialized field -- either because the market is a natural monopoly market or because of the entrepreneur's skill, foresight and industry.

The general rule that has emerged is that a monopolist, unlike a normal businessman, cannot refuse to deal in order to maintain or extend his commercial position.[8] A related formulation of what is essentially the same principle is the rule that those controlling access to an essential facility must grant access on reasonable and non-discriminatory terms to all in the trade.

The latter rule grew up in connection with local transportation facilities.[9] It was subsequently applied to a whole series of local produce markets.[10] In more recent years, the same general principle has been applied to national institutions such as the Associated Press [11] and the New York Stock Exchange.[12] What is required is that there be some unique resource under the control of the defendant.

The *Associated Press* case is probably the leading authority on the character of competitive advantage which, if denied to competitors, may give rise to an unreasonable restraint of trade. The Associated Press was a cooperative news-gathering service, which prohibited its newspaper members from making news gathered by themselves available to non-members prior to publication by Associated Press.

Moreover, Associated Press's membership requirements were restricted and discriminatory. An applicant in any area already served by an existing AP member was faced with an effective barrier to membership (although the admission of a non-competing applicant was relatively easy). These exclusionary arrangements did not commend themselves to the Supreme Court. AP argued that, since other wire news services were available to non-members and because non-member papers could gather their own news, AP service was not strictly necessary in order to operate a competing newspaper. The Supreme Court squarely rejected that argument. It held that the inability to buy news from the largest and, in the opinion of some, the best news agency, or one of its members, could not only seriously affect the publication of competitive newspapers, but raised barriers to new entry as well.[13]

It is clear, therefore, from this aspect of the Associated Press decision that the competitive advantage need not be indispensably necessary to competitive survival, but it is sufficient that without it the excluded competitor is at a significant competitive disadvantage.[14] This was stressed by Judge Learned Hand for the three-judge District Court in Associated Press, in a passage quoted with approval by the Supreme Court:

> "Most monopolies, like most patents, give control over only some of the means of production for which there is a substitute; the possessor enjoys an advantage over his competitors, but he can seldom shut them out altogether; his monopoly is measured by the handicap he can impose. . . . And yet that advantage alone may make a monopoly unlawful."[15]

The Associated Press case involved collective activity, and can be interpreted as limited to such activity. But the heart of the problem was that the collective effort tended to perpetuate AP's monopoly power and that of its members -- the very end condemned by Section 2 of the Sherman Act.[16]

The most recent application of the general access principle has occurred in a case now on appeal to the Supreme Court in the electric power industry. This Government case was directed against a variety of practices by Otter Tail Power Company, a regional power company in Minnesota.[17] The District Court found that the defendant had a monopoly position in the transmission and sale of electric power within its region, and that it used this position to perpetuate its monopoly by preventing certain small communities from terminating its retail power franchises. It did so by refusals to sell power to proposed municipal systems and by refusals to wheel, i.e., transmit power across its transmission system for such proposed systems. The defendant was also found to have used vexatious litigation to prevent the establishment of these systems. The District Court found that these activities constituted illegal monopolization under Section 2 and ordered the defendant to sell wholesale power and provide wheeling service to municipalities within its area.

The District Court stressed that "it is well established that the unilateral refusal to deal with another, motivated by a purpose to preserve a monopoly position, is illegal."[18] The court also applied what it referred to as the "bottleneck" theory of antitrust law -- i.e., that it is an illegal restraint of trade for a party to foreclose others from the use of a scarce facility. The court applied both these related principles to Otter Tail's transmission facility, which clearly was an effective local monopoly and a scarce facility.

The rationale for the access and refusal to deal rules is not hard to find. A group of firms, or even a single firm, which controls a scarce facility can use it as a means of foreclosing competition with those who require use of that facility. Thus, for example, in Otter Tail, the finding was that the power company had used its control of wholesale transmission as a means of foreclosing new competition in local retail power supply. In Eastman Kodak, the defendant enjoyed a monopoly in photo supplies, and it refused to sell to the plaintiff (a photo supply dealer) at the usual dealers' discounts in order to promote the business of its own supply houses. In Associated Press, the restrictive membership rules were designed to favor each local AP member vis-a-vis any other newspapers in their same local areas.

I firmly believe that these principles will be applied to computer systems which enjoy some degree of monopoly power in specialized fields of activity. The problem, as with Associated Press, is most likely to arise in situations where some joint arrangement exists among computer users who want to exclude additional competitors; or, as in Southern Photo and Otter Tail, where the firm controlling the essential computer facility seeks itself to compete with those requiring access to that facility.

The basic principle is equal treatment for all comers. Late comers cannot be excluded if the facility can be fashioned to accommodate them, although it may be "only fair that the newcomer should pay rather more for the new facility than those who have invested over a long period."[19] These are practical problems which will have to be worked out in the context of the hard facts posed by particular schemes.

Even so, as I stressed earlier, fair access to a monopoly facility is very much a second best solution from the antitrust standpoint. Where it is economically possible to have direct competition between two or more competitors for the same service, the economic results would likely be better and all these difficult legal questions can be avoided.

---

[1] See generally, the Response of the United States Department of Justice, FCC Computer Inquiry, Dkt. 16979, March 5, 1968.

[2] 28 FCC 2d 267 (1971). See also the Commission's decision in Carterfone, 13 FCC 2d 420 (1968), a case in which the Department participated as amicus curiae.

[3] Small Business Act of 1958, §2(2), 15 U.S.C. §631.

[4] United States v. Philadelphia National Bank, 374 U.S. 321, 372 (1963); see also Northern Pac. R. Co. v. United States, 356 U.S. 1, 4 (1958).

[5] Section 1 provides: "Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with Foreign nations, is declared to be illegal. . . ." 15 U.S.C. §1.

[6] Section 2 provides: "Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons, to monopolize any part of the trade or commerce among the several States, or with Foreign nations, shall be deemed guilty of a misdemeanor. . . ." 15 U.S.C. §2.

[7] See, e.g., United States v. Grinnell Corp., 384 U.S. 563 (1966).

[8] See, e.g., Eastman Kodak Co. v. Southern Photo Materials Co., 273 U.S. 359 (1927); Lorain Journal Co. v. United States, 342 U.S. 143 (1951).

[9] United States v. Terminal R.R. Ass'n., 224 U.S. 383 (1912) (terminal railroad controlling a vital bridgehead); United States v. Great Lakes Towing Co., 208 Fed. 733 (D. Ohio 1913), 217 Fed. 656 (D. Ohio 1914), appeal dismissed, 245 U.S. 675 (local towing facilities in several Great Lakes ports).

[10] These include a produce exchange building, Gamco, Inc. v. Providence Fruit & Produce Bldg., 194 F.2d 484 (1st Cir. 1952), certiorari denied, 344 U.S. 817; a tobacco market, American Federation of Tobacco Growers v. Neal, 183 F.2d 869 (4th Cir. 1950); a fish market, United States v. New England Fish Exchange, 258 Fed. 732 (D. Mass. 1919); and a sponge market, United States v. Tarpon Sponge Exchange, 142 F.2d 125 (5th Cir. 1944).

[11] Associated Press v. United States, 326 U.S. 1 (1945).

[12] Silver v. New York Stock Exchange, 373 U.S. 341 (1963).

[13] 326 U.S. at 13.

[14] See also Gamco, Inc. v. Providence Fruit & Produce Bldg., Inc., 194 F.2d 484, 487-488 (1st Cir. 1952), certiorari denied 344 U.S. 817.

[15] 326 U.S. at 17, n. 17.

[16] See Standard Oil Co. v. United States, an early antitrust landmark, where Chief Justice White generally made this point — namely, that Section 1 is aimed at the means of achieving that which is prohibited by Section 2. 221 U.S. 1, 61 (1911). In Gamco, supra, where joint control was not a key issue, the Court of Appeals treated "exclusion of competitors from the market" as conduct "condemned per se by Section 2," at least where space in the market was available. 194 F.2d at 486-7.

[17] United States v. Otter Tail Power Company, 331 F. Supp. 54 (D. Minn. 1971).

[18] 331 F. Supp. at 61, citing Eastman Kodak Co. v. Southern Photo Materials Co., 273 U.S. 359 (1927) and Lorain Journal Co. v. United States, 342 U.S. 143 (1951).

[19] Neale, A.D., The Antitrust Laws of the U. S. A., 2d Ed. (Cambridge: Cambridge University Press, 1970), p. 69.

THE PROMISE AND PERIL OF COMPETITION IN INTERCITY COMMUNICATIONS

Kenneth A. Cox, MCI Communications Corporation, Washington, D.C., U.S.A.

Introduction

Like the Phoenix of legend, competition in common carrier communications has risen again from its ashes. The chaotic and admittedly inefficient competition in basic telephone service in the early days of the industry was replaced by the concept of the single area-franchised utility, whose monopoly is now to be modified slightly by allowing competition in the field of business and data communications. The Federal Communications Commission has initiated this rebirth in a series of innovative and courageous decisions that have shattered popular preconceptions as to the resistance of regulatory bureaucracy to sharp and rapid change.

What the Commission has been willing to do is to act on the considered judgment that the public interest will best be served by a revival of competition in limited sectors of an industry which is daily growing more essential to business and government. That it has not done so lightly or rashly speaks well of its wisdom. That it has done so at all testifies to its genuine dedication to the goals for which it was created and its desire to discharge the responsibilities the Congress has given it.

Certainly, the Commission's decision to permit competition in a corner of the regulated common carrier industry cannot have been an easy one. While few would contest that competition is normally the least cumbersome regulator of the marketplace, many have asserted that the communications common carrier industry is particularly unsuitable for its normal and proper functioning. To support this contention, they have marshalled arguments based on long communications practice; the mix of monopoly and non-monopoly services which poses serious problems of cost allocations; the complications involved in the two-tier regulation of rates and practices; the mandate to insure adequate national service, which might be adversely affected by the abandonment of nationwide average rates; the magnitude of the communications industry's capital requirements; the risk of uneconomic duplication of plant; the masses of small investors in the Bell System, etc. All of these, it was claimed made regulated monopoly better than competition in the common carrier communications industry.

Each of these aspects of the situation contributes to the basic dichotomy implicit in the "promise versus peril" title of this paper. Does the offering of a mix of monopoly and non-monopoly services by the same company advance of retard competition? Clearly, it can do either. It can advance effective competition if the regulatory agency can define a clear distinction between the competitive and non-competitive services and fairly assign costs to each element of the total service; or it can retard competition by obscuring the differences between services to such a degree as to permit subsidy of the services in which the carrier faces competition. One of the FCC's most urgent responsibilities is the development of principles of rate making which will prevent such cross-subsidy and thereby keep predatory pricing from stifling the new competition which the agency has found to be in the public interest.

The same question exists in connection with the fact that communications is subject to divided regulatory control under our dual governmental system. The problem of divided jurisdiction exists in many areas, but nowhere has it been more difficult to deal with than in the regulation of telephone communications. The problem of separations--the assignment of responsibility, as between the intrastate and interstate regulatory bodies, for the costs of commonly used communications plant--has been an acute one for at least twenty-five years, and has resulted in continuing disputes as to usage characteristics and changing technology in the complex telephone network. Now it seems to be becoming a philosophical-legal dispute as to the subsidization of local service out of interstate revenues. In both the areas of interconnection of non-carrier supplied terminals and authorization of the specialized new carriers, federal authority has encouraged competition, while state regulators, or at least some of their spokesmen, have seemed wary of it.

What is meant by the emphasis which some put on "national" service? Does it simply mean that carriers should offer their services nationally, so that they are available to people in all parts of the country; or does it also include an obligation to price the same service equally throughout the country, irrespective of regional differences in costs and revenues? This becomes important because the established carriers have charged that the specialized carriers seek to serve only the high density, profitable routes--thereby, it is claimed, "skimming the cream" of the common carrier business. AT&T and Western Union contend that their rates over such routes are based on nationwide average costs (including those of sparse, less rewarding routes) and are therefore higher than would be the case if they had set them strictly in relation to the per circuit costs of these segments of their systems. They say that, in order to meet the competition of the new carriers, they must abandon such nationwide averaging and go to a system of route-by-route pricing. While regulatory agencies have generally recognized the social advantages of uniform national rates based on averaging of nationwide costs, the FCC was not persuaded that this practice was so vital as to exclusion of possible new competition. It pointed out that the carriers would not have to abandon averaging in their basic monopoly services, because no one seeks to compete in those areas. However, it said that the existing carriers could abandon nationwide averaging in the business and data communications field, if they so desired, and could participate in the specialized communications market so long as their participation is not a burden on their other services and is based on prices which "realistically and reasonably reflect economic advantages, if any, that are inherent in the plant and operations of those carriers."[1] Again, the matter of the proper costs to be used by the established carriers in pricing their competitive services is the crucial issue.

Does competition promote or impede the decision-making processes of the communications industry? Will competition compel telephone and telegraph executives to make a cost-effective allocation of the limited capital resources of the existing carriers, so as to stimulate new and more meaningful services, or will it lead to diversion of funds to services where these carriers are not the most efficient suppliers and result in too many trips to the pump, so that the well may run dry and keep the carriers from financing needed expansion of their basic services? AT&T has been faced with the necessity of raising more than $4 billion in new capital in each of the last couple of years--at a time when interest rates are at their highest level in recent times. It has been suggested that Bell should welcome the ability of the new carriers to tap new sources of capital to finance part of the needed growth in the overall communications system, but the Company does not seem to accept this view. It argues that <u>any</u> construction of intercity transmission plant will constitute wasteful and uneconomic duplication of facilities and will impair its vaunted economies of scale. Again, however, the FCC does not agree.

Finally, who will benefit from competition? If the public will, as the Commission concluded, then demand for all forms of communications should increase, so that the investors in the existing carriers will also benefit, rather than suffer, as a result of the authorization of new carrier entry. But if competition does not prove to be feasible and the new carriers fail, the investors in the established carriers will end up in an even stronger position. This will leave only the entrepreneurs who have financed the new carriers--and who entered the market knowing they had what Chairman Burch has aptly termed "the freedom to fail"—to meditate on man's fate.

A Bit of MCI History

To those who have planned and worked toward the creation of new carrier service, promise has obviously overshadowed peril, though how Jack Goeken sustained his lonely battle to launch Microwave Communications, Inc. for six years, living <u>only</u> on promise, mystifies me. The specialized common carrier applicants seem generally to be aware of the risks of their enterprises, but they have preferred to concentrate on getting their licenses to compete, leaving on the back burner the problems of the competitive environment in which they would have to succeed or fail. But now the back burner is bubbling over.

Much has been written of the development of the specialized carrier concept.[2] Similarly, the implementation of that concept has also been well chronicled.

Jack Goeken sold and serviced mobile radio systems in Joliet, Illinois. He knew that the value of his "product" to potential customers was contingent upon the extent of the range of service they could enjoy, and he began to explore methods of expanding the coverage area of his radio systems. In the process, he devised the idea of a radio relay system, amplifying and repeating signals from one microwave tower to another. He configured this system to parallel his desired service area along Route 66 and the Illinois River between Chicago and St. Louis, and filed the required applications with the FCC.

I gather that this generated a certain amount of consternation; it certainly produced a lengthy series of administrative proceedings. But all during the six years between initial filing and the grant of his construction permits, Jack Goeken, belittled and beset by the established carriers, worked to improve and expand his original concept. He was joined by Bill McGowan in 1968, and when the grants came in August 1969,[3] they had begun to develop support for a nationwide system of interconnected regional companies to offer private line service on a national basis.

The MCI grants proved that a small company of limited resources but great ingenuity and patience could take on the established carriers and win. There followed applications from thirty-four companies--of which seventeen were MCI affiliates--seeking to enter the specialized common carrier field. To deal effectively with this volume of applications--envisioning some 2,000 individual microwave sites, with their associated frequency coordination problems--the Commission grouped them all into one rulemaking proceeding and addressed itself to the threshold questions common to each of them, and to the whole concept of the new industry they wanted to establish. In briefest form, these questions were (1) whether competition would be desirable; (2) if so, should only one license be granted in a geographic area in which there was more than one applicant; (3) how could frequency use be best coordinated among the applicants; (4) what policies should be prescribed to protect satellite service and promote efficient technical operation; (5) how could the customers of the new carriers be ensured quality service; and (6) how could their signals best be distributed locally.

This rulemaking proceeding, Docket No. 18920, was begun in July 1970.[4] In just ten months it had been concluded,[5] and the specialized common carrier industry was born. Considering the enormity of the issues involved and the magnitude of the battle which was joined over them, such speed and decisiveness were unparalleled. The fact that only the established carriers opposed the FCC staff's tentative conclusion that competition was feasible and desirable, and that both industry (including significant representation from the Fortune 500) and government (principally the Department of Justice and the Small Business Administration) supported it, doubtless expedited and shaped the Commission's affirmative action by substantiating the common assertion of all the applicants that there was a need for the various services they proposed which the existing carriers had not met.

The "minor miracle" of Docket 18920 permits any technically, financially, and legally qualified applicant to offer its specialized services to the public, and at the same time recognizes a right on the part of the established carriers to full and fair competition.[6] That Solomonic decision embodies the promise or peril dichotomy which seems to be implicit in the concept of competition in common carrier communications.

The Problems of New Competition

Competition in a field that has long been treated as a natural monopoly, with regulation supposedly serving as the surrogate for market forces, is not going to be a simple matter. Hopefully the new carriers will be able to raise the necessary capital; they have, or can find, people who know how to build, maintain, and operate microwave systems; they can recruit salesmen, who should be able to find customers among the many businesses and industries who have indicated dissatisfaction (in Docket No. 18920, in the FCC's Computer Inquiry, etc.) with the services heretofore provided by the existing carriers. They can no doubt open up shop and attract subscribers--some perhaps diverted from AT&T and Western Union; but more of them seeking means of satisfying unmet needs. I think the question is not how the new carriers will go about competing for a small corner of the overall communications market, but how the existing carriers will respond. The issue is what they will do--and what the regulatory commissions and the Department of Justice will let them do--by way of competitive response.

The Commission has said the existing carriers can participate in the market for specialized services "so long as their participation is not a burden upon or significantly detrimental to their other services."[7] But how does one tell if their efforts to meet competition are a burden or detriment? Again, the Commission's order recites that the existing carriers may depart from uniform nationwide prices, based on averaged costs, in the services where they now face competition, and goes on to say that it is the Commission's intention "to permit the existing carriers to price their competitive services in a fashion that will realistically and reasonably reflect economic advantages, *if any*, that are inherent in the plant and operations of those carriers." (Emphasis supplied). Well, how is the Commission going to tell whether an existing carrier's new competitive rate reflects legitimate economic advantages which it enjoys, or is made possible only by cross-subsidization out of the carriers' monopoly revenues? And the Commission said "there should not be any 'protective umbrella' for the new entrants or 'any artificial bolstering of operations that cannot succeed on their own merits'." But how will the Commission know whether a specialized carrier alleging predatory pricing practices by an established carrier is seeking a "protective umbrella" or is validly complaining about a hidden cross-subsidy? Or how will it tell whether a new company which is on the verge of failure is in trouble because it could not succeed on its own merits, or is being done in by anticompetitive conduct on the part of a competing carrier which doesn't have to worry about umbrellas in the business and data communications markets because it is assured of receiving millions--or even hundreds of millions--of dollars every year from safe, secure monopoly operations? I do not doubt the Commission's sincerity in authorizing the existing carriers to compete or in expressing an intention to see to it that their competition is fair, but I have some concern about the tools the Commission has to do the job with and about the speed and effectiveness with which it can act.

The Commission's Past Handling of Competitive Issues

The FCC has not had noticeable success in dealing with problems of this kind in the past. Let us consider the case of Telpak and the same ten or eleven years of litigation in which it has been involved. I think its history is revealing and more than a bit frightening.

When the Commission allocated frequencies for private microwave operations in 1959--in the face of the same kind of concerted opposition from the existing carriers as they later mounted in Docket No. 18920--AT&T was evidently fearful that some of its larger private line customers might cancel their service and build private systems of their own. It therefore filed a new service offering called Telpak, which purported to provide bulk communications in a discrete broad bandwidth, but was found actually to be simply a method of repricing ordinary private line circuits. Telpak A offered a discount to users of 12 voice channels between two points, while Telpak B, C, and D provided increasing discounts to users of 24, 60 and 240 channels, respectively. In fact, the discount on a Telpak D cross-section was 85% of the single channel rate.

Two classes of competitors alleged that they were hurt and asked the Commission for relief in 1961 or 1962. Motorola said that it and other manufacturers of private microwave equipment were damaged because the very attractive rates offered to large users removed any incentive, in most cases, for the construction and operation of private systems. And Western Union, which had had to follow AT&T by offering a Telpak service of its own, alleged that it had lost millions of dollars in revenues due to diversion of its business to Bell, or due to its repricing of facilities it had previously been furnishing its customers at higher rates. These parties alleged that Telpak involved discrimination against the small users of private line service, who were still charged the old, higher rates for their circuits, while the large users were getting comparable circuits at the discounted rates.

The FCC designated the matter for hearing. That is its normal method of proceeding, and one expressly authorized by the Communications Act. However, in recent years this procedure has often involved terribly long, expensive and often inconclusive hearings, so that it is not really adequate to the needs of the Commission, injured competitors, or the public.

In 1964 the Commission issued a decision in the Telpak case.[8] It found that Telpak and the regular private line offerings were, indeed, "like services," that Bell had not shown any savings in connection with Telpak to justify its lower rates; and that the rates were therefore discriminatory. Since the Act prohibits discrimination, that might seem to have been the end of the matter.

The Doctrine of Competitive Necessity

But Bell claimed that its Telpak rates were justified by "competitive necessity," a vague concept which I don't think the Commission has ever defined. The Communications Act really bars only "unjust or unreasonable discrimination," and apparently some utility commission, at some point now lost in antiquity, decided that discrimination is not "unjust or unreasonable" if it is practiced in order to hold business in the face of competition--hence the name, "competitive necessity." I really think this was, and is, a false and pernicious concept. A common carrier is usually protected from competition because it is believed that this will permit it to generate economies of scale, with resulting savings to the public. So one would think that a well-run telephone company, with hundreds or thousands of channels between a pair of points, would be able to provide private line circuits, whether singly or in quantities, at rates lower than anyone else not enjoying comparable economies of scale. If not, then it would appear that the company is not the most efficient, lowest cost supplier of such service and should concentrate its resources in areas where it is more efficient.

But the doctrine has won some measure of acceptance, and the Commission applied it in the Telpak case. It first concluded that there was no competitive necessity for Telpak A and B, because it didn't think anyone would build a private microwave system to handle only 12 or 24 private line circuits. But it decided--without adequate analysis, I believe, although I voted for the result--that there was a competitive threat that some of Bell's customers with requirements for fifty or more voice circuits might build their own systems unless AT&T offered attractively low rates to hold their business.

I say the analysis was inadequate because we simply accepted AT&T's assertion that it had to keep its private line business, without scrutinizing that claim. For example, we did not consider the possibility--which I believe to be fact--that the company was, and is, under such pressure to expand its basic switched services that it never had any significant surplus capacity which it had to dispose of at bargain rates, and that it could, therefore, have used any vacated channels for switched service--and would have earned more for its shareholders than it did by continuing to devote them to private line use, at bulk rate discounts.

Rates Must Be Compensatory

In any event, the Commission conluded that Telpak C and D were competitively necessary--but that didn't close the case either! The agency said that even if the rates meet this test, they still had to be "compensatory." It said that AT&T had failed to prove this point, but should be given another chance to do so in a further hearing. I dissented on the ground that it was Bell's fault that it had not made its case, and that the tariff should have been cancelled. The sad truth is that eight years later, the Commission has still not resolved the question of whether the Telpak rates are compensatory.

The Commission first considered the question in further hearings in the original case. Then in 1965 it began its first full-scale investigation of AT&T's rates, and announced that it would develop ratemaking principles and standards in that proceeding for evaluating the relationship among the rate levels for the Company's various services. Although Phase I of the hearing (dealing with rate base, rate of return, and jurisdictional separations) was completed very promptly, Phase IB which was to consider the ratemaking issues became bogged down in a welter of conflicting theories and testimony. Finally the Commission concluded, in 1969, that it could not resolve these questions in the abstract, but should consider them in the context of particular rates. It therefore "noted," without approving, a statement of principles agreed to by parties--which really resolved nothing--and transferred the whole controversy (including a massive record) to Docket 18128, which had been instituted a year earlier to consider the protests of the larger users against an increase in the Telpak rates which Bell had filed in early 1968. (There have been two other rate increases since--so apparently Bell, itself, recognized that its rates were too low.) The matter is still at issue there, along with specific issues as to cross-subsidization and why all of Bell's 1970 request for increased revenues was loaded onto its monopoly Message Telecommunications Service which were added when the FCC instituted Docket No. 19129 in January 1971. The proceeding is not really well adapted to resolve all these questions since the principal contenders are the large users, who want lower Telpak rates, and AT&T, which is trying to justify its present rates. MCI believes even those rates are too low and involve subsidy, but it got into the proceeding only toward the end. So only the Commission's staff has represented a position throughout the case which is possibly adverse to the concept of discounted rates for bulk users. The record may be closed in the near future, but a final decision is probably still a year away.

I think that part of the problem is that on one is really certain what "compensatory" really means. It seems to imply, at the very least, that the rates for a given offering must "cover the costs of providing the service," plus some return or profit on the plant investment used in rendering it. But, again, the question is "What are the relevant costs?" For years AT&T said that its rates should be tested on the basis of fully allocated costs. But when it was required in 1965 to make a fully allocated cost study of all its services at once, this showed that it was earning only .3 of 1% on Telpak, as compared to about 10% on WATS. It then announced that fully allocated costs were not the significant measure. It has propounded a number of theories since, and is now urging, in Docket No. 18128, that the true test is long range incremental cost, incorporated in a new "burden theory" propounded by one of its economic experts. As noted above, when the Commission instituted Docket No. 19129 in January, 1971, to evaluate the largest request for increased interstate rates in Bell's history, it raised issues as to whether those increases will burden MTS and whether Bell is using its monopoly revenues to cross-subsidize its competitive services. But it put these questions in Docket No. 18128, which has posed some very real procedural problems.

The terms "burden" and "cross-subsidy" are somewhat new in the regulatory lexicon. They do not appear in the Communications Act and are not clearly defined. But it would seem that rates which "burden" the users of another service give an "undue or unreasonable preference" to the users of the favored service, and subject the users of the other service to "undue or unreasonable prejudice or disadvantage"--and these practices are barred by the Act. Similarly, "cross-subsidization" of competitive rates would seem to be a form of predatory pricing which the Commission can reach under its mandate to enforce the antitrust laws. There is no doubt that AT&T has enormous revenues from its secure monopoly operations which make it possible for it to subsidize its competition with the new specialized carriers--who have no comparable collateral resources. The problem which faces the new carriers and the FCC is to be able to tell if and when the Company crosses the line from potential to actual subsidization. And Western Union, on a smaller scale, has some potential for corss-subsidy. So the old saying can be paraphrased, for the specialized carriers, to read "Eternal vigilance is the price of survival."

What Will Be the Existing Carriers' Response to New Competition?

MCI is concerned about the existing carriers' competitive response. In the face of the mere prospect of competition, AT&T instituted its experimental series 11,000 service, announced that it would construct a functionally discrete but physically integrated digital network serving sixty cities by late 1973, and just last fall it announced a "new development by Bell Labs" which would permit it "to handle all the requirements for digital data service up through about 1977 on our existing radio systems." Meanwhile, it has been negotiating with the parties to Docket No. 18128 in an effort to settle the case and develop a new bulk rate private line service to replace Telpak. Western Union has developed its Datacom service and announced its own plans for a data network. And on April 6, 1972, it filed tariff changes reducing its rates for voice and medium speed data channels between Chicago and St. Louis by from 14 to 32%, thereby matching MCI's rates for comparable service. This is the first instance of abandonment of nationwide average pricing by an existing carrier, and the Commission, on a protest by MCI, has set the matter for hearing. Thus it is clear that whether it represents an efficient use of their facilities and serves the interests of their shareholders or not, the existing carriers do intend to take advantage of the Commission's statement that they can compete fully in the specialized communications market. Our concern is whether they will also comply with the injunction to compete fairly--and whether the FCC can detect and correct unfair pricing or other practices, if any, in which they may engage.

The Procedural Dangers

Our problem is complicated by the fact that, in most cases, carriers can file any tariffs they please and the Commission, if it doubts their lawfulness, can only suspend them for ninety days and put them into a hearing. But that hearing may take years to complete, and can be very time-consuming and costly for a specialized carrier which is not in a position to pass its expenses on to a large class of users. And since the concepts discussed above are still in a state of flux, such proceedings can be very frustrating and inconclusive. We have urged the Commission aggressively to enforce its new rules requiring more detailed showings in support of tariff filings. We have supported its proposal requiring advance approval before new services can be put into effect. We have asked it to deal with one problem--the resale of AT&T's communications capacity by unregulated third parties under the joint user provisions of the carrier's private line tariff--by rule making. We feel that it must reject tariffs which are designed to block or destroy competition unless they clearly show that the rates are justified in terms of the proper costs. And we would like the Commission to experiment with expedited rate hearings, and to make greater use of its authority to enforce antitrust policy in these matters.

In short, we think that the FCC acted in Docket No. 18920 in a most forward looking way, by authorizing a whole new competitive communications industry in one small corner of the overall field. But we think that it is going to have to be very aggressive and very imaginative in its procedures if it is to insure that the competition it has found to be in the public interest will really have a chance to come into being, get room and time to grow, and become a significant factor in serving the public's exploding communications needs.

What is Full and Fair Competition?

The Commission is going to have to make clear what it meant by "full and fair competition." That phrase is subject to a number of interpretations; or perhaps I should say that it is the *object* of a number of *subjective* constructions. I would not want to impute any sanguinary motives to my opposite numbers in the Bell System, for instance, but there is a certain chop-licking rhythm to some of their recent statements.

William Ellinghaus, President of New York Telephone Company, told the Conference on the Revolution of Business Information in August 1970: "The grand old monopoly I work for is in fact a highly competitive outfit and I don't think anybody need worry about our ability to take care of ourselves in fast company."

As Vice Chairman of the AT&T Board, John deButts told the New York Society of Security Analysts in October 1971: "If [the FCC finds] it in the public interest to open up [terminal equipment and private line services] to competition, then it's all right with us so long as . . . the ground rules are the same for all parties . . . I haven't the least doubt about the Bell System's ability to give a good account of itself."

A little more bullishly--of course, he was now Chairman of the Board--Mr. deButts told the AT&T annual meeting in April 1972: "Make no mistake about it, our costs are lower--in most cases, a whole lot lower--than those of the new carriers, and did we choose to price on a route-by-route basis, our rates would be lower too on just about every mile of the routes proposed. But nationwide average pricing has been a major factor in extending high quality communications throughout this great country of ours--to small towns as well as large--and we are understandably reluctant to abandon it without sufficient occasion. However, if we must, we will."

None of these statements, standing alone, would particularly worry me--if I thought they really reflected an expectation on the part of their authors that Bell and the new carriers would engage in vigorous, but mutually beneficial, competition in this new market. But much closer, I believe, to Bell's view of the future is the remark John deButts made as Chairman-designate at a press conference on February 16, 1972: "Competition posed by new, smaller companies, especially in the data transmission business, doesn't bother me in the slightest."

I'm afraid the thought of competition doesn't bother him because he doesn't really believe there will be any significant competition. There will be no "new, smaller companies" to compete with--at least not many and not for long. Competition, that wild idea of the Commission, will die before it can really begin!

I think the truth of the matter is that the Bell System can't really tolerate the thought of competition in its domain. Its people genuinely consider competition to be disruptive and contrary to the public interest. In their view, I'm inclined to believe, to be anti-competitive is not merely to be pro-Bell, but is actually to be *pro bono publico*. The fundamental operating principle of the Bell System seems to be that there is no communication service that any subscriber might reasonably require that the Telephone Company cannot provide. While there is no corresponding commitment to provide it when the public wants it, or at rates the public considers reasonable, the Bell people remain convinced that there is no service you could want that Bell could not potentially give you--and so no one else should be allowed to supply it, even though Bell hasn't gotten around to doing so.

John deButts put it this way in his first press conference as Chairman-designate: "Well let me repeat the position that we have taken in connection with the specialized common carriers and that is that actually we believe that the service is being provided today." Thus, in the face of the statements by more than a hundred parties in Docket No. 18920 that the existing carriers' service has been inadequate, Mr. deButts holds fast to the view that there is really no need for the new competitive carriers.

In AT&T's view, I'm sure competition also seems disruptive. Since the services and terminals the new companies propose to offer are regarded as redundant, Bell probably does not expect them to survive for long in the marketplace. But during their brief existence, Bell is afraid that they will have drawn customers away from its operating companies and Long Lines Department. When these subscribers fail to find in the competition's services the panaceas they had hoped for--as Bell asserts they must--they will return to Bell and demand service, which Federal and State regulators will require Bell to provide. The process of undertaking to serve them again may disrupt service to other subscribers, since the huge, but not limitless, resources of the Bell System must be allocated on some system of priorities, and someone is always at the bottom of the list.

And if competition is disruptive, the Bell people presumably think it is also destructive. For as customers desert the new competitive enterprises, the latter can no longer survive. As they close down, they cast employees on the job market, fail to meet their debts, cancel contracts on which suppliers had relied, and so on through successive rounds of economic cause-and-effect. The net result is a dislocation of the economy without, in Bell's view, any offsetting public benefit.

Isn't it, then, the Bell people must feel, much better to halt this prodigal and pernicious process before it starts? And this can be done very subtly, while seeming to welcome and praise the competition. The new company line seems to go something like this: Competition is unnecessary and harmful, but if we have to have it, then we are overjoyed about the whole thing. Mr. H. I. Romnes, Bell's immediate past Board Chairman, has said: "I already sense--from the imagination and zest with which our competitive response is being shaped . . . that . . . (competition) is going to be good for our business and good for the public." And Mr. deButts has announced that "Competition kind of puts us on our toes. It puts us in a position where we learn better marketing methods. . ."

From these--and the somewhat more sanguinary quotations cited above--can be gleaned the following between-the-lines assertions: "Now is the wrong time to move, Mr. Customer, just when we're ready to make those changes you've wanted all along. You'll be disappointed if you pull your marbles out now. Only we are big enough to take care of you. And anyway, how long are those guys going to be around?"

Conclusion

There would clearly be nothing wrong with these implied promises of new wonders to come--if they were really credible. But Bell has had something of a history of announcing new services with great fanfare, and then delivering them belatedly, or in greatly reduced quality. Picturephone was to have been the great achievement of the 'forties; data transmission, the breakthrough of the 'fifties; electronic switching, the miracle of the 'sixties; and now data networking is the watchword for the 'seventies. That delivery of these services is running a little behind schedule says less about inadequacies at Bell Labs or Western Electric than it does about the lack of the competitive urgency, until recently, needed to stimulate prompt and sustained action toward the implementation of these noble goals.

But what may be accomplished by this simple strategem is the impoverishment of competition before it can really develop. Potential customers who might have turned to the competitive services may be dissuaded or intimidated from doing so. Likely investors in the new ventures will perhaps foresee even greater risks than their investment philosophies will support. Normally available pools of capital can dry up. State regulators may increasingly question the viability of competition as a concept, and ponder the possible impact on separations, or local exchange rates, or other matters which concern them. And before you know it, it can all be over!

This, then, is the peril which may, or may not, extinguish the promise of competition in communications.

The promise is that if the Commission can generate and maintain full competition and can keep it completely fair, then the public may judge and choose, on the basis of comparative performance, between the established carriers and the purveyors of new services. Hopefully, then, both groups of competitors will be stimulated to provide innovation, flexibility, reliability, economy and improved quality for the customer.

But that there is peril is equally clear-and that is the danger that if the Commission is not vigilant and creative in performing its regulatory functions, competition in communications may never fully come into existence.

---

[1] Specialized Common Carrier Services, First Report and Order, Docket No. 18920, 29 F.C.C. 2d. 870, at 915 (1971).

[2] See, for example, Mathison and Walker, Computers and Telecommunications: Issues in Public Policy, Prentice-Hall, Inc. (1970; Walker and Mathison, "Specialized Common Carriers," Telephone Engineer and Management, October 15, 1971.

[3] Microwave Communications, Inc., 18 F.C.C. 2d 953.

[4] Specialized Common Carrier Services, Notice of Inquiry to Formulate Policy, Notice of Proposed Rule Making, and Order, 24 F.C.C. 2d 318.

[5] Note 1, op. cit.

[6] Id. at 915.

[7] Id. This and the other statements quoted or paraphrased below are from Paragraphs 89 and 90 of the Commission's First Report and Order in Docket No. 18920.

[8] Telpak, 37 F.C.C. 1111.

COMPETITION IN THE FIELDS OF COMPUTERS AND COMMUNICATIONS IN JAPAN

Yasuo Makino, Ministry of Posts and Telecommunications, Tokyo, Japan

1. Preface

In modern societies, it is essential to make the benefits of technological progress equally available to as many people as possible for the achievement of peace and prosperity of mankind. Policy makers should set their goals of establishing a social system to this end. Japan pursues the policy that competition under a democratic social system on an international scale must be carried on so as to encourage worldwide techno-economic progress and social development.

This paper starts from this standpoint in discussing the role of competition in Japanese computer-communications.

2. The Need for International Cooperation

The progress of both current remote access data processing and electrical communication is based on a number of remarkable technological developments in electronics.

The coverage of systems recently developed is increasing from a local to a nationwide scope and still further to an international one. This leads to the need for international relationships in many fields in order to achieve:

1) Free exchange of technological information, and
2) Establishment of international technical standards.

In the telegraph and telephone fields, efforts have already been successfully made through CCITT activities and have played an important role in developing a global communication system.

3. Areas and Conditions of Competition

Although it is self-evident that free competition generally promotes economic development, some kind of regulatory measures may be necessary in order to ensure fair competition. However, they should be applied to the minimum extent possible. The minimum conditions will be:

1) Assurance of security,
2) Elimination of excessive competition, and
3) Prevention of unnecessary competition in public utilities which are naturally monopolistic in nature.

Therefore, areas of competition may be limited to the following:

1) Competition in service quality and price,
2) Competition only in service quality, or
3) Competition within restricted local areas.

Because of its natural monopoly nature, telecommunications service is usually offered by a locally monopolyzed enterprise and no competition can exist. Protection of privacy and satisfactory operation of the system are ensured by regulations.

4. Competition in the Fields of Data Processing and Telecommunications

Marked distinctions exist between data processing and telecommunications. However, in the field of remote access data processing systems, where telecommunication forms an integral part of the system, it is difficult to draw a clear demarcation line between the two and treat them as distinct systems.

In Japan, the same manufacturers supply hardware for both systems, and the common carrier is allowed to provide both services. Therefore, we Japanese are likely to have a similar concept of remote access data processing and telecommunications in the technology.

4.1 Research and Development (R&D)

No one denies that competition is an essential condition in the R&D field. However, there exist some exceptions in a large-scale development activity or in the development of a large system.

When developing a large-scale system, the development activity itself has a characteristic of natural monopoly. This is because much consideration will necessarily be given to the compatibility of the system components, even when the project is conducted by a single enterprise, and the performance of a system will have a great influence on later-developed systems.

In Japan, such development work in computer-communications systems, however, is too big to be achieved in a single private enterprise. Therefore, it is being taken up as a national project, avoiding competition on a national scale among enterprises.

A number of Japanese telecommunications equipment manufacturers have collaborated with NTT, the common carrier in Japan for domestic telecommunications service, in developing large-scale systems such as DIPS, a large-scale data processing system designed especially for remote access use, and DEX, an electronic switching system. The DEX will be used as an integral part of the future international network as well as in the national or local networks.

One of the primary objectives of the DIPS project is standardization. Therefore, I expect that DIPS will be used not only in NTT but also in many areas of business enterprise in Japan and other countries. Main items of standardization are:

1) Machine language for domestic use, and
2) Input/output interface (channel interface) for international use.

The latter was proposed at ISO/TC-97/SC-4 in 1969. I do hope that collaboration of the United States and other countries will be helpful for its further progress.

The characteristics of DIPS are shown in Figure 1, in which a comparison is made with the IBM 370/165 computer system.

In the development of a large-scale system in the field of data processing and telecommunications, competition is possible and also effective only on an international scale. Furthermore, I'd like to stress that international cooperation on standardization is a necessary condition in order to make a huge international system.

As far as terminal equipment and mini-computers are concerned, maximum creativity and innovation can be encouraged only if free competition is maintained. However, standardization is required at the point of connection, when such equipment is connected to a large system and workable as a subsystem. Competition in development and interface standardization may be viewed as two opposite sides of the development of huge systems.

4.2 Manufacturing

Free competition is the underlying principle on which the manufacturing of equipment for data processing and telecommunications is based. Although equipment supplied to NTT is purchased at the same price from many manufacturers, competition is encouraged with respect to quality. Similar competition with respect to quality exists among manufacturers which supply computers to NTT.

In order to maintain competition in the manufacture and sales aspects of the equipment on an international scale, it is necessary to discourage unfair competition so that all interested parties stand on an equal footing. In the case of DIPS and DEX either on a national or international scale, competition is likely to occur regarding system quality.

4.3 Telecommunications Service

Telecommunications service is an integrating enterprise because of its inherent nature of serving as wide a sector of the general public as possible and because of its high capital investment. Let us focus our attention on the fast advancing data transmission field, and particularly on the digitalization of transmission media. In Japan, PCM systems constitute 20% of the total carrier transmission systems. Application of PCM systems to long-distance transmission is now under way, utilizing multiplexing techniques of higher-order PCM signals (see Figure 2). This is based on the following reasons:

1) Digital systems provide a more suitable means of signal transmission than conventional analog systems when transmitting video or data signals.

2) Guided millimeter wave and laser transmission systems are now at the laboratory stage of development.

3) Marked progress in solid-state devices has made reliable and economical digital modulation and regeneration of signals feasible.

Greater advantage will be obtained if telephone and data signals are sent over the same transmission path than will be obtained if the path is separated for transmission of data signals only.

From the foregoing discussion, it is expected that, in the near future, data transmission service will be provided by efficient digital systems and that it will be possible to transmit telephone, data and video signals over such digital transmission media. If digitized signals are economically switched through by time-division switching systems, an integrated network will be realized.

In Japan, since high-density telecommunication facilities are concentrated only between large cities, active competition cannot be expected in the field of data transmission service, as compared with the United States.

However, Japan and the United States are closer together in their policies regarding interconnection to the telephone network. In Japan, restrictions on the use of foreign attachments to the telephone network were lifted by the recent revision of the Public Telecommunication Law, thus allowing various competitors to seek approval for interconnection of their equipment if it meets the specified requirements.

4.4 Remote Access Data Processing Services

In the remote access data processing field, I believe that competition is desirable. Therefore, common carriers may also be admitted to this field as long as competition is maintained. NTT can provide remote access data processing service on a separate financial accounting basis from the general telecommunications service. NTT is a Public Corporation whose budget and installation plans are subject to the approval of the National Diet. Whenever unfair competition exists between NTT and private companies, it could be judged by the general public.

In view of the growing scale of remote access data processing systems, it is anticipated that, in Japan, distinctions between service areas of NTT and private companies will be gradually clarified. A nationwide-scale remote access data processing system will be developed and provided by NTT, while local or special purpose small-scale projects will be implemented by private companies.

NTT and KDD, as Common Carriers, are capable of taking responsibilities in providing either nationwide or international systems, with DIPS as an effective tool for achieving this purpose.

In this respect, it is desired to encourage free, fair competition on an international scale.

5. Summary

The foregoing describes the effects of competition and factors influencing competition in the fields of remote access data processing and telecommunications, which can be summarized as follows:

1) The development of large-scale systems and the provision of remote access data processing service on a national or international scale should be encouraged by fair competition among big businesses. For this purpose, fair activities and security must be maintained and international cooperation toward standardization becomes a more important matter.

2) No competition will be expected in the field of telecommunications service, including data transmission.

3) Competition should also be encouraged in the equipment manufacture and sales activities on as wide a scale as possible, but in harmony

h the national interest. This requires andardization of system components or terfaces as thoroughly as possible.

e provision of remote access data processing rvice should be based upon competition, gardless of system size, either on a tional or international scale.

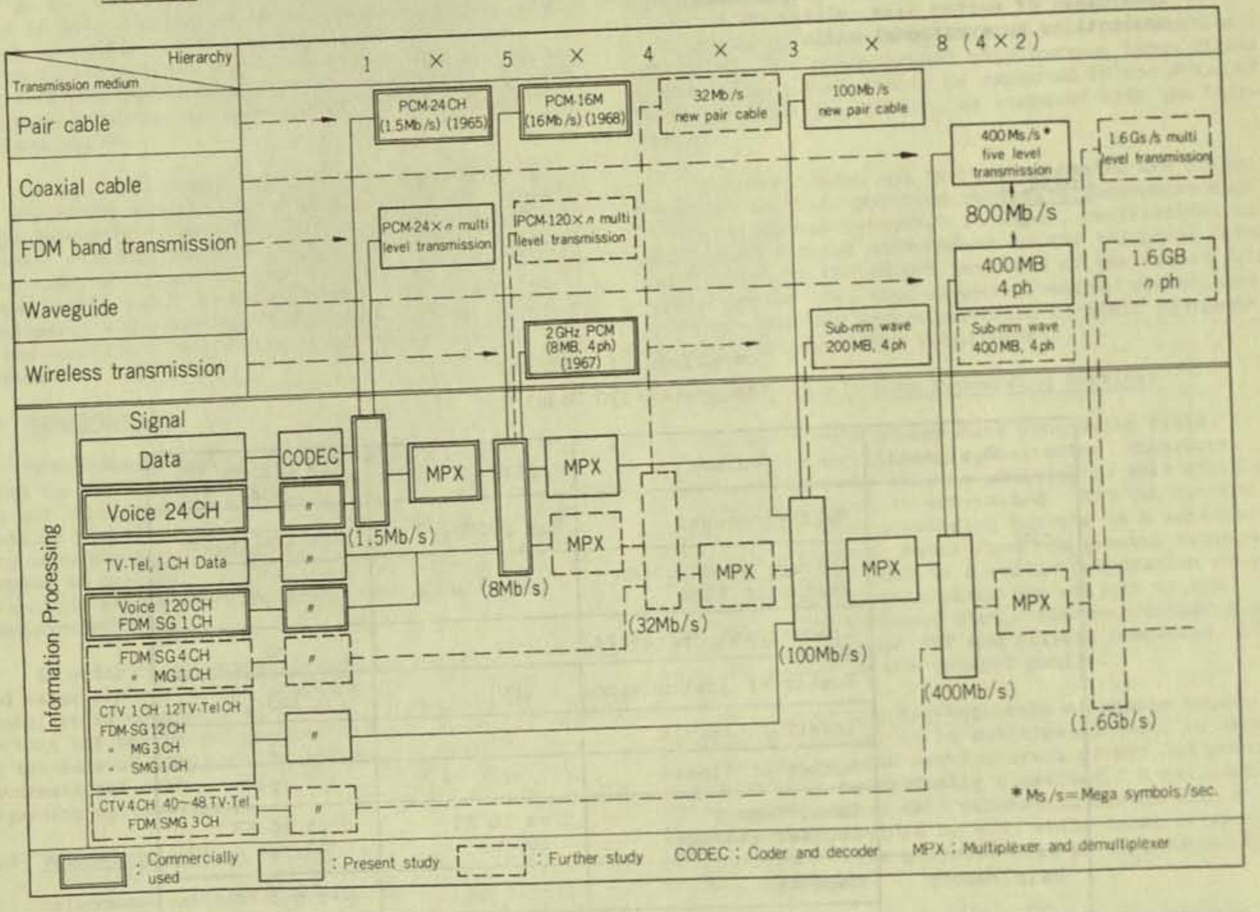
FIGURE 1

Comparison of DIPS and
IBM System/370 Model 165

Equipment	Feature	DIPS	IBM System/370 Model 165
CPU	Multiprocessor	Max 4 CPUs	Not Available
	Average Instruction Execution Time	0.63 μs	0.5 μs
	Operational Registers	36	20
	Number of Instructions	160	163
	Interrupt levels	8	5
	Number of Timers	4	2
	Local Memory (Buffer storage)	8 or 16 KB 100 ns	8 or 16 KB 80 ns
Main Memory	Capacity	1 - 16 MB	0.5 - 3 MB
	Cycle time	2 μs	2 μs
	Paging	4 KB/P Logical Space 256 MB	Not Available
	Memory protection	Ring level Access Restriction	Key-Lock
Channel	Number of Channels	Max 96	Max 12
	Total through-put	Max 72 MB/S	Max 8 MB/S
Communication Control Unit	Number of CCUs	Max 36	?
	Number of lines	256/CCU	176 ?

FIGURE 2

Progress of PCM transmission systems in Japan.



INTERCONNECTION: IMPACT ON COMPETITION - CARRIERS AND REGULATION

William H. Melody, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

Summary

Efficient interconnection is crucial to the effective implementation of virtually all public policies opening competitive opportunities in communications. Interconnection has been at issue since the earliest days of telephony and restrictions played a key role in building the existing monopoly structure in both service and equipment markets. The history of the interconnection issue demonstrates that it has always been an economic and not a technical one, based upon market protection and not protection from harm.

Since the Carterfone decision, new markets have been opened, a new industry created, user facility and service options expanded, and no demonstrable "harm" to the system created. Yet, market response has not been as great as expected because of the delays in implementing Carterfone and the obstructionist tactics of the Bell System. Contrary to common belief, the Carterfone decision signalled the beginning of an interconnection policy debate at a new level and not a resolution of the issue. As the interconnection standards and certification issues move closer toward resolution, developments in the independent terminal equipment and interconnect segments of the market will be supplemented by growth in carrier systems interconnection with new specialized carriers, and eventually a communications-based service industry.

Interconnection is an area where monopoly power has far exceeded any possible rationalization on the basis of technologically determined natural monopoly. There is a serious problem that technical standards may be created as economic barriers to entry replacing the former interconnection tariff restrictions. Successful implementation of interconnection policy will require both recognition of the benefits as well as the harms and risks, and effective regulation to prevent Bell's predatory pricing practices in both communication services and equipment markets.

I. Introduction

More than at any time since the turn of the century, telecommunications public policy is confronted with issues of competition and monopoly that could bring fundamental changes in the structure of telecommunications markets and the telecommunications industry. In significant part, these opportunities for altering the enormous degree of monopolization in the industry have occurred because of the growing interaction, interrelationship and interdependence between the computer and telecommunications industries. The collision between a dynamic competitive industry, operating in a virtually continuous state of major technological and market change, and a stable regulated monopoly that has been insulated from R&D through manufacturing and consumer services has called into question the applicability of the entire spectrum of traditional regulatory policies to the new computer-communications environment.[1] Over the past several years, the Federal Communications Commission (FCC) has been engaged in a comprehensive evaluation of the applicability of traditional public policies in light of alternative opportunities that have been opened to the Commission.

The first in a series of recent landmark FCC policy decisions opening opportunities in communications came in the Carterfone case, June 27, 1968.[2] In this decision, the FCC found Bell System tariff restrictions against "foreign attachments" unreasonable, unlawful and discriminatorily applied against the Carterfone device. In its decision the Commission went beyond the Carterfone device itself and addressed the principle of interconnection as follows:

> ...Our conclusion here is that a customer desiring to use an interconnecting device to improve the utility to him of both the telephone system and a private radio system should be able to do so, so long as the interconnection does not adversely affect the telephone company's operation.[3]

Interconnection provides the cornerstone for the entire structure of new regulatory policies because virtually all public policies opening competitive opportunities in communications require efficient and effective interconnection if they are to be successful. Thus, FCC policy decisions relating to new specialized common carriers and domestic satellites have emphasized increasingly the importance that local telephone operating companies interconnect with all long haul carriers on reasonable bases.

The Carterfone decision did not establish specific terms, conditions and standards for interconnection. Rather, it required the Bell System to re-examine its interconnection policies and file revised tariffs. In December 1968, Bell filed modified tariffs that permitted customers to provide terminal devices subject to two conditions: (1) that the customer purchase an interface device between the customer facilities and the telephone network; and (2) that Bell supply the network control signalling devices with the customer equipment. Independent manufacturers and users objected that the new tariffs were not responsive to the Carterfone decision. However, the FCC permitted the new tariffs to go into effect on January 1, 1969, without approving them as being fully responsive to the Carterfone decision.

It is now more than three years since Bell filed its revised tariffs, a long enough period to assess the changes, and the rate of change, in the competitive nature of communications markets, the response of the telephone common carriers (particularly the Bell System), and the response of regulation in implementing the Carterfone policy. The experience since Carterfone provides a unique illustration of government-business relations in communications that can be instructive for evaluating the issues and problems that must be addressed in the implementation phase of related public policies that attempt to open new opportunities in data and other specialized communications markets.

II. Historical Perspective

The Evolution of Policy and Practice[4]

Interconnection has been an issue in communications since long before Carterfone. In fact, it has been uniquely tied to the development and structure of the communications industry since its infancy. After the expiration of the basic Bell patents in 1893, interconnection restrictions became the fundamental weapon through which the Bell System monopolized both the long distance network and local exchange service in selected areas. Together with vertical integration, interconnection restrictions enabled Bell to virtually monopolize the equipment market as well. Thus, the monopoly structure of the communications industry was not founded upon public policy determinations that such an industry structure was more efficient than alternative, more competitive, industry structures, or necessarily in the public interest. Rather, the monopoly structure was the inherited condition. By the time the appropriate roles of federal, state and local regulatory authority were clearly delineated, the ownership patterns and industry structure were already established. Had the earliest state policy makers actively pursued the interconnection issue, a markedly different industrial structure would likely have developed.[5]

This conclusion is reinforced by the Carterfone decision which stated not simply that Bell's tariff restrictions against interconnection were unlawful, but that they "have since their inception been, unreasonable, unlawful and unreasonably discriminatory...."[6] In retrospect it would appear that the communications industry has been structured, at least in significant part, by the application of unlawful interconnection restrictions. Moreover, even though the most recent tariff filings by Bell are considerably more liberal than those that were replaced, we still do not have a public policy determination as to whether they are lawful and in accord with Carterfone.

The interconnection issue was initially addressed at the federal level in 1913 when the Independent telephone companies, responding to Bell's accelerating acquisition program, filed an antitrust complaint with the U.S. Attorney General. The intervention of the Justice Department led to the Kingsbury Commitment under which Bell agreed not to acquire control over any competing company and to interconnect the Bell network with Independent systems if they met Bell System equipment standards.[7]

It is interesting to note that the Independent sector of the telephone industry, for financial and other reasons, permitted substantial use of subscriber terminals on their lines. But when Bell took over Independent companies, Bell's absolute interconnection restriction against subscriber furnished terminals was applied. In many instances, the terminals remained in place but were simply sold back to the telephone company. When necessary, Bell was generally successful in eliciting the support of state commissions for this policy.[8]

At the level of federal regulation, the interconnection issue rose several times prior to Carterfone. Manufacturers of recording devices sought to enjoin Bell from applying its "foreign attachment" tariff restriction to stop them from selling their products. The FCC found in favor of the manufacturers but yielded to the position of Bell and the National Association of Railroad and Utility Commissioners (NARUC) that FCC should not impair State Commission freedom of action. Therefore, the FCC conditioned its authorization for interconnection with customer recording devices by requiring that all recorders be equipped with plug-jack arrangements to enable customer removal during the placement of intrastate calls.[9]

The Hush-A-Phone attachment was a simple, cup-like device that snapped on to the telephone and confined the speaker's voice within its inclosure, thus reducing noise and providing privacy. The FCC sustained Bell's interconnection restriction on grounds that the Hush-A-Phone affects the quality of service throughout the telephone network,[10] but was overruled by the Court. The Court of Appeals concluded that the interconnection restrictions in the tariff "are an unwarranted interference with the telephone subscribers right to use his telephone in ways which are privately beneficial without being publicly detrimental."[11]

In 1950, manufacturers sought clearance from AT&T for the interconnection of their electronic answering systems to the Bell network. Despite the earlier Recording Devices decision, which had appeared to establish the right of a telephone user to interconnect his own recorder, Bell viewed the devices as a violation of its tariff. In its decision in the Jordaphone case (1954), the FCC rejected the Jordaphone petition on grounds that there was no demonstration of demand for the product. "Complainants have not demonstrated the need for this device for interstate telephone calls. The Commission has no objection where local authorities have determined that a local need for them exists."[12]

Prior to 1959, the Bell System had interconnected with railroad communication systems under private contracts outside the tariffs. In 1959 Bell proposed to cancel its outstanding contracts with the railroads and restrict railroad interconnection rights to those special circumstances applicable to right-of-way companies. It will be recalled that 1959 was the year of the FCC's landmark Above 890[13] private microwave decision which permitted private microwave development beyond those communications uses that could not be met by the common carriers and permitted the sharing of private microwave systems by certain classes of users, including railroads. Thus, Bell's attempt at imposition of its interconnection tariff restrictions on the railroads was a competitive response to the railroads activity in pressing the FCC for expanded opportunities for private microwave development.

After an investigation by the FCC and a comprehensive examination of the issues on both sides, AT&T and the railroads negotiated a compromise which was accepted by the Commission.[14] The compromise agreement extended the interconnection rights of the railroads to microwave facilities. The FCC did not address the discriminatory character of the new agreement with the railroads. Bell then shifted its competitive activity from interconnection to pricing by introducing its Telpak private line tariff with rate reductions ranging from 51% to 85% of regular private line rates but with discriminatory eligibility requirements paralleling the FCC's new private microwave rules.[15]

With the advent of television broadcasting and the development of microwave technology, the provision of video communications became recognized as an important communications market in the late 1940's. Under Bell's initial tariff, it would not furnish intercity video transmission channels to a customer who planned to connect such channels either directly or indirectly with intercity video channels not furnished by Bell, except where Bell did not have the required facilities available. Then interconnection privileges would prevail only until Bell was prepared to provide the service.

As a common carrier, Western Union had gone ahead and constructed a microwave system between Pittsburgh, Washington and Boston. Bell refused to interconnect and Western Union sought an FCC order to interconnect the two carrier networks. In the ensuing FCC investigation,[16] Bell defended its interconnection on economic grounds that unrestricted interconnection has an adverse economic effect upon the business of the carrier and that monopoly supply of the service would be the most economical.[17] The Commission's decision noted that "it is admitted by the Bell System that one of the policies underlying the /interconnection/ regulations is to discourage development of competition in the field."[18]

The FCC based its decision reserving the television transmission market for common carriers on grounds of economy in the use of microwave frequencies. It upheld Bell's refusal to interconnect with Western Union on grounds that the latter had failed to establish a demonstrable need for the services. However, the Commission found Bell's interconnection restrictions to be unlawful in part, and stated that it would evaluate proposals for interconnection between common carriers on an individual proposal basis and determine whether the interconnection was "necessary or desirable in the public interest." Recognizing Bell's stated objective of monopolizing the intercity television transmission market and the failure of its initial effort at market entry, Western Union did not attempt entry again.

In the FCC's Domestic Telegraph Investigation of the early 1960's, Western Union petitioned that there be an industry-wide interchange of facilities with full interconnection for all services. Bell responded that interconnection between carriers is possible, but at an unwarranted price that must be paid in terms of impaired service quality and increased costs.[19] However, the FCC did not rule on the issue in the Telegraph Investigation because Western Union compromised on leasing interexchange channels from Bell. The Commission next came to grips with the issue in the *Carterfone* case.

Implications for Current Interconnection Issues

It is apparent from this historical review that *Carterfone* was not a precedent and policy shattering decision. It was a reinforcement and strengthening of the direction of past decisions. It addressed the interconnection issue as a matter of policy rather than as an *ad hoc* examination of only the *Carterfone* device.

Yet when contrasted with carrier practices, the *Carterfone* policy does appear revolutionary. This is because of the enormous gulf between policy and practice. Prior to *Carterfone*, despite the continuing regulatory and judicial declarations that tariff restrictions against interconnection were unlawful and that expanded interconnection must be permitted, there was little interconnection. Tariffs were applied almost as if there had been no change. Bell interpreted the *Jordaphone* and *Hush-A-Phone* rulings as exceptions to outright prohibitions against "foreign attachments" in the tariffs. The exceptions were interpreted as narrowly as possible and the general rule was still to refuse to interconnect with customer-provided attachments or systems.

In circumstances where it appeared that interconnection would be examined as a matter of general principle and policy, such as in the case of railroad interconnection, a negotiated agreement limited the liberalized interconnection privilege to a select group of users while avoiding a decision on the broad policy issue. Thus, it is evident that the probable consequences of a decision cannot be evaluated on the basis of the policy or principle enunciated. The policy pronouncement signals the beginning of the battle for change, not the end, as the *Carterfone* experience has verified. It provides official recognition that change is possible. Whether, and to what degree, the policy will be implemented depends upon the continuing advocacy of that policy by the regulatory agency and those who will benefit from the policy change at each stage of implementation. If both parties adopt passive roles under the assumption that the issue was decided in the policy decision, practices will not change significantly. Moreover, the status quo has a vested interest in delay.

The history of the interconnection issue shows quite clearly that it has always been an economic and not a technical one.[20] The fundamental motivation for restrictions has been market protection or monopolization. Over the years, it has been, perhaps, the most effective anti-competitive weapon in the Bell System's arsenal. The earlier debates focused directly on the basic economic interests of the contending parties, which is in contrast to the post-*Carterfone* debate where the emphasis has shifted to matters of technical harm.

Technical problems will arise in the supply of any product or service that is undertaken without planning, the development of standards, or adequate information flow and coordination among component suppliers, whether they be baking bread or preparing a space probe. This seems to have been recognized prior to *Carterfone* as a normal industry problem that was neither unique to communications nor central to the interconnection policy question. The issue was plainly one of competition and monopoly. Because the *Carterfone* decision declared as unlawful interconnection restrictions that are artificial economic weapons against competition, the economic debate now has been shifted by the Bell System to a framework of technical issues, standards, constraints and terminology. The result of the present debates on interconnection will determine whether the standards, certification procedures, tariff conditions and other terms of engagement are to be based upon maximizing or minimizing competitive opportunity.

III. Post Carterfone Responses

Certainly one of the most interesting observations subsequent to the *Carterfone* decision is that the dire consequences of expanded interconnection that were forecast by Bell have not occurred. In retrospect, it is difficult to find Bell representatives who will defend Bell's position in the case as a valid one. Apparently, it was all part of the continuing game of manipulating the regulatory process to one's economic advantage in preserving markets and blunting attempts to change policy that would reduce the scope of the Company's monopoly power. The only remaining advocate for the pre-*Carterfone* tariff restrictions is NARUC, which bases its position on a monopoly ideology rather than an analysis of either the economic or the technical aspects of the problem. However, noteworthy responses have occurred in the market, in the carrier's approach to the interconnection problem and in the regulatory role.

Market Response

Although opportunities for interconnection have not been as great and new market development has not been as fast as one would have expected following Carterfone, we have still seen the formation of a new industry. The variety and diversity of terminal devices has expanded enormously. Opportunities for innovative specialization have been pursued that have not only enabled users to make more efficient use of their terminals, attachments and computers, but also more efficient use of the telecommunications network. Under the new liberalized interconnection policies, the supply side of the market has been more responsive to the particular characteristics of consumer demands at reduced costs to consumers and with benefits to both independent suppliers and the telecommunications common carriers.

The "interconnect" industry, which presently is supplying equipment ranging from speaker phones to data modems to large private automatic branch exchanges (PABX) is presently installing facilities at a rate of more than $100 million annually. Some analysts have predicted that the market will grow to $2 billion by 1980, with a total potential of 6 to 8 billion.[21] After 40 months of interconnection, the Bell System had furnished 77,000 interconnecting units and entrance facilities to private microwave systems in 32 cases.[22] A modem suppliers' industry association /Independent Data Communications Manufacturers Association/ has been formed. Several tariff surveillance and analysis services have developed. User telecommunications network planning, management and marketing functions are beginning to receive detailed examination of a type never experienced in the old monopoly market environment.

Liberalized interconnection has the potential to open up as large a telecommunications equipment market through carrier systems interconnection as through terminals and attachments. Existing private microwave systems, the competitive development of which has been virtually foreclosed by interconnection restrictions and most of which are being used only at a fraction of their capacity, now have enormously expanded opportunities for development and expansion. Miscellaneous common carriers, which were developed solely to transmit television signals from Bell's major intercity routes to smaller towns and rural areas, and which previously could not send signals into the Bell network because of interconnection restrictions, can now compete for other markets, including the developing data markets. New specialized carriers such as Datran and MCI are likely to provide a substantial equipment market for independent equipment suppliers that could exceed $1.5 billion by the early 1980's if forecasted schedules of development hold up.

However, estimates of market size and growth rates must be quite uncertain at this stage of development for several reasons. First, the line of demarcation between the protected monopoly portion of the communications equipment market and that portion open to competitive suppliers has not been clearly drawn. AT&T presently has $15 billion invested in equipment on customers premises. If this were all exposed to competition from the interconnect industry, it would give independent suppliers access to 20% of Bell's present equipment and facilities investment and an even larger percentage of network system investment in the future when the data, specialized carrier, satellite and interconnect markets become more fully developed. The process of defining the nature and scope of the interconnect market as well as the terms and conditions of competitive engagement with Bell and the other vertically integrated common carriers will be going for many years.

Second, developments in the interconnect market will be influenced by regulatory policies in related areas such as specialized carrier competition, interservice subsidy, vertical integration, undue discrimination and pricing practices. In particular, the application of policies in these areas to developing computer-communication services and the ultimate criteria adopted for separating regulated communications services from unregulated data processing services will be crucial factors.[23] The regulatory rules for dividing communication and computer service markets carry with them an implicit division of the communication and computer equipment markets with attendant effects on the characteristics of the independent communications equipment market.

Thus, in several respects the equipment market represents the invisible portion of the communications iceberg.[24] It is heavily monopolized as part of an enormous utility holding company system. Yet, as a resource market, the equipment market has not been subject to regulatory scrutiny which has been traditionally limited to communication service markets. Further, as demand and technology change, the equipment market is not only broadened but it also can provide a base for the expansion into new equipment markets as a foundation for monopolizing new communications-related service markets. Indeed, Bell's equipment manufacturing decisions are now defining the scope of its future teleprocessing and computer-related services. If Bell elects, as it apparently has, to convert its specialized voice system into a specialized digital data system, then the range of communications-oriented teleprocessing services will be much greater than they otherwise would.[25]

Third, as the various dimensions of the interconnect market become defined at their more liberalized level, much depends upon the competitive response of the Bell System. If Bell, responds by specializing in those portions of the market that it can supply efficiently and establishes prices based upon costs, market development will be much different than it will under circumstances when Bell follows its traditional philosophy of trying to be everything to everybody and to impose all possible barriers to competitors. Finally, the nature of market development will depend significantly on how the federal and state regulatory agencies deal with the inevitable predatory competitive responses of Bell. Eventually, the issue will be reduced to pricing policy, cost justification and interservice subsidy practices. An examination of the experience with Bell's "competitive necessity" rate reductions in the Telpak and Series 11,000 tariffs does not bode well for competitors to Bell in the interconnect market. After a decade the FCC has yet to establish interservice subsidy standards or come to a decision on Telpak. The state commissions have yet to begin to come to grips with the issue.

The actual sales of communications equipment in the interconnect market by independent manufacturers have been generally substantially below the optimistic forecasts that were made by companies after the Carterfone decision. The effect of Bell's interface tariff has been to impose a cost disadvantage of varying magnitudes upon non-carrier equipment. This cost disadvantage is most significant for small communications users where there are fewer opportunities for independent suppliers to demonstrate a greater countervailing cost advantage. It is estimated that this alone excluded approximately one-third of the subscriber equipment market. For about another one-third of the market, user requirements for supporting services with their systems has constrained manufacturers from going directly to users and opened opportunities for interconnect companies.[26]

The new communications market that has yet to develop, but which may ultimately offer the greatest potential for innovation development is the communications-based service industry. The needs and opportunities for communication are dependent upon a community of interest between senders and receivers. Any particular communications user may be a member of several common interest groups or demand networks of communication and information. For most networks, the value to any individual member depends upon who else is a member of the market network. Actual communication service offerings may be inadequate because they fail to encompass the real communities of interest that underly demand networks or because they attempt to serve too great a conglomeration of diverse and specialized sets of demand networks.

User demand networks are quite different from the facilities networks that are engineered to provide various kinds of services within certain technological constraints. In the past, under basic monopoly conditions, communications users have had to adapt their demands to the terms and conditions of the facility systems and service offerings established. In economic terms, the supply side of the market has not had to be responsive to the demand side. Demand was forced to accommodate itself to supply.

Liberalized interconnection opportunities and the development of new special service common carriers can be partially responsive to the changing economic circumstances. But because these developments are essentially responses from the supply side of the market, they cannot be complete. Market opportunities cannot be fully developed without someone assuming the role of discovering evolving and potential communities of interest in specialized communications, organizing demand networks and determining their most efficient application to existing and planned communication facility systems and services. Demand network management is necessary for the maintenance of the supply/demand interactive process that is crucial to efficient market operation.

A communications-based service industry will likely develop along lines paralleling the evolution of computer-based services. The service industry will build on and follow changes in the equipment industry. It will focus on economical ways of using and adapting the new facilities and services and thus provide direction for the supply of future facilities and services. At present, some demand network management is being performed in certain circumstances and activity is increasing. However, paralleling interconnection, we find that tariff conditions limiting opportunities for circuit and rate sharing, brokerage, resale and the development of "authorized user" networks have been enormously restrictive. Fortunately, artificial tariff restrictions in these areas have begun to crumble as the FCC examines their justification and effects. Recently an RCA proposal to create a joint user network and act as network manager was permitted to proceed by the FCC staff. As movements in this direction intensify, the FCC will be forced to address the resale issue (in all of its forms) as a matter of regulatory policy.[27] Resolution of this issue will significantly influence the course of direction of demand network management and the communications-based services industry.

Carrier Response

After the Carterfone decision, Bell System pronouncements on interconnection indicated not only acceptance of the FCC policy decision but forecasts of the benefits of interconnection to all affected groups. In the 1968 Annual Report, AT&T Chairman Romnes stated: "Since customers now have more options in using the network, this should further increase usage and enhance the growth of our business." Elsewhere he stated: "We want to make the connection of such equipment as easy as possible, and the rules and regulations as few as possible....[28] In 1970, W. M. Ellinghaus, President of the New York Telephone Company stated: "We think these changes are good because they will stimulate use of the nation-wide telecommunications network while at the same time provide more options to our customers and more business for the manufacturers of communications terminal equipment."[29]

Statements such as these are more than diametrically opposed to the position adopted by Bell in every interconnection proceeding before regulatory agencies. They are clear statements of principle indicating that in its own view, Bell was following policies that were contrary to the efficient economic interests of both consumers and the Company. Apparently the only benefits of the restrictions were as a predatory device to limit competition and minimize the risks to be assumed by Bell management. Now that public policy finally has overcome Bell's monopoly power and influence of the regulatory process, in essence, the Company has announced that the FCC has forced it to act more efficiently in a way that directly benefits the Company.

However, once again principle and practice are turning out to be far apart. Interconnection refusals based upon extremely narrow interpretations of its new tariff, delays in service to interconnect companies and delays in the progress of the technical committees are the order of day. A running dialogue has developed between interconnect companies, the FCC and Bell regarding inconsistencies between Bell's new tariffs and Company practices. Yet, this should not be surprising since Bell has a vested interest in seeing that its decline in monopoly power takes place as slowly as possible and with maximum discouragement to competitors. To illustrate, identical equipment requires connecting arrangements if supplied by competitors, but does not if supplied by the telephone company. Recently Bell advised the FCC Common Carrier Bureau that formation of a data interconnect advisory committee is "premature and unnecessary" and would "strain the limited resources of all concerned..."[30] This position follows on the heels of Bell's revision of its 1980 data market forecast from $2 to $5-6 billion, the creation of a specialized in-house group for computer-communications and data services, the speeding up of installation of its data transmission capacity as well as its manufacturing capacity on competitive lines of equipment, and the filing of a 25% price reduction in certain data modems because of competitive necessity.

Although Bell is continuing to make the interconnection issue more and more a technical one of increasing complexity, other carriers have continued to view it as basically an economic matter. Western Union's response to the Department of Defense complaint regarding AUTODIN interconnection focused on the fact that substantial carrier investment might be lost to competition if interconnection restrictions were liberalized. The Rochester Telephone Company has filed a tariff in New York State with more liberal interconnection terms than Bell's tariffs, but has translated the interconnection competitive barrier directly into a differential pricing barrier between carrier-supplied and competitor-supplied equipment. Unfortunately, Bell's approach to this rapidly diversifying market is still one of trying to be all things to all people by fencing off all communications markets from others as best it can.[31] This approach may have worked during the era of Theodore Vail, but today it can only be harmful and inefficient for all concerned, including Bell.

Regulatory Response

Following the Carterfone decision, a special panel of the National Academy of Sciences (NAS) was appointed to assess the technical factors affecting common carrier/user interconnection. The panel reported on potential "harms" which can arise from interconnection, concluding that the network can be protected by means of connecting arrangements or direct connection under a program of standards and certification and that opportunities for innovation would be expanded with reduced restrictions on interconnection.[32] This was followed by a report to the FCC from Dittberner Associates suggesting alternative approaches to implement a standard and certification program that were less extensive than that suggested by the NAS report.

In March 1971, the FCC established its first technical advisory committee to study private branch exchange (PBX) interconnection. PBX equipment was selected from the wide range of terminal devices to establish a framework and guidelines for other equipment. It is expected to report to the FCC during the fall of 1972. To date, one other advisory committee to examine automatic dialers and recording and answering devices has been created.

On June 16, 1972, the FCC issued a Notice convening a Federal-State Joint Board to submit recommendations to the FCC on the issue of whether and under what conditions customers should have the option of furnishing network control signalling devices and any needed connecting arrangements.[33] Creation of the Joint Board is premised on the fact that interconnection affects local exchange and intrastate toll services which are subject to state regulatory jurisdiction. However, NARUC has never really accepted the Carterfone decision, has generally been opposed to every FCC policy change permitting opportunities to anyone other than the established telephone company, and apparently desires to retry the basic interconnection issues. Hopefully, the FCC headed off this retrenchment possibility by emphasizing in its Notice that "the soundness of our Carterfone decision has been amply demonstrated," new markets have been opened, the public has benefited, and "there has been no actual demonstrable harm to the system or its users." The current issue, the Notice stated, is "whether, and to what extent, there is public need for us to go beyond what we ordered in Carterfone."

Nevertheless, it must be recognized that the creation of the Joint Board will very likely slow down substantially the process of implementing interconnection policy changes from their already extremely slow and supercautious pace. In fact, the Bell-NARUC axis is capable of bringing the process of change to a complete halt, or even reverse it, unless advocacy from the other side is strong and the FCC plays an even more active role than it has since Carterfone.

IV. On The Engineering Of Economic Policy

Monopoly: Natural and Unnatural

Economic analysis of public utilities recognizes them as a special category of business wherein conditions of natural monopoly may exist. Natural monopoly exists when the technological alternatives are such that competition cannot work efficiently and must inevitably lead to monopoly because large scale monopoly supply is more efficient than smaller scales of competitive supply. The monopoly is "natural" because it is determined by the state of the technological alternatives available at a point in time. Legal barriers to the entry of competitors are frequently established. In order to protect the users of the monopoly services from economic exploitation, regulatory agencies attempt to ensure that the monopoly carrier charges reasonable rates and operates in the public interest. Local exchange telephone service is generally believed to be characterized by conditions of natural monopoly. However, interconnection restrictions have never been based upon the natural monopoly thesis.

It is important to distinguish between technologically determined "natural" monopoly and the acquisition or extension of monopoly power beyond the natural monopoly. Just because a carrier has monopoly power over certain markets does not mean that the existing degree of monopoly reflects the limit of natural monopoly.[34] This is particularly true of the Bell System and of the interconnection issue. In a recent text A. E. Kahn attempts to apply the natural monopoly thesis when he states: "The process of defining and redefining the area of natural monopoly is nowhere more clearly illustrated than in the evolving policies of the Bell System with respect to the kinds of equipment it permits customers to attach .... to the network and the circumstances, if any, under which it permits interconnection...."[35] It is not the area of natural monopoly that is being defined but the area of total monopoly, natural plus unnatural. Indeed, monopolists have never been known to graciously limit the scope of their power to that required by technological conditions.

The foregoing analysis clearly demonstrates that the interconnection issue has never had anything to do with the definition of natural monopoly and economies of scale. It has been directly concerned with the application of unnatural monopoly power to prevent the realization of economies of specialization. The debate has been over the ownership of equipment and the control of markets. Regulation will not have begun to define the limits of natural monopoly in this area until all artificial restrictions are removed and interconnect opportunities are equal for both common carriers and competitive equipment and systems.

Technical Standards and Economic Barriers

The problem of determining technical standards, certification procedures and other requirements is economic in that the issue is not that the functions cannot be performed, but which of a number of alternative sets of standards and procedures will be adopted. The standards adopted will establish barriers to entry to some potential suppliers. Thus, if Bell defines the standards, they will be extremely rigid so as to limit competition and protect Bell from all possible risks. In contrast, the interconnect industry would prefer much lower standards and the assumption of greater risks. To date, interconnection has been approached from the viewpoint of maximum protection and the minimization of risk to Bell. However, if interconnection policy is to be truly successful, it must be viewed as an opportunity. The benefits must be weighed with the costs and risks so that the resulting standards reflect the interests that make up the public's interests and not simply Bell's interest. If they are not, monopoly power and barriers to entry can be readily built into standards and procedures.

The structure of the U.S. telecommunications equipment market makes it entirely possible that interconnection standards could be established at levels that would constitute substantial barriers to entry. Aside from those companies affiliated with telephone holding company systems and Stromberg-Carlson, General Electric and ITT, the companies tend to be small and specialized. Recognizing that the ultimate standards will be significantly influenced by the advocacy positions of the various interest, it is entirely possible that the resulting standards could impose substantial competitive barriers on these smaller firms, the ones that could be most likely to innovate with their specializations.

Pricing: The Direct Competitive Response

As Bell's interconnection restrictions are overcome and competitors begin to have an impact upon markets, they will inevitably be subject to Bell's "competitive necessity" pricing response. Under existing regulation where the basic monopoly services provide all necessary residual revenues to cover Bell's revenue requirements, Bell is actually provided with an incentive to subsidize its competitive markets with revenues from its monopoly markets. Moreover, the interservice subsidy practices need not be confined to the company's service markets. It can be readily transferred to the manufacturing affiliate by means of the affiliates pricing practices in selling to the operating company. Vertical integration multiplies the cross-subsidy possibilities enormously.

We have recently seen the first major competitive pricing response by Bell when it reduced rates for its Series 201B data set from $72 to $55 per month, and installation charges from $100 to $75 per month. The price reduction was justified in terms of a decline in Western Electric's price for the modem by 26%, a sudden reduction in the annual charge factor for maintenance to less than a third of the previous factor, and a depreciation life estimate of 8 years for rate purposes even though service life experience is currently 3.9 years. Such an approach to costing is only possible when a firm has a captive market to supply all residual revenue requirements. Interconnection policy cannot succeed until regulatory standards are established that will enable commissions to detect and prevent anticompetitive interservice subsidy pricing practices both among communication services and lines of equipment.

V. Conclusion

Just as the antitrust laws provided the initial step in the evolution of the Carterfone interconnection policy, it is likely that they will be the best weapons to keep interconnection markets open and anticompetitive practices of telephone companies limited. The real development of interconnection markets is still on the horizon awaiting the development of standards which is likely to progress at an even slower pace than in the past.

The issue of technical harm has been generally blown out of proportion to its significance. "The greater speeds and complexity of computer peripheral devices makes the interconnection problems in telecommunications simple by comparison."[37] The standards and certification problems should be recognized as essentially economic in nature and developed in terms of consideration of the benefits and costs of maximized interconnection opportunities subject to the assumption of reasonable risks. A continuing challenge for standards development will be to ensure that they will accommodate changing technology and not simply rigidly impose today's technological constraints upon tomorrow's technology. Factoring out the influence of monopoly power in standards determination and preventing predatory competitive tactics by Bell, especially in the pricing area, will continue to be the crucial regulatory problems.

The nature of the interconnection problem is directly tied to the vertical integration of the telephone companies. Standards development would be proceeding much faster and in a much more favorable environment if the carriers did not have a strong economic interest in maintaining a favored market position for a manufacturing affiliate. Indeed, were it not for vertical integration, interconnection would be a problem of much smaller magnitude than it is today.

Developments in interconnection demonstrate that regulation in communications can no longer proceed on the basis of ad hoc policy making with the individual firm as the focal point of regulatory examination. Regulation must develop a competition policy that specifically addresses the technological limits of natural monopoly and the structure of communications markets. Market structure policies which delineate the scope and type of competition desired, the terms of entry that will be promoted, and the anticompetitive activities that will be declared illegal, must be developed. Then regulation can establish a framework for the implementation of specific policies in interconnection and other areas that will govern the growth and development of computer communication services.

When viewed in total perspective, the interconnection issue provides a very telling experience about the efficacy of traditional communications regulation. Artificial restrictions that should never have been have structured the industry. The monopoly power created has prevented and is retarding a full and speedy adjustment today. Interconnection developments illustrate all too clearly how the regulatory process has been abused as an extra-market force to be influenced and manipulated to advantage in accordance with the economic power positions of the firms in regulated markets. It is time for a searching examination of the regulatory decision making process with a view toward neutralizing the influence of established monopoly power, creating a permanent advocacy force for unrepresented interests (particularly users of monopoly services), and requiring continuous policy analysis and justification by the regulatory agency of the extent of monopoly power warranted by technological conditions and the public interest.[38]

References

1. Regulatory and Policy Problems Presented by the Interdependence of Computer and Communications Services and Facilities, Notice of Inquiry, FCC 66-1004, Nov. 10, 1966. Tentative Decision and Notice of Proposed Rule Making, Docket No. 16979, FCC 70-338, April 3, 1970. Memorandum Opinion and Order, FCC, March 28, 1972. For an analysis of the issues see Mathison, S. L. and Walker,

P. M., Computers and Telecommunications: Issues in Public Policy (Prentice-Hall, 1970).

2. Carterfone Corp., 13 FCC 2d 420. See related decisions at 14 FCC 2d 571; 15 FCC 2d 605; 18 FCC 2d 871; 19 FCC 2d 1068.

3. 13 FCC 2d 420, 424.

4. For a more detailed review and analysis of many of the cases, see Gabel, Richard, "Telecommunications Interconnection - Wherefrom and Witherto?" International Symposium on Communication: Technology, Impact, Policy, University of Pennsylvania, March 1972.

5. Op. cit., p. 5.

6. 13 FCC 2d 420, 426.

7. AT&T Annual Report to Stockholders, 1913.

8. Gabel, Richard, op. cit., pp. 10-11.

9. Use of Recording Devices in Connection With Telephone Service, FCC Reports, Vol. 11, 1036; Vol. 12, 1005.

10. Hush-A-Phone Corp., FCC Reports, Vol. 14, 282 (1949).

11. 238 FCC 2d 266 (1956).

12. Jordaphone Corp., FCC Reports, Vol. 18, 644 (1954).

13. 27 FCC 359 (1959); 29 FCC 825 (1960).

14. Railroad and Right-of-Way Interconnection with Telephone Companies, FCC Reports, Vol. 32, 337.

15. See Melody, William H., "Interservice Subsidy: Regulatory Standards and Applied Economics," in Harry M. Trebing, ed., Essays on Public Utility Pricing and Regulation (East Lansing, Michigan: Institute of Public Utilities, Michigan State University, 1971), p. 170 ff.

16. FCC Docket No. 8963, 5 RR 639.

17. Trebing, H. M. and Melody, W. H., An Evaluation of Domestic Communications Pricing Practices and Policies, in President's Task Force on Communications Policy, The Domestic Telecommunications Carrier Industry (Staff Papers, June 1969), Chapter VII.

18. Charges and Regulations for Television Transmission Services and Facilities, Report of the Commission, Docket No. 8963, p. 62.

19. Telpak Rates and Charges. FCC Reports, Vol. 37, 111 (1964).

20. Gabel, Richard, op, cit., p. 37.

21. See, e.g., "The Revolution in the Phone Business," Business Week, November 6, 1971. Communications Investor (Paul Kagan Associates), April 20, 1972. Frost and Sullivan Report on developments in the communications industry, January 1972. Interconnection, newsletter of the new interconnect industry.

22. Telecommunications Reports, August 14, 1972, p. 12.

23. Irwin, Manley R., "The Computer Utility: Market Entry in Search of Public Policy," Journal of Industrial Economics, July 1969.

24. Irwin, Manley R., The Telecommunications Industry: Integration vs. Competition (New York: Praeger, 1971).

25. This, of course, raises critical and complex issues of interservice subsidy. See Melody, William H., "Interservice Subsidy: Regulatory Standards and Applied Economics," op. cit.

26. International Telephone and Telegraph Co. vs. General Telephone and Electronics Co. and Hawaiian Telephone Co., Decision of the U.S. District Court for the District of Hawaii, July 14, 1972, p. 35.

27. Microwave Communications Inc. has recently petitioned the FCC to address the resale issue as a policy matter. Telecommunications Reports, June 5, 1972, pp. 4-8.

28. 195 Magazine, September 16, 1968.

29. Ellinghaus, W. M., "Meeting Future Communications Requirements," Conference on the Revolution in Business Communications, New York, New York, August 31, 1970.

30. Letter from AT&T Vice President Thomas W. Scandlyn to FCC Common Carrier Bureau, July 28, 1972. See Telecommunications Reports, July 31, 1972, p. 21.

31. See interview with AT&T Chairman John D. deButts, Telephony, April 10, 1972, p. 19.

32. National Academy of Sciences, Computer Science and Engineering Board, "Report on Technical Analyses of Common Carrier/User Interconnection," June 10, 1970.

33. Notice of Inquiry, Proposed Rule Making and Creation of Federal-State Joint Board, FCC Docket 19258, June 16, 1972.

34. Melody, W. H., "Technological Determinism and Monopoly Power in Communications," American Economic Association Meetings, New Orleans, La., Dec. 28, 1971.

35. Kahn, A. E., The Economics of Regulation, (Wiley, 1971) Vol. II, p. 140.

36. Irwin, Manley R., op. cit., Chapter VII.

37. Gabel, Richard, op. cit., p. 38.

38. See Melody, William H., "The Role of Advocacy in Public Policy Planning, Symposium on Communication: Technology, Impact and Policy, University of Pennsylvania, March 1972.

BEYOND THE COMPUTER INQUIRY
(Who Should Be Regulated in Computer/Communications)

Charles R. Cutler
District of Columbia Bar
Washington, D.C., U.S.A.

Summary. The efficacy of common carrier regulation has come under increasing challenge in the past decade. However, there is little likelihood that regulation will soon be abolished in the field of computer/communications and it is timely, now that the FCC's decision in the *Computer Inquiry* has been released, to reexamine the applicable criteria for regulation.

The FCC's decision in the *Computer Inquiry* provides only limited guidance. It concluded that pure data processing need not be regulated, but left other questions unanswered such as, in what instances will message switching and other communications activities actually be subjected to regulation?

Past precedents of the FCC do provide guidance for determining who should be subjected to regulation in the computer/communications field. Taking into account these precedents, and the traditionally accepted conditions and goals for utility regulation, the paper suggests standards to be applied in regulating computer/communications activities.

* * * * *

FOR many years, physicists examined smaller and smaller particles of matter/energy in order to determine mass, location, momentum, velocity, etc. They still do. But, in the early part of this century, one could be led to believe that if we just knew where all the atomic particles were, and which way they were headed at any instant -- like billiard balls -- theoretically we could predict future events with certainty. This belief was shattered by the now accepted principle of "indeterminancy" or "uncertainty" proffered by physicist and 1932 Nobel Prize winner, Werner Heisenberg. His proposition was that, *as a matter of principle*, one cannot know *simultaneously* the precise: (1) location; and (2) velocity, of a particle.

As I have watched regulators, along with regulatees, potential regulatees, and interested consumers of communications struggle year after year in FCC hearings as the FCC seeks to determine where each is, and where each is going, I have been continually reminded of Heisenberg's principle. Perhaps we can regard the familiar "regulatory lag" in utility regulation as just a special case of Heisenberg's theory, because it often seems impossible to fashion regulatory tools to ascertain the location and trajectory of regulatory objects in time to permit the regulator to act effectively.

In earlier decades, there was a confidence in government regulation which has faded just as surely as has the assurance of the proponents of determinism in the physical sciences. Indeed, there are many arguments being heard today on both practical and theoretical grounds to the effect that agency regulation is ineffective and that it should be curtailed, revamped, or abandoned -- and these arguments are generally coupled with the suggestion that increased reliance be placed on antitrust enforcement and deterrent.[1]

In this paper, I shall avoid that theoretical debate in that I shall assume that the established common carriers such as AT&T, Western Union, and GT&E will continue to be regulated much as at present. This discussion, then, relates to persons who are not now common carriers who render, or wish to render, computer/communications services. Of these, who should be regulated? And, by "computer/communications services,"[2] I mean to include such operations as data processing, message switching, and multiplexing, together with the associated utilization, sharing, or brokerage of channels of communications obtained from an established carrier.[3]

We are now, of course, in a position to examine this question with the benefit of the FCC's official views as of March 18, 1971, the date it issued its Final Decision in its inquiry into the *Regulatory and Policy Problems Presented by the Interdependence of Computer and Communications Services and Facilities*. [4] That decision, while commendable for what it did say, left more questions unanswered than is generally supposed. Let us look at it briefly.

The FCC's Decision in Its *Computer Inquiry*

The issues in the *Computer Inquiry* were wide ranging and included: the adequacy of common carrier rates and services, the threats posed to the right of privacy, and the question of what computer/communications activities should be subjected to FCC regulation. With respect to the first two issues, the Commission's decision deferred action to other proceedings and tribunals, but it did purport to indicate what activities in the computer/communications fields should be regulated. It is that aspect of the Computer Inquiry which we discuss below.

"Data processing," the FCC said in its decision, is a sufficiently competitive business so that the forces of the marketplace should be allowed to prevail without imposing regulation. However, to help the market stay competitive, it told the major established communications common carriers that they would each have to render their data processing services to the public through a separate, dissimilarly named, corporate entity. And, the Bell System was to forbear altogether. As for "message switching," that was a different matter.[5] Message switching, the FCC concluded, had always been considered to be a "communications" function, and, therefore, is "subject to regulation." If a hybrid service involved both data processing and switching, then the service should be examined on an ad hoc basis to determine whether the service was "primarily data processing" or "essentially communications" in order to determine its vulnerability to regulation.[6] And, as an enforcement tool to help the FCC learn of any non-common carriers who were trying to provide "essentially communications" services, the FCC suggested that the established common carriers use their anti-resale tariff clauses as a means of drawing this to the attention of the Commission.[7]

This was a logical and rather appealing disposition of a complex problem. But, when one turns away from the written page to the complicated world, one realizes the scope of what was left unsaid.

The decision not to regulate pure data processing remains clear enough. But the Commission's view that message switching is a "communications" function and is therefore "subject to regulation" raises a host of questions.

Even setting aside the troublesome "hybrid" systems, the regulatory status of pure message switching really cannot be determined by a reading of the FCC's decision in the Computer Inquiry, although at first reading one might assume so. For example, would anyone suppose that the FCC has any intention -- or desire -- to regulate a computerized message switch operated by, say, General Motors simply for internal company purposes, even though it embraced scores of plants, locations, and divisions? Or, would the Commission plunge in if the system were expanded to provide two or three of GM's regular suppliers access to the switch? Or, suppose the electric power industry decided that a message switching system was desirable for all power companies in the East, with one company operating the switch for the industry? Or, suppose the industry forms its own communications company to do the job on a cost-sharing basis? Or, even retains an independent computer service company to do the switching for the industry?

The FCC's decision in the Computer Inquiry does not tell us at what point the pure message switching functions described above would be placed under regulation, if ever. In short, the FCC announced what might be regulated, but not who, when, or how.

We need not limit our examples to message switching. The rationale applies even to communications circuits and channels, which are surely "communications." The simple fact is the FCC has never tried to regulate every person who makes a communications circuit or channel available to another. The question always is, at what point will the FCC decide that the provision of a communications service by one person to another is subject to regulation?

Earlier FCC Precedents

Accepting the fact that the decision in the Computer Inquiry is of limited value in deciding "who" should be regulated, we will take a look at the Communications Act of 1934 itself and what the Commission has said in the past.

The bare language of the Act is of little help to us. It merely defines a "common carrier" as "any person engaged as a common carrier for hire, in interstate or foreign communication ... "[8] We will, however, have occasion to note the use of the phrase "for hire" later.

More useful than the bare language of the Act is its legislative history. The Conference Report to the 1934 Act enunciated a principle which, though seemingly innocuous, is significant:

> "It is to be noted that the definition [of common carrier] does not include any person if not a common carrier in the ordinary sense of the term ... ."[9]

The importance of this language is that Congress apparently did not intend the Commission to seek out every possible activity which could be embraced under some regulatory precedent, but rather included only one who is a common carrier in the "ordinary sense of the term." The Commission has taken note of that limitation in its decisions.[10]

Looking to past actions of the FCC and its staff, one can formulate some principles as to whether a person is acting as a common carrier "in the ordinary sense of the term." Here are the basic ones:

1. Scope of Those Served. A common carrier is one who offers its facilities or services to the public. Although one can debate how large a class of users must be before it becomes "the public," it is clear that one is more likely to be held a common carrier if services are offered to the public at large than if service is available only to a restricted group. For example, in the Computer Inquiry when discussing its regulatory obligation toward those hybrid services which were "essentially communications," the FCC referred to "a public offering."[11]

Since the concept of common carriage involves service to the "public," if an organization is providing communications to itself, then the presence of common carriage is negated. Extending that reasoning, if an entity's primary purpose is to render communications service essentially to those who own or control it, the concept of providing service to "one's own" stands in contrast to the provision of communications services to the "public." Implicit also in the rendition of services to customers who control the vendor is the fact that it is likely that the charges and terms of service will be fair and equitable, else the management of the vendor will be turned out. Thus, whether the entity is a corporation, or an association, or merely a group engaging in "sharing," the provision of the services should be less likely to be viewed as common carriage if there is some relationship between those served by the entity and those who own it.[12]

2. Are the Communications Services Rendered For Profit? The Commission has generally construed the concept of "for hire" which is contained in the statutory definition as meaning "for profit." In a wide range of formal and informal actions and inquiries, the Commission and its staff have for many years indicated that the presence or absence of a profit motive is a strong, and sometimes decisive, factor in determining whether the function is that of common carriage.[13] The most recent language of the Commission juxtaposing the concept of "for hire" and "for profit" occurred in a Memorandum Opinion and Order released August 24, 1971, in which the Commission noted that generally where private microwave facilities are authorized, "The licensee of the facilities, the person to whom the privilege of use of the channel is accorded, is not permitted to profit from the use of the facilities authorized to it, unlike the case of common carriers, where the licensee is permitted to make the facilities authorized to it available to the public 'for hire.'"[14] While the Commission is not likely to concede that not-for-profit operation guarantees exemption from common carrier regulation, it is surely a most persuasive negative factor.

3. Interest In or Control Over Content of Transmissions. The Commission has indicated that one distinction between a common carrier and a non-common carrier is whether or not the "carrier" has an interest in or control over the content of the transmission. If the organization handling the transmission does have an interest in the message itself, as distinguished from merely providing the facilities over which it is sent, common carriage is negated. As the Commission said in the course of a comprehensive statement in 1966, "[A] carrier provides the means or ways of communications for transmission of such intelligence as the customer may choose to have transmitted so that the choice of the specific intelligence to be transmitted is the sole responsibility or prerogative of the customer not the carrier."[15] Indeed, in 1937 the FCC observed, where an organization sought to obtain common carrier circuitry to serve others in the process of fulfilling its own bona fide business purposes, that "It must be concluded that no interest in the content of the messages handled is necessary, but it is sufficient if the fact that the message is sent aids the lessee [of common carrier circuits] in carrying out purposes of its own distinct from those of a communication carrier for hire."[16]

There is no reason to expect that the FCC will ignore in future ad hoc examinations of computer/communications services the three criteria which we have just discussed. But, before pursuing that line of thought further, let us examine the question of regulation itself.

Does It Matter Whether You Are Regulated?

The preceding discussion has assumed that if a communications service is to be regulated, it will be regulated as a common carrier activity under Title II of the Communications Act of 1934. The FCC indicated that this would be the case in its Computer Inquiry decision.[17] However, the Commission also asserted that it had wide latitude within the framework of common carrier regulation, saying, "It also should be pointed out in this regard that, within the statutory scheme of regulation applicable to interstate common carrier services, the Commission may exercise appropriate discretion as to the methods and policies that will best serve the public interest in the regulation of hybrid communications services."[18] Thus, the Commission seems to be saying that because a communications service may warrant regulation as a common carrier does not necessarily mean that the full impact of traditional common carrier regulation need fall on the hapless provider of "hybrid" services. This would be welcome news because Title II of the Communications Act of 1934 provides a pervasive scheme of regulation of common carriers which could be quite onerous.[19] It includes, for example, rate regulation whereby rates for services must be "just and reasonable," and must not constitute an "unreasonable discrimination." (§201(b), §202 (a)). In enforcement of this power, the FCC is empowered to suspend rates and practices for as long as three months, to investigate rates, and, after investigation, prescribe new rates and practices for any interstate communications common carrier service (§§204, 205). And, of course, common carriers must file complete tariffs with the FCC describing the nature of the service and the charges and practices applicable to them (§203). A critical question would be the manner in which the provisions of §214 of the Act were applied by the FCC. These require a common carrier to apply for and obtain from the FCC, after public notice, a certificate of public convenience and necessity before it "shall acquire or operate any line, or extension thereof." The definition of "line" includes any interstate "channel of communication established by the use of appropriate equipment" of more than ten miles in length. Indeed, the Commission may even direct a carrier to expand its service by means of additional facilities (§214(d)). These specific powers of the Commission are in addition to its general management controls over common carriers which it exercises through its promulgation of accounting practices, financial reporting, etc. Further, as a common carrier, one is not automatically entitled to obtain the tariffed services of another common carrier, such as AT&T. These services must be obtained by negotiation. If service is refused, the FCC may order the interconnection of service, but this may require a hearing (§201).

Thus, the concern over whether or not an operation is subject to common carrier regulation is not a trivial one. The Commission's previously quoted assurance that it has "appropriate discretion" provides some basis to hope that the Commission would try to minimize what it considered to be unnecessary inconvenience and expense for a regulatee.

If the choice were between full-blown common carrier regulation and some general "standards of behavior" promulgated under the Commission's general authority to regulate interstate communications,[20] the latter is preferable. However, it may be that by artful fashioning of its common carrier tools, it can, in those areas where regulation seems necessary, set up a scheme of regulation which is not unduly oppressive.

Who Should Be Regulated?

Having discussed some of the principles which the FCC has applied in determining who is a common carrier, and having considered briefly what "regulation" might mean, it is timely to give closer thought to who really should be regulated.

We can first take a look at the conditions which are generally accepted as requiring or justifying common carrier or utility-type regulation, forgetting for the moment any statutory mandates requiring a regulatory agency to impose regulatory restraints. The conditions relate to the possible emergence of monopoly, either because monopoly may be feared, or because it may be felt desirable. This sensitive condition is said most likely to exist where the field of operation has such high threshhold costs that entry as a practical matter is limited to a few firms. Further, if the nature of the services offers substantial economies of scale, not only is monopoly more likely to occur, it may be in the public interest to suffer monopolistic operation in order to gain the efficiencies and economies that will result. Once the conditions are ripe for a natural monopoly, the argument goes, regulation is necessary in order to keep the beast under control. Commonly accepted goals of this control are: (1) prevention of unreasonably discriminatory pricing; (2) assurance of a satisfactory level and scope of service to the public; (3) assurance that public resources are efficiently utilized; and, perhaps, (4) elimination of artificial restraints which the dominant firm(s) might impose upon new entrants into the field.[21]

Now let us compare the above factors, i.e., the "conditions" which are generally said to suggest that a field is ripe for regulation and the "goals" of such regulation, as they apply to the computer/communications world.

The FCC found in its *Computer Inquiry* that the "conditions" for regulation did not exist in the data processing industry -- no high threshhold costs or economies of scale, and no tendencies yet discernible toward monopoly or oligopoly.[22]

But, what of the "natural monopoly" characteristics of "message switching," or the selling, sharing, or arbitraging of communications channels obtained from established common carriers, or the furnishing of facilities (such as multiplexing equipment) -- with or without data processing? The Commission made no specific findings directed to that question, reciting only, in effect, that because switching was a "communications" function, it may have no choice but to impose regulation under its governing statute.[23] However, the subject has been considered more fully elsewhere. For example, the President's Task Force on Communications Policy concluded that "neither pure store-and-forward switching nor hybrid services seem to involve natural-monopoly elements"; nor did the Task Force find any persuasive obstacle to unregulated "line sharing and brokerage" in connection with such services.[24] Similarly, the Department of Justice in its Comments filed with the FCC in the *Computer Inquiry* found "no policy reason" for regulation of pure message switching, noting that it "involves none of the attributes of a rate-regulated monopoly."

If the conditions in the computer/communications field are characterized more by competitive equilibrium than natural monopoly, what then would be the purpose of regulating one, say, who switches messages over shared lines? Would it be to prevent discriminatory pricing? To assure satisfactory service? To assure efficient utilization of resources? To prevent artificial restraints to market entry? These traditional goals of regulation don't seem to fit the existing situation. Paradoxically, where the conditions for natural monopoly can be argued persuasively, that is in an unrestricted nationwide network of intercity circuitry, the FCC has embarked on a policy of "open entry" through its *Specialized Common Carrier* decision.[25]

Yet, a *user* of common carrier facilities not only finds himself in a labyrinth of tariff restraints (notwithstanding the blessing of regulation), but also is intimidated by the spectre of becoming a common carrier himself if he "switches" more than he "processes," or even "switches" separate from what he "processes." He fears a similar fate if he resells rather than "jointly uses" a circuit obtained from a common carrier; or markets the use of equipment (such as a multiplexer) which is considered a communications "*facility*" under the Communications Act of 1934.[26]

This leads us to a philosophical question. Should the decision to regulate be based primarily on *what* is done (e.g., whether or not a traditional "communications" function is performed), or, rather, *who* does it (a little non-profit fellow vs. a big profiteer), or, perhaps, *how* it is done? In the *Computer Inquiry*, the FCC seemed to emphasize the *what* when it said that "hybrid services which are 'essentially communications' ... warrant appropriate regulatory treatment as common carrier services under the Act." Yet, even in that same paragraph of its decision, the FCC implied that its obligation to regulate would depend on whether there was a "public offering" of the services, and that it could exercise "appropriate discretion" in deciding *how* to regulate, and, in any event, "we see no need at this time to attempt a definitive resolution" of the extent of the regulatory obligation.[27]

Having announced that it believes it has discretion in how to regulate, though not unlimited discretion in *whether* to regulate, the Commission said it would prefer to implement its discretion on an *ad hoc* basis rather than by formulating elaborate guidelines. The latter, it feared, might inhibit ingenuity and responsiveness of interested parties who felt constrained by hypothetical cases posed by the guidelines.[28]

One can grant the reasonableness of that conclusion and still believe that ingenuity and responsiveness may be inhibited by the absence of guidelines. Therefore, with appreciation for the complexity of the subject, I would make some modest suggestions.

First, the FCC should make it clear that it is continuing to apply its traditional "scope of service," "for profit," and "interest in or control over transmissions" tests to the computer/communications field in deciding whether a proposed operation really is common carriage under the statute. These criteria for exemption should be interpreted broadly so as to free as many activities as possible from regulation. This would enable industries which in the past have not taken full advantage of the new technology to create service organizations, as needed, without undue fear of falling under a net of red tape in the form of common carrier regulation.

Second, for other service organizations, not so exempt, providing computer-associated "communications" functions by means of interstate channels of communications obtained from established common carriers the absolute minimum of regulatory reporting and compliance should be required, and the nature of these obligations should be made known so that they may be properly weighed by those entering the business. Nothing more than the bare minimum would seem justified since the traditional regulatory goals of conservation of public resources, assuring adequate service, and curbing restraints on entry would not seem to be in grave jeopardy as long as the field were competitive. That is, since the interstate channels of communication are being provided by established common carriers, we can assume that the regulatory process has assured conservation of public resources (e.g., radio spectrum). A good level of service would be preserved by adequate technical standards of interconnection. So long as the field remained competitive, as assumed, new entry would not be threatened. The possibility of predatory practices would exist, but these are subject to antitrust deterrents as well as prohibitions under Title II of the Communications Act.

If the Commission had concluded in the Computer Inquiry that it was free to treat these activities outside the scope of common carriage, but felt that some regulatory attention was necessary, it could under its general authority to regulate interstate communications have promulgated general standards of conduct which it felt were necessary.[29] In effect, these could have been phrased in terms of precluding anti-competitive practices, giving added specificity to existing antitrust principles. Having decided that such an approach was either undesirable or unavailable under the statute, the Commission should now devote its ingenuity to making its Title II regulation as diaphanous as possible.

FOOTNOTES

1/ For some of the recent discussions which question the efficacy of agency regulation with frequent emphasis on the alternative of increased reliance on antitrust deterrents, see: Kohlmeier, The Regulators ch. 7 (1969); Computers, Communications, and the Public Interest ch. 7 (M. Greenberger ed. 1969) (discussion among Richard Posner, Nicholas Johnson, Lee C. White, and Kingman Brewster); Hector, Problems of the CAB and the Independent Regulatory Commissions, 69 Yale L.J. 931 (1960); Address by Lee Loevinger, The Sociology of Bureaucracy, International Conference on Communications of the IEEE, June 13, 1968; The Crisis of the Regulatory Commissions (P. MacAvoy ed. 1970); President of the United States, Economic Report (1970); President's Task Force on Communications Policy, Final Report (1968); Posner, Natural Monopoly and Its Regulation, 21 Stan. L. Rev. 548 (1969); Turner, The Scope of Antitrust and Other Economic Regulatory Policies, 82 Harv. L. Rev. 1207, 1231 (1969); Comments of the Department of Justice, esp. at 68-71, FCC Computer Inquiry (1969) (Docket No. 16979); Address by Clay T. Whitehead, Director, Office of Telecommunications Policy, The Conference Board, Feb. 15, 1972. For a somewhat less pessimistic view of regulation, despite its title, see Trebing, Common Carrier Regulation -- The Silent Crisis, 34 Law & Contemp. Prob. 229 (1969); see also, 2 A. Kahn, The Economics of Regulation: Principles and Institutions 108-12, 324-29 (1971).

2/ Recognizing the awkwardness of "computer/communications services" as a descriptive term, I cannot yet quite bring myself to use the term "compunications" offered by Anthony G. Oettinger, in his Communications in the National Decision-Making Process in Computers, Communications, and the Public Interest (M. Greenberger ed. 1971). "Teleprocessing" is a more palatable word, but may suggest to some that only data processing is done at the end of the line -- so, the old terminology is used.

3/ It is appropriate to observe here that line sharing or arbitraging which might be involved raises more than merely the question of whether the one who is doing it should be regulated. There are also the questions of whether unfettered sharing or arbitraging destroys the common carrier's competitive or cost justification for offering bulk-discount rates, and of what restrictions against sharing or resale the common carrier may properly include in its tariff. It is not intended here to address those questions. The fact is that various common carrier tariffs now do permit different kinds of shared, joint, and derivative use of their facilities and services.

4/ Tentative Decision, 28 F.C.C.2d 291 (April 3, 1970) [hereinafter cited as Tentative Decision]; Final Decision, 28 F.C.C.2d 267 (March 18, 1971) [hereinafter cited as Final Decision] (Docket No. 16979). The decision is now on appeal to the United States Court of Appeals for the Second Circuit, pursuant to appeals by common carrier parties.

We should note that the Common Carrier Bureau initiated this proceeding in 1966, hoping to neutralize "regulatory lag" by anticipating problems before they became serious and imbedded. Thought provoking and informative speeches by the FCC's Common Carrier Bureau Chief, Bernard Strassburg, provided a valuable backdrop to the Inquiry, e.g., The Marriage of Computers and Communications -- Some Regulatory Implications, Oct. 20, 1966; The Communications Carrier and Management Information Systems, Oct. 21, 1965.

5/ The FCC's definitions of data processing and message switching appear in Appendix A to its Final Decision:

> "(1) 'Data processing' is the use of a computer for the processing of information as distinguished from circuit or message-switching. 'Processing' involves the use of the computer for operations which include, inter alia, the functions of storing, retrieving, sorting, merging and calculating data, according to programmed instructions.
>
> "(2) 'Message-switching' is the computer-controlled transmission of messages, between two or more points, via communications facilities, wherein the content of the message remains unaltered."

6/ Tentative Decision 304-305; Final Decision 276, 278. (Because the Final Decision accepts so much of the reasoning of the Tentative Decision in this sector, reference to both documents is useful.)

7/ Final Decision 279. One organization has since sought to have these anti-resale tariff restrictions removed. See Letter from counsel for Electronic Data Systems (EDS) to Kelley E. Griffith, Chief, Domestic Rates Division of the FCC's Common Carrier Bureau, Dec. 6, 1971; the response of AT&T to Mr. Griffith, Feb. 10, 1972; and the rejoinder by counsel for the Business Equipment Manufacturers Association (BEMA) directed to Mr. Griffith on April 21, 1972.

8/ Communications Act of 1934 §3(h), 47 U.S.C. §153(h) (1970).

9/ Conference Report on Communications Act of 1934, H.R. Rep. No. 1918, 73d Cong., 2d Sess. 46 (1934).

10/ E.g., Frontier Broadcasting Co., 24 F.C.C. 251, 254 (1958) (regulatory status of CATV systems). In that connection, we should observe that the body of precedent established by the FCC in determining common carriage under its statute is unique to it, and no attempt is made in this paper to survey the general field of common carriage. For an earlier discussion of FCC precedents in this area, see Computer Services and the Federal Regulation of Communications, 116 U. Pa. L. Rev. 328 (1967). See also Smith, The Interdependence of Computer and Communications Services and Facilities: A Question of Federal Regulation, 117 U. Pa. L. Rev. 829 (1969).

11/ Final Decision 278. In other contexts, the FCC has sought to learn of an organization's "relationship to the public" or whether "any" party may participate, in the course of deciding whether the operation was common carriage. See FCC's correspondence with Scantlin, Timeplex, Inc., and Series 11,000, Inc., cited note 12 infra. It emphasized a "public offering" to "all members of the public" in its Frontier Broadcasting and Industrial Radiolocation opinions. This does not mean, however, that a common carrier may not restrict its offering to a "special classification of service." See Industrial Radiolocation Serv., 5 F.C.C. 2d 197 (1966); Allocation of Frequencies for Theatre Television Service, 9 P&F Radio Reg. 1528, 1538 (1953).

12/ See Cooperative Sharing of Operational Fixed Stations, 4 F.C.C. 2d 406, 417 (1966); Multiple Licensing -- Safety and Special Radio Services, 24 F.C.C. 2d 510, 511 (1970); accord, Special Emergency Radio Service, 24 F.C.C. 2d 310 (1970).

13/ See, e.g., recent letters from the Common Carrier Bureau to: Scantlin Electronics, Inc., Feb. 5, 1971; Timeplex, Inc., Feb. 5, 1971; Series 11,000, Inc., Aug. 20, 1970 (in each case inquiring about the profit aspects of a proposed activity). For earlier actions relying on the "not-for-profit" nature of the operation, see FCC letters to: counsel for Florists' Telegraph Delivery Association, reported in FCC Public Notice 70819, Dec. 6, 1951; and counsel for the Chicago Board of Trade, reported in FCC Public Notice 45049, Jan. 11, 1950. See generally Aeronautical Radio, Inc. v. American Tel. & Tel. Co., 4 F.C.C. 155 (1937); Cooperative Sharing of Operational Fixed Stations, 4 F.C.C. 2d 406 (1966); Multiple Licensing -- Safety and Special Radio Services, 24 F.C.C. 2d 510 (1970).

14/ Preston Trucking Co., 31 F.C.C. 2d 366, 372 (1971) (Docket No. 19309) (emphasis added).

15/ Industrial Radiolocation Serv., 5 F.C.C. 2d 197, 202 (1966) (emphasis added), citing Frontier Broadcasting Co., 24 F.C.C. 251 (1958). See also Western Union Telegraph Co., 14 F.C.C. 1026, 1028, 1042 (absence of a property right in the transmissions confirmed that common carriage was involved), and Moore v. New York Cotton Exchange, 270 U.S. 593, 605 (1926) (emphasizing the passive role of a communications common carrier).

16/ Aeronautical Radio, Inc. v. American Tel. & Tel. Co., 4 F.C.C. 155, 164 (1937) (emphasis added). In this case, the issue was whether Aeronautical Radio, Inc. was a proper "customer" of common carrier circuitry on behalf of the airline industry as distinguished from becoming a common carrier itself.

17/ Final Decision 278. The common carrier route to regulation was the most logical one for the FCC to follow and the one it apparently felt compelled to follow. However, in United States v. Southwestern Cable Co., 392 U.S. 157 (1968), the United States Supreme Court concluded that the broad language of §2(a) of the Communications Act of 1934 bestowed upon the Commission the necessary jurisdiction to regulate interstate wire communications which were neither common carriage nor radio transmission. That decision was in the context of wire community antenna systems, but the rationale of the decision had implications broader than CATV.

18/ Final Decision 278.

19/ See generally Communications Act of 1934 tit. II, 47 U.S.C. §§201-223 (1970); and 47 C.F.R. ch. 1, subch. B (1972). Sections cited in text are to the Act. See also Tentative Decision 300 where the Commission recites its full spectrum of powers over common carriers.

20/ See note 17 supra.

21/ For a discussion of these goals see, e.g., Trebing, Common Carrier Regulation -- The Silent Crisis, 34 Law & Contemp. Prob. 299, 301-02 (1969); and on a broader, but more critical plane, Posner, Natural Monopoly and Its Regulation, 21 Stan. L.Rev. 548 (1969).

22/ Tentative Decision 297-98; Final Decision 270.

23/ Final Decision 278.

24/ President's Task Force on Communications Policy, Final Report 31-35 (1968). See also Dunn, Policy Issues Presented by the Interdependence of Computer and Communications Services, 34 Law & Contemp. Prob. 369, 376-83 (1969) (summarizing the findings and recommendations of the Stanford Research Institute's Report submitted to the FCC in the Computer Inquiry); Irwin; The Computer Utility: Competition or Regulation? 76 Yale L.J. 1299, 1313-17 (1967).

25/ 29 F.C.C.2d 870 (1971) (Docket No. 18920).

26/ In a letter of February 5, 1971 to Timeplex, Inc., cited note 13 supra, Mr. Asher Ende, as Acting Chief of the Common Carrier Bureau, indicated that the furnishing of multiplexing service at a profit might constitute common carriage even though no profit would be sought on the associated interstate lines.

However, on April 28, 1972, the Bureau, through Kelley E. Griffith, Chief of the Domestic Rates Division, advised counsel for RCA that the "Bureau staff is willing to refrain from recommending to the Commission that it exercise jurisdiction" over a network of jointly used lines connected by multiplexers owned by RCA and with respect to which RCA apparently would bear the risk of loss as well as the possibility of profit.

Microwave Communications, Inc. (MCI) opposed this relaxed position of the Bureau staff and on June 13, 1972 sought formal FCC policy resolution of this and a number of related issues by filing a Petition for Rule Making (R.M. No. 1997) which it characterized as, "In the matter of Amendment of Rules and Regulations, Part 21, with respect to alleged sharing of circuits, resale of bulk communications facilities, single user concepts, and other methods which have been employed to secure point-to-point communications at reduced rates by providing common carrier services without undergoing regulation." As this article goes to press, Commission action on the petition is pending

27/ Final Decision 278.

28/ Final Decision 278, 279.

29/ See note 17 supra.

COMPUTERS, COMMUNICATIONS AND DISTRIBUTED HEALTH CARE SYSTEMS

CHAIRMAN:

Martin Elliot Silverstein, M.D.
President, Health Analysis, Inc.
Bethesda, Maryland, U.S.A.

PROGRAM COORDINATOR:

Lynn Hopewell, Vice President
Network Analysis Corporation
Glen Cove, New York, U.S.A.

PAPERS:

Martin Elliot Silverstein, M.D., *"Computers, Communications and Distributed Health Care Systems"*

Dr. Maxine L. Rockoff, Chief, Logistics Branch, Health Care Technology Division, National Center for Health Services Research and Development, Department of Health, Education and Welfare, Rockville, Maryland, U.S.A., *"Health Care Communication Systems"*

SPEAKERS:

Morton E. Ruderman, President, Medical Information Technology, Inc. Cambridge, Massachusetts, U.S.A., *"A Modular Hospital Information Utility"*

Melville H. Hodge, Executive Vice President, Technicon, Medical Information Systems, Corporation, Mountain View, California, U.S.A., *"A Regional Hospital Information Utility"*

Richard K. Tompkins, M.D., Department of Medicine, Dartmouth Medical School, Hanover, New Hampshire, U.S.A., *"The Use of Interactive Clinical Algorithms for the Education and Performance Audit of Physician Assistants"*

Edward F. Vastola, M.D., Department of Neurology, State University of New York, College of Medicine, Brooklyn, New York, U.S.A., *"Effective Use of a Computer for Neurological Patient Screening"*

DISCUSSANTS:

Dr. Alex Reid, Long Range Studies Division, Telecommunications Headquarters, London, England

Gerald S. Cohen, Associate Director, Health Care Technology Division, National Center for Health Services Research and Development, Department of Health, Education and Welfare, Rockville, Maryland, U.S.A.

The assistance of Dr. Bruce D. Waxman and his staff in the initial organization of this session is gratefully acknowledged.

THEME: This session delineates the impact that computers and communications are likely to have on distributed health care systems of the future. This session will also address those highly applied applications of computers and communication to health care delivery. Major attention will be focused on those obstacles which hinder large scale deployment and diffusion of technological systems in the health care field.

COMPUTERS, COMMUNICATIONS, AND DISTRIBUTED HEALTH CARE SYSTEMS

Martin E. Silverstein, M.D.
Health Analysis, Inc., Bethesda, Maryland, U.S.A.

The health service industry presents the largest unexplored frontier for computer technology and computer marketing today.

A science based industry, health services is now the third largest aggregate industry in the United States, encompassing a wide variety of institutions, organizations and personnel who will this year spend 75 billion dollars or 7% of the gross national product. One in twenty members of the American work force is employed in health. This huge industry is still characterized by manual operation, fragmentation, disorganization and inefficiency due to regional duplication.

Despite the obvious need for computer automation, few markets of this size have made so little use of machine based methods. This is a deficiency of both the supplier and the user. Enamored of manual eye-hand expertise and organized into a guild system, the decision-makers of the health industry regard man machine systems as foreign. The suppliers have provided few hardware and software items tailored for the medical man and for the medical manager.

1972 marks, I believe, a year of change of attitude. The managers of the health industry of the nation as a whole have become aware that there is no manual solution for the economic and manpower waste resulting from duplicated facilities dictated by barriers of time and geography. Society represented by the Federal government now views as unconscionable, the denial of optimal services to large segments of our population because of these barriers.

Computer communication systems are accepted, indeed sought after, as the answer; and we are making some progress.

Hospitals have begun to sacrifice their competition for prestige in favor of economy by regional sharing of a single hospital CPU and software. Individual physicians in small group practices appreciate the economy of subscribing to service bureau operated networks for financial procedures. One of my own groups used telephone line communication between our office and a service bureau 200 miles of deserts and mountains away.

But, financial oriented networks are the least sophisticated of the health industry's opportunities for computer communication systems in the health industry. The most exciting trends are those which promise to enhance the ability of the health care profession to heal their patients at a lower cost. In this connection, there is a growing use of medical instruments as computer peripherals in the network. The best known of these is computer analyzed and stored electrocardiographic tracings. Large numbers of physicians in any part of the United States who are not cardiologists and who do not have access to cardiologists can direct their patients' cardiographic voltage patterns into the computer which gives them a statistical analysis and a differential diagnosis. This pattern of machine consultation at a great distance is a trend which will increase as analytical programs for other medical parameters become available. It is only a matter of time before the same services include electroencephalographic analysis.

A great limitation on the majority of the 200,000 practitioners in the United States is the lack of availability of medical center libraries and other sources of consultation. Recently, the National Library of Medicine began a policy of allowing private terminals access to the Medinet bibliographic system. As an avid user, I can testify to the value of this utilitarian system.

The Regional Medical Programs funded by HEW and the National Library of Medicine are the predominant supporters of general medical information utilities. Universities and private groups of practioners are, however, producing some innovative, exciting registries. The Missouri Division of Mental Health designed and operated a massive special records system of psychiatric patients, which provides access to clinical record information, computer aided patient care suggestions (a landmark innovation in computer treatment), and individual case and program evaluation. The San Joaquin Valley Foundation of physicians has reported an index of drug interaction gleened from data reported by California physicians. The payoff has been priceless. For example, one in thirteen prescriptions set up a potential drug reaction, 30% of patients on oral diabetics received an additional, conflicting drug, 53% of post-heart attack patients received prescriptions for anticoagulants along with medication which changes anticoagulation into uncontrolled hemorrhage. On the other hand, 72% of patients receiving drugs for cardiac failure failed to receive replacement of potassium.

At Edinburg, there is a registry of inherited disease so that marriage counselors and obstetricians can have the opportunity of preventing genetic tragedies.

We are now beginning to see the appearance of "decision" utilities. The Vishinsky Institute of Surgery in Moscow is at the center of a network of remote Russion cities that analyzes the clinical findings of local hospitals and doctors and returns a set of potential diagnoses.

Before I introduce your speakers, I would like to direct your attention to the most direct and useful way in which sophisticated computer communication systems can save lives and fill a gap in our health system which has so far eluded solution. Heart attack patients require the institution of monitoring within a "golden" ninty minutes. Hemorrhage and the wounds of violence should be treated within the same period. But our present system of ambulance dispatch and return followed by delay in fully occupied emergency departments, postpones the beginning of definitive treatment for from three to five hours. The needed system is one in which a central computer maintains constant inventory of traffic routes and emergency department status, switches and analyzes incoming physiological signals from portable diagnostic equipment and directs the returning vehicle through the right traffic pattern to the emergency department with the appropriate facilities and the occupancy capability to handle the patient with that physiological array with dispatch. In Boston, there is at Logan Airport an early edition of such a system utilizing telediagnosis without regional traffic or emergency room data. In New York, a prototype system for all of these characteristics was prepared as long ago as 1968, but awaited the development of appropriate portable physiological peripherals; but they are available now. Los Angeles has a similar comprehensive prototype plan. To the best of my knowledge, these desperately needed systems have been user designed. Like so many needed and marketable systems for the health industry, this represents any untouched frontier for the professionals with hardware and software expertise.

HEALTH CARE COMMUNICATION SYSTEMS

Maxine L. Rockoff
Health Services and Mental Health Administration
Rockville, Maryland, U.S.A.

Summary

The broadband communication systems of the future will link people as well as computers, just as the narrowband telephone systems of today are used for person-to-person, person-to-computer, and computer-to computer communications. The change in our communication systems from narrowband to broadband is happening at the very time our Nation's health care system is undergoing fundamental organizational changes. This paper explores the potential impact of these synchronous events.

The Aggregating Health Care System

Demand for Health Care

The Nation's health care system is under mounting pressure to provide more health care of better quality to more people at less cost. This is due in part to an increased public awareness of advances in medical research, and in part to a fundamental change in our society's attitude toward medical care: access to the best medical care, long a privilege of the well-to-do, is increasingly being regarded as a basic human right. Public clamor for health care has made health a political issue, and some form of legislation (such as national health insurance) seems imminent. The dollars thus directed will transform *desire* for health care into *demand* for it.

Several plans have been proposed to meet the growing demand for health care. One approach is to expand our health care resources by training many more doctors and building more hospitals. A second approach is to increase the productivity of our existing resources through their reorganization and rational allocation, and through the introduction of technology and manpower innovations. A third approach is to reduce the demand for the scarce health care resources (doctors and hospitals) required to *cure* illnesses by putting greater emphasis on the *prevention* of illness, or more properly, the maintenance of health. (The economic benefits of preventive maintenance for machines have of course long been recognized!) In this paper we will be concerned with the implications of the second two approaches, namely increasing productivity, and health maintenance.

Increasing Productivity and Maintenance of Health

Various schemes have been proposed for increasing the productivity and efficiency of existing health care personnel and facilities. On the technical side, for example, clinical laboratories have been automated, electrocardiograms are being read by computers and much of the tedious monitoring of physiological signals in intensive care units can be done automatically. On the manpower side, physician-extenders of various kinds have been introduced. A close look at the economics of these productivity-enhancers shows that their principal benefits can be ascribed to the economies of scale which they make possible. To be used profitably, however, they must be used in systems which are large enough to take advantage of these potential economies of scale. If a doctor with two nurse practitioners in his office cares for the same number of patients that he could handle by himself, then the cost per patient seen is obviously greater. Simply stated, if X dollars are invested to increase productivity and if that increase is not consumed, then the unit cost rise. Hence, health care resources must be reorganized to serve large populations in order to utilize the increased productivity made possible by successful manpower and technological innovations. Further, a systems approach in the reorganization is required, with rational allocation of whatever resources are available.

Most proposals for the reorganization of health care resources (facilities and personnel) into efficient delivery systems call for hierarchical systems, with graduated levels of care matched to health care needs. From the providers' point of view, three levels of care have been identified: the primary care level is the most basic level of patient/health care system interaction. General practitioners, family practice physicians, pediatricians, and internists, to whom the patient has traditionally turned when he seeks health care, are considered primary care physicians. Secondary care is provided at the next level, and includes services provided by specialists (such as radiologists and pathologists), usually by referral or request from the primary care physician. Tertiary care encompasses the most highly skilled (and specialized) medical services, including such things as organ transplation.

The elaboration of this hierarchical system in itself draws attention to the related concepts of total or *comprehensive care and continuity of care*. Vertical linkages must be developed to allow patients to transfer smoothly from one level in the system to another. The very expression "health care system" implies concern for the functional integration of the totality of the facilities and resources required for the delivery of comprehensive health care. The concept of a health maintenance organization (HMO) which provides comprehensive care on a pre-paid rather than a fee-for-service basis has been introduced. The HMO is intended to stimulate health care providers to organize themselves efficiently, and to encourage them to provide preventive care.

Communication for Aggregated Health Care Systems

Logistical Problems of Large Health Care Systems

Both the hierarchical organization and pre-payment concepts lead to health care systems designed to serve large numbers of people. Since the number of people requiring a particular level of care decreases as the level increases, hierarchical system calls for large population base in order to be efficient. Pre-payment calls for a large population base in order to spread the risk and expense associated with catastrophic illness. Health care systems designed to provide comprehensive health services to large populations are inherently faced with logistical problems because such systems necessarily have geographically dispersed facilities and personnel; one million patients can't be served in a single building. The twin technologies of communication and transportation have the potential to solve these logistical problems. Both overcome distance, transportation physically and communication functionally. In the remainder of this paper we explore the uses to which communications technology might be put in order to solve some of the pressing logistical problems of health care delivery.

Inter-institutional Broadband Communication

Broadly speaking, communications technology has the potential to:

1. facilitate the functional aggregation of dispersed health care resources;
2. increase the productivity of skilled health professionals and distribute their expertise throughout the health care system;
3. mitigate the depersonalization of health care which might otherwise accompany increased automation, efficiency and size, for example by allowing visual communication between a patient and "his" nurse-practitioner, regardless of the distance between them;
4. improve the quality of care available at any point in the health care system by making expert consultation immediately available; and
5. increase the access to health care of the medically underserved (both rural and urban) by facilitating the creation of decentralized entry points into the health care system, such as neighborhood health clinics.

The National Center for Health Services Research and Development has initiated an exploratory program in which a two-way visual communication capability is being added to a number of existing health care delivery sites now relying on telephone or travel. The seven funded projects include both physician/physician links and physician/physician-extender (e.g. nurse anesthetist, Medex, nurse-practitioner) links. All of the links are broadband, point-to-point; two are microwave, two are Picturephone, two are bi-directional cable, and one is a laser; five are in urban settings, two are in rural areas.

Obviously a broadband communications link has the capacity not only to carry video and voice signals between people, but also to carry data between people and computers and between computers. Therefore, the communications networks for health care systems of the future can reasonably be anticipated to be carrying medical information of all kinds, including visual information, voice information, and computerized information. We can speculate about "universal terminals" that allow person-to-person communication or person-to-computer communication. The switching requirements will undoubtedly include the ability to trade frame rate for resolution so that a given bandwidth can be used at one time for reading a high resolution X-ray, and at another time for diagnosis of a gait disorder.

As the communication networks for health care systems are designed, the transportation alternatives will also have to be considered, as well as the extent to which resources are to be centralized and decentralized. Expensive equipment and highly skilled personnel should undoubtedly be centralized, to achieve economies of scale; primary care facilities should be distributed in order to provide a system which is responsive to human needs.

Cable Television

To this point we have been discussing the inter- and intra-institutional communication requirements for delivery of health care. No discussion of the application of communications technology to health would be complete without mentioning the newly developing two-way cable television systems which can carry as many as forty channels into every home. The home television set will also be a computer terminal, and may even someday be the principal entry point into the health care system, via automated branched questionnaires and computer triage. An HMO's computer-produced "tickler lists" which today result in reminder letters about appointments scheduled and check-ups due, will be able to be sent instead to the family's own "health message box" which the family can dial up. Highly specialized and personalized programs on various health subjects (such as venereal disease, care of newborns, sanitation, and nutrition) will be able to be viewed on request from the subscriber, or as "prescribed" by a member of the health team.

This model for the future is based upon the "Garfield plan" under experimentation at the Kaiser-Permanente Medical Group in Oakland, California. In the Kaiser program, patients entering the system by calling for a doctor's appointment are triaged, via automated questionnaires, laboratory tests, and physical examinations administered by paramedical personnel, into categories of sick, early sick, asymptomatic sick ("worried well"), and well. Patients in the last two categories are given health education and counseling and a "prescription" may result for participation in a group counseling program for nutrition, say, or for viewing a specialized TV program on eye care in a library carrel equipped with TV cassette monitors. Initial successes have been reported in increasing physician throughput and reduced waiting times for physician appointments for sick patients. These gains are ascribed to the rational allocation of health care resources made possible by a "systems approach" to health care delivery.

Sociological Consequences of Communications Technology

The inter-institutional use of communications technology discussed above not only facilitates the re-organization of our health care resources to increase their efficiency and productivity, but also may bring about a change in their traditional behavior pattern which is to provide care in response to a patient-initiated request for services. In addition, cable systems have the potential to change the behavior patterns of health care consumers in even more fundamental ways. Thus, as we introduce these technologies we must be keenly aware that we are doing social experiments, with some potential for doing harm. As we perform intervention experiments that are technologically appealing, we must maintain an awareness of the resulting sociological and psychological effects, some of which may be counter-intuitive. To illustrate, let us consider a few of the issues relating to communication systems for health.

1. Will an increase in alienation and isolation result if we substitute communications for travel? Perhaps a vist to a centralized clinic facility has indirect benefits (social contact with other patients, for example) that would be lost if efficient outreach facilities were available.
2. Will people think that outreach facilities staffed by paramedical personnel represent inferior medical care? Might they resent the doctor who consults via television rather than in person? Might they treat such care lightly, and fail to keep appointments with a physician-extender, or not bother to take the medicine he prescribes?

3. Computerized patient records instantaneously available anywhere in the health care system will greatly increase the quality, continuity, and comprehensiveness of care; what loss of privacy and confidentiality will result, and is the benefit worth the cost?

4. If medical care is easily available at outreach facilities, will not the system soon be clogged with people who don't require the "sick care" services of physicians and hospitals but rather "well care" services such as education and counseling?

5. If the feasibility of remote consultation is established, and if good quality wide-band communication links are available nationally, won't all the "local" resources be bypassed as consultations are sought with the best known and most glamorous medical centers? If consultative back-up is readily available, won't health providers tend to rely too heavily for decisions on the next higher level expert in the system, crippling their own development?

We raise these negative issues not to be pessimistic. Rather, we raise them because our health services delivery system is inextricably interwoven with the entire fabric of our society. Therefore, we can reasonably anticipate that any technological fixes applied to it will have long range social consequences that may be difficult to prognosticate. There are far too many glaring examples of technology applied without adequate concern for the immediate and long-term consequences. To include sociological concerns in the technological planning is not only prudent, but mandatory in this era of public recoil from technology. We do not urge that the introduction of technology be curbed because it may have harmful effects. Rather we urge that the technology be exploited and developed as quickly as possible, but with a somewhat wider awareness, as technologists, of the larger context in which we operate.

[illegible]

1. [illegible]

2. Increase the amount [illegible] health care [illegible]

[illegible] National Center for Health Services Research [illegible]

[illegible]

As the communication networks for health care services are designed, [illegible] in order to provide a system which is responsive to human needs.

Data Transmission

In this paper we have been discussing [illegible] communication requirements [illegible] delivery of health care, [illegible]

[illegible]

Organizational Implications of Communications [illegible]

The [illegible] not only facilitate the organization of the health care [illegible] their efficiency and productivity, but also [illegible] change in their traditional behavior [illegible] In addition, [illegible]

1. [illegible]

2. [illegible]

COMPUTERS AND LIBRARIES OF THE FUTURE

CHAIRMAN:

Dr. Lawrence P. Grayson, Director,
Division of Technology Development,
National Center for Education Technology,
U.S. Office of Education, Washington, D.C., U.S.A.

PROGRAM COORDINATOR:

Charles R. Fisher, Director,
Switching Systems Engineering,
Data Transmission Company,
Vienna, Virginia, U.S.A.

PAPERS:

Dr. Lawrence P. Grayson, *"Computers and Libraries of the Future"*

Dr. Carlos A. Cuadra, Manager, Education and Library Systems Department, System Development Corporation, Santa Monica, California, U.S.A., *"Computer Technology and Libraries of the Future"*

Warren L. Ziegler, Co-Director, Educational Policy Research Center, Syracuse University, Syracuse New York, U.S.A., *"Notes on Social Considerations that May Affect Future Libraries"*

Frederick G. Kilgour, Director, The Ohio College Library Center, Columbus, Ohio, U.S.A. *"Library Economics – The Future"* (abstract only)

SPEAKER:

Dr. John B. Farmakides, Member Atomic Safety and Licensing Board Panel, Washington, D.C., U.S.A., *"Legal Considerations that May Affect Future Libraries"*

DISCUSSANTS:

Frank K. Cylke, Executive Secretary, Federal Library Committee, Library of Congress, Washington, D.C., U.S.A.

Dr. Ruth M. Davis, Director, Center for Computer Sciences and Technology, National Bureau of Standards, Gaithersburg, Maryland, U.S.A.

Dr. Donald L. Katz, Alfred Holmes White University Professor of Chemical Engineering, University of Michigan, Ann Arbor , Michigan, U.S.A.

THEME: This session explores the fundamental questions relating to the application of computers to libraries.

COMPUTERS AND LIBRARIES OF THE FUTURE

Lawrence P. Grayson
U.S. Office of Education
Washington, D.C., U.S.A.

It is evident that man can plan, create and change, in significant ways, the kinds of library services he wishes to have in the future. The fundamental questions relating to the application of computers to libraries, and indeed to society at large, revolve not around what can be done, but rather what should be done.

The fact is that we now possess or soon could develop the means to revolutionize library and information services by achieving virtually any ends that we desire in the storage, access, delivery and display of information. The ends, which can be achieved through the combination of computer and data transmission services, can relate to the more traditional library functions of acquisition, cataloging, serials control, circulation, and bibliographic and reference access, as well as to newer services including computer printed book catalogs, on-line search and retrieval of information both from a listing of the library's holdings and from large data bases such as census information and patent files, selective dissemination of information based on individual user profiles, and much wider access to larger collections of holdings through cooperative arrangements and networks of libraries. One can let his imagination run and conceive of a host of other interesting services that are not now existent, but technically could be available in the future.

This range of possibilities brings to the fore the importance of the goals that are to be achieved. Goals are not solely a consequence of library theory or computer practice, but must be based on the needs, desires and values of the people to be served, and are constrained by economic, social, organizational, legal and political considerations. The goals will help to determine the future.

Even though the range of technical options can be determined, any prediction of the future is uncertain because of the interaction that takes place between a technology and the society in which it acts. In the short-term, computers will be adopted and used because they can improve the effectiveness or efficiency of a given existing service. As computers become widely used, however, they will create new options or make previously unattainable opportunities feasible. Change may then take place in what is done and, more importantly, in why it is done. The technology thus can modify the original goals for which it was adopted, and major changes can occur in the purpose and functions of libraries. It is because of this interaction that prediction is difficult, and it is precisely because a social institution may change that the needs, desires and values of society and those of the users are important to identify and incorporate in any planning for the future.

The papers in this session will be predictive, first by projecting the types of computer-based library services that could be available over the next few decades, and then by discussing the real world factors that will select and shape the services that eventually are developed.

COMPUTER TECHNOLOGY AND LIBRARIES OF THE FUTURE

Carlos A. Cuadra
System Development Corporation
Santa Monica, California

Introduction

Several times in my professional career, I have been asked to take on the role of oracle and make prognostications about the future of libraries. It is a rather dangerous role, in general, because it involves the serious risk of eventually being found wrong. It is especially difficult for me, because I'm better at extrapolating from the present than at taking wild conceptual leaps into the future. It's always easier to see the handwriting on the wall when the letters are three feet high and one is standing very close.

Another challenging aspect of prediction-making is that there are a number of alternative futures open to libraries and other information facilities, and would-be soothsayers have to decide how optimistic a view to take. This is particularly difficult to do at this moment because libraries are at what may be the most critical point in their existence, facing problems that are partly technical but mostly political, social and economic.

It is a well known fact that there are severe pressures on libraries of every type in most parts of the country. Many public libraries are faced with budget cutbacks not only for capital expenditures (such as buildings and bookmobiles) but also for materials and operating expenses. Academic libraries, both ublic and private, are undergoing severe economic ress, and some special libraries are disappearing. short, virtually everyone is feeling the pinch.

There is a sense in which these apparently economic problems are really political and social problems. The American taxpayer ultimately supports almost every library, in one way or another, and there are signs that he is placing less and less value on traditional library services. This is not necessarily the fault of the libraries: some believe that it is largely a function of the increasingly important role of television as a communications and entertainment medium, or of population shifts that have resulted in having fewer people in urban areas who either need or want traditional library services. A sizable--and perhaps growing--segment of the urban population not only believe that traditional library services are not truly relevant to their needs: they distrust the library as an instrument of "the establishment."

The problem is actually worse than the public believes. For the most part, the public and their representatives are not generally aware of the immense cost of duplicative cataloging effort throughout the United States; they are not aware that most of the holdings of most libraries are infrequently used and could be shared much more than they are; and they are not aware of the potential contribution of networks, consortia, and other arrangements for improving services and reducing costs. If they were, the general attitude toward libraries might be even less supportive than it now is.

It is not surprising, in view of these kinds of problems, that some people--including librarians--feel that the public library as it now exists may be on its deathbed. The Federal government is very much aware of this possibility and is struggling to determine what position it should take in the national interest: to continue its present level of support to public libraries, to reduce or cut off this support, or to greatly increase it, perhaps to the extent of having a completely Federally supported public library system. The choice depends on the kind of alternative future one can see for public libraries and on the extent to which it is necessary for the Federal government to intervene in bringing one or the other of the future possibilities about.

This conference and this session are, of course, concerned with computer technology, and in the remainder of this paper, I expect to concentrate on this topic. But it is important to recognize at the outset that the effective application of computers and other technology to libraries will depend on the recognition and solution of several major economic, social, and political problems that are very much intertwined with technology.

Computers in the Total Context of Library Technology

As all of you know, computers are not the only kind of technology that is important to libraries. Library technology includes a lot of conventional and non-conventional hardware and gadgets: computers, terminals, communications gear, microform copiers and readers, file cabinets, mechanical conveyors, and so on. It also includes a host of techniques related to the processing of textual information, from the time-tested techniques now in use in most library and information facilities to more exotic techniques such as format recognition, automated indexing, automated classification, machine translation, and question-answering systems.

What kinds of new technologies do libraries need, and which ones will prove most useful? That seems like an easy question, but it is not. The answer depends on what the library itself is likely to become. If libraries of the future perform only the kinds of functions that most of them now perform, they may have little use for technology, because their place in tomorrow's information-oriented society may have been taken by other types of information service facilities, such as information clearinghouses, information analysis centers, and community learning centers.

But many libraries are beginning to think of themselves in new terms: not as passively-oriented repositories of books but as agressive educational and recreational forces, reaching out into the community to perform a wide range of services, centered around information but not strictly limited to information. If more libraries begin to operate in this new mode--assuming that they can afford to do so--they will need to make extensive use of technology, including computer technology.

Assuming an optimistic point of view about the future of libraries, one can foresee an important role for several kinds of technology. Libraries have long used microfilm, for example, and a number of government agencies and even some scientific journals regularly distribute information in microfilm. We can certainly expect this technology to expand in sophistication and use. Up till the present time, library microform collections have tended to be quite small and the equipment to read microform images has been relatively poor, but in the future, we can certainly expect to see the establishment of large-sized microfiche collections and significant improvements in microform technology including stable color microfiche, with sophisticated microform-handling equipment, and high quality readers.

The marriage between computers and microform technology has important implications for libraries. We already have a technology for direct information transfer between microimage and computer systems. If libraries kept their master bibliographic files in machine-readable form, they could prepare updated microfiche versions of their catalog directly under computer control. The combination of microform and computer technology also has important potential for nationwide on-line retrieval services, such as those associated with the ERIC and NTIS data bases. Although it is technically feasible to put document abstracts or even complete documents in digital form, for on-line search, it may be economically more attractive to use on-line computer search techniques to identify the documents of interest and to use locally or regionally held microfiche to obtain the actual hard copy of abstracts or full text.

Another major area of technology that is very important for libraries is communications. Today, the U.S. mail is still the most practical means for the transmission of textual materials, and it is likely to remain so for many years to come. But electronic transmission of materials from library to library has been tried, in a limited way, in a number of places, and we can certainly expect to see the volume of such transmission grow in future library systems.

The increased use will stem in part from improvements in communications technology and in part from the lowering of communications rates. A year ago, if a library in Washington, D.C., wanted to search a file in California from an on-line terminal, even for a directly dialed, station-to-station call, the telephone line time would cost about $30 per hour of connect time. This year, it can do so for well under half that cost. In another two years, when the microwave transmission companies recently unleashed by the FCC have their equipment in place, we can expect the cost to drop much lower. I don't know when U.S. libraries can count on the availability of satellite communications service, but we will almost certainly see it within the next decade.

One final type of technology that deserves mention, as part of the total context of potential library tools, is the technology of moving materials. Most of us who spend a great deal of time in association with computers tend to think in terms of digital data and, if we forget to count the number of bits involved in the collections of even rather small libraries, we can delude ourselves about the ease of transmitting the contents of, say, 1,000 books from a library on the East Coast to one on the West Coast.

Those who recognize the enormous information transfer problems associated with library-type materials have begun to think in terms of regional libraries whose resources could quickly be tapped by individual libraries in their geographical area. Although this kind of system would be costly, because the collection in each regional library would almost certainly have to duplicate the collection in the other regional libraries, a regional library system could reduce the delays involved in moving materials from, say, a single national lending library to the point of request.

A recent report from the RAND Corporation suggests that technology may eventually provide a different solution. The head of the physical sciences department at RAND has indicated that it is possible to develop a rapid transit system to move people from Los Angeles to New York in half an hour for a $50 fare. The system he has devised involves a tunnel more or less along the present routes of U.S. Highways 66 and 30. The tunnel would contain several large tubes for the east-west travel of trains floating on magnetic fields, moving at top speeds of 10,000 miles per hour. The trains would actually be single cars and would be big enough to carry both passengers and freight, including large containers and automobiles. These could rapidly be delivered to intermediate stops along the main line and to other cities, through feeder lines. The tunnels themselves would carry pneumatic tubes, which could carry mail, and laser beams for the instantaneous transmission of messages.

The 10,000 m.p.h. tunnel train may not be with us very soon, because tunneling across the United States would be a very expensive undertaking. However, it is important to recognize, in looking ahead toward the year 2000, that there may be alternatives to physically transporting materials by airplanes, trucks, and surface trains, on the one hand, and digital transmission, on the other.

Current Uses of Computers in Libraries

Most libraries are not presently using computers in their operations. In fact, I would estimate that not more than 10 percent of this country's more than 25,000 libraries are making use of computers. This is not from lack of interest or awareness of the potential of computers, but, rather, from the lack of resources in any given library to move from manual operations to computer-aided operations.

Most of the use of computers has been in large academic libraries, some of which have had the benefit of extensive Federal support. These and some of the public and special libraries have used computers in three major ways:

(1) To support one of several major functions such as acquisitions, cataloging, new materials, circulation control, management of serial publications, and reference. These uses of the computer are technically quite successful but rather questionable from the economic standpoint.

(2) To support all of these major functions in the context of a "total" system for a single library. Very few libraries have a total system in operation, though many have accepted the concept and are working toward it, component by component. Total systems have more to offer a library, from the standpoint of cost-effectiveness, because a given keyboarding of library data serves a variety of useful purposes.

(3) To support the operations of a group or system of libraries that have banded together to develop--or at least use--computer resources that none of the libraries could afford individually. There are only a very few such systems in operation today, and they support individual library functions, rather than operating as total systems.

Most of the applications of computers in libraries have tended to be for the functions in the "back room"--acquisitions, cataloging, and serials management, together with circulation control. There has been relatively little emphasis on the reference and retrieval function, even though for many libraries that is the primary interface between the library and the customer. The major exception to this emphasis has been in medical libraries, many of which have been enjoying on-line retrieval access to large medical literature files for several years.

Future Uses of Computer-Related Technology in Libraries

Computers, along with other related technology, will play a role in four aspects of future libraries: (a) their holdings, (b) their techniques of pointing to the holdings, (c) the transferring of information to the user, and (d) their internal processing.

Holdings

It is generally accepted that there will be major changes in the variety of holdings of libraries, with a shift away from books and other printed materials toward other media. I have already mentioned microforms. Several companies now offer microfiche libraries, and some of the recently introduced viewers to read these fiche are very impressive. We can also expect to see holdings in the form of digital and analog data bases. Good examples are the large data bases already available on census and voting information. The University of California at Los Angeles has had a project for some years to provide reference service for this type of data, and the availability of computers and rapid communications now makes it reasonable to think of such information as properly being general library information in the next decade or so.

In the next decade we can also expect modern libraries to have much larger holdings of quite different materials: common materials such as films and phonograph records, and new media such as videotape cassettes and holograms. There are already several systems, in various states of marketability, that record TV programs on cartridges or cassettes and allow playback on one's home TV set. These media will be increasingly important in our society because they carry certain types of information that the printed word cannot carry well, if at all.

I do not mean to imply that books are going to disappear from libraries. Far from it; they are exceptionally efficient and convenient devices for holding sizable stores of easily accessible information. But the bookstores, drugstores, supermarkets, and airport magazine counters have helped to erode the book-providing function of libraries. Since videotapes and other new media are not likely to be cheap enough for mass sale for some time, they may become the major attractions of the library's holdings, particularly if, through the advent of cable TV, library users can enjoy the use of some of these holdings without even going to the library facility itself. Who knows?--ten years from now one may even find audio- or videotape recordings of conferences such as this one available through the library.

Techniques for Pointing to Information

One of the functions that technology is changing and will continue to change is the way in which the library points to its holdings. The major tool currently used for this purpose in most libraries is the card catalog, which provides several routes of access to the library's holdings. Computers have been used for several years, in a few libraries, to produce catalogs in book form so that multiple copies of the library catalog can be available in several locations within a library or other places, such as the offices of individual scientists or groups of scientists. The next stage may be to take advantage of photocomposition--computer controlled preparation of reproduction masters with electro-optical techniques, to produce high-quality copies of catalogs, either in book form or on microfiche, through the COM (computer output on microfilm) technique. Either way, more people will be able to search the library's holdings without having to travel to the library to use the card catalog.

Another, more interesting use of computers to point to library information, is to provide on-line search from remote terminals. Not many libraries are able to offer this kind of service yet, but the few services that exist, in special libraries, are highly successful. For example, the National Library of Medicine has established two on-line search services serving more than 80 terminals, for about 10 hours a day, in medical settings throughout the United States. From these terminals, one can search files containing over 650,000 bibliographic citations to recent medical literature. Any good on-line system provides tutorials and diagnostics to its users, to help them use it better. The good library of the future will provide its users with some amount of computer-aided instruction, to help them learn how to use the library itself better.

Within another decade, I expect to see at least 5,000 libraries and information centers using on-line searches of a variety of data bases, such as MARC, ERIC, Government Reports Announcements, and Chemical Abstracts Condensates. At least 1,000 of these libraries will be libraries operated by the Federal government. These libraries tend to be small and quite limited in staff size and budget and lacking in computer support, but these very limitations, together with their role in support of our institutions of government, make them natural candidates for on-line search services operated from a Government-sponsored computer-based library service bureau.

There is another technique for pointing to information in which computers have long played an important role: selective dissemination of information. This type of computer application began in the late 1950s and today there are probably scores of computer-based SDI programs in operation. Most of them are in information centers concerned primarily with scientific and technical information, but there are a few operating in libraries, including one in the National Science Library of Canada (NSL). NSL regularly acquires new-literature tapes from a number of private and public agencies, compares the descriptive data on the literature with hundreds of stored individual user profiles, and sends out notices to these users located throughout Canada.

Most libraries in the U.S. have not considered this kind of service to be a priority target, even though for a group of libraries in a city, region, or state, it could be very cost-effective. There is a likelihood, however, that the Federal government will take some initiative in operating some kind of centralized SDI services on behalf of its more than 2,600 libraries. That would greatly improve the ability of these libraries to point their customers to new information of particular interest to them.

From SDI announcements on individual documents, it is only a small step to the semi-automatic production of personalized newsletters or collections of abstracts, tuned to the interests of very small groups of researchers, and perhaps only one person. Producing something like a personalized book on demand won't be inexpensive for many years, but that is a direction in which information services seem to be moving.

Both the personalized dissemination services and the on-line retrieval services are going to help foster the development of "information brokerage" enterprises. The average library or information center user will not wish to subscribe to one SDI service covering one type of literature and another service covering a different type. He will want, instead, to tell someone nearby--an information broker--what he needs and let the information broker serve as a middleman in the necessary transactions.

Transfer of Information

Libraries help to transfer information in three ways. They operate a circulation system, which permits materials to be taken outside of the library and later returned, and they permit use of reference materials that cannot be taken outside of the library. The third role, which is presently of minor importance, is to distribute--or facilitate the distribution of--materials that the user can keep. This role is evident mainly from the copying machines now widely available.

I expect computers to be used increasingly often for both the circulation function and the reference function. Computers have been used for keeping circulation records for many years and we are likely to see more automated circulation systems as time goes on. Some library users, particularly among the so-called "disadvantaged" groups, would like to see more materials--for example, paperbacks--simply given away by libraries. While bookstores might very well be concerned about this as a national library policy, it is a fact that under some circumstances the cost of a paperback is less than the cost of operating the circulation control mechanism. Perhaps libraries will need to limit their distribution function and give away not the printed materials themselves but microfiche copies. That solution raises some social issues, because the people most in need of free materials would be least likely to have a microfiche reader in the home. At any rate, libraries will still need to lend--rather than give away--the more expensive materials, such as audio and video cassettes, so there is no way in which the circulation function as such will disappear. If anything, it will expand, placing new demands on the computer system that supports it.

Earlier, I mentioned the use of on-line systems for pointing to information. Obviously, these systems can support the library's reference function by displaying abstracts, full text, graphic materials, and other holdings of the library. The dormitory rooms in tomorrow's modern universities may well include a terminal or two, to provide the student with access to many different kinds of materials formerly available only at the library itself or at the media center. Even more exciting, if cable TV continues to develop rapidly and the "wired nation" becomes a reality, we can envision almost every home with the same kind of access to a substantial fraction of the nation's knowledge store. No well equipped household will be without two-way cable TV service providing reference materials, television instruction for students unable to attend school, and many kinds of community information, such as bus schedules and doctors' names, rates, and available appointment hours. From their own home TV set--probably one of two or three in the household--users will be able to examine the catalog of the library and select the material they would like to see, either on-line or delivered to their home the following day through a simple ordering process.

What will become of the library facilities themselves when users will have access to many of its holdings directly from their homes? One of the persistent findings in studies of information use is that the use of one information channel increases, rather than decreases use of other information channels. My guess, therefore, is that increased remote use of the library will breed new respect for the library as an institution and lead to increased use in person, particularly if the facility is well equipped with the newer media and equipment to use them, if it is comfortable, inviting, and stimulating, and if it has a program that makes it an active part of the community life. I include this last item not for social reasons--although they are important--but for public relations reasons. The library needs to be visible in order to be used.

Internal Processing

The three aspects of library operation that I have been discussing--the nature of their holdings, their techniques of pointing to the holdings, and their means of transferring information--are primarily the aspects that relate directly to the user. But behind these aspects are several kinds of activities necessary to identify, acquire, and catalog materials and to make them part of the holdings. How will these internal processing activities be carried out in future libraries?

I foresee two primary differences between today's internal processing and those of 10 to 20 years from now. First of all, there will be a great deal more use of computers then than now. Most of the current uses of computers are in the internal processing areas--acquisitions, cataloging, and serials management--and because of the experience that has already been gained in these uses, these uses will continue and expand.

The second difference is that there will be much greater centralization of processing effort than at present. While the movement toward automation has been one of the most publicized aspects of libraries for the past decade, the movement toward cooperative library processing operations has really been the most dramatic. I happened to be involved in a recent study of academic library consortia. We discovered that of the 125 library consortia that met the criteria for inclusion in our study, 2 percent had been established during the 1930's; 3 percent had been established during the 1940's; 4 percent had been established during the 1950's; and about 90 percent had been established in the 1960's, most of them between 1965 and 1970.

There can be no doubt that libraries have seen the handwriting on the wall: in order to maintain and improve their services to users, they must extend the areas of cooperation with other libraries to both share resources and minimize duplicative operations. The kind of centralized cataloging involved in the Library of Congress' MARC service will, I believe, be the standard mode of operation, and this centralized cataloging will be speeded up and otherwise improved through the assistance of publishers, who will themselves be producing literature in machine-readable form and routinely arranging for cataloging prior to publication.

I am not sure whether the 1970's will see the continued growth of consortia. Some of the consortia that exist are carrying out activities that, although necessary and very important, might not need to be carried out at the local or regional level if there were a national library system in this country. The establishment of the National Commission on Libraries and Information Science represents a significant step in exploring the potential of the idea of national-level library operations. I believe that there can be a more rationally articulated library system in the United States by 1980, but it will take great planning effort and, more particularly, significant changes in both the national and grass roots leadership in the library/information science profession. I think that the field will begin to move about five years after most of the schools of library and information science in the country introduce courses in cost-effectiveness analysis and network and consortium planning and make them mandatory.

Conclusion

I have touched on many topics in this paper. It would have been tempting to stick more closely to the role of technology in libraries of the future. But technology is merely a tool that may or may not help to solve a problem or achieve a goal, depending on what the goal is. Library and information science are woefully short of explicit goals, and I often think we are seduced by visions of what can be done, rather than forcing ourselves to define what ought to be done. A few months ago, an analysis was performed to identify, in the library and information science literature, any clear and specific statements relating to the library and information needs of the nation. There was virtually nothing in the literature, in spite of the fact that knowledge of the needs should be a prerequisite to planning future library services.

The technologically-flavored view that I have given of future libraries is only one of many possible ones, and there is a very large element of choice open in the directions that libraries may take. It is not necessary or inevitable that libraries expand their holdings to include digital information or videotapes. It is not necessary that libraries become either users or providers of remote-access, on-line services. It is not necessary that they begin to operate as knowledge resource centers or community information centers. But these functions are certainly going to take place in tomorrow's society, and the agency that carries out these functions will certainly occupy the central role in the knowledge business. Present-day libraries certainly have the information-handling background and some of the information resources to move toward this central role. Whether they will actually do so effectively is a prediction that I will leave for others to make.

NOTES ON SOCIAL CONSIDERATIONS THAT MAY AFFECT FUTURE LIBRARIES

Warren L. Ziegler
Syracuse University Research Corporation
Syracuse, New York, U.S.A.

INTRODUCTION: METHODS

The task of identifying social considerations affecting the future of libraries suggests a refined capacity for forecasting in the "soft" domain of social behavior and institutions. Such a capacity--i.e., forecasting techniques and methodologies--is in short supply, not widely distributed among librarians, social scientists or educators. Moreover, no techniques now available enable us to make scientific claims about relevant social conditions with the kind of certitude to which we are accustomed in the "hard" sciences, by which I mean to also exclude economics and demography, although these two fields have their predictive, or at least extrapolative, models best in hand among the social sciences.

Thus, our task is less to set forth hard forecasts and more to define the relevant domains of social conditions--some, but not all--about which we can try to ask the right questions. A hundred forecasts--about demographic patterns, levels and distribution of educational attainments, levels and distribution of income, character of future transportation systems, spatial and socio-ecological characteristics of future metropolitan areas, etc.--may be less useful, as a beginning, than some good questions.

Finally, let us not in this brief inquiry continue with the mythology which pretends objectivity or value-neutrality for any claims or propositions about the social future. Discussion about the future of man is an exercise in hope--or perhaps despair. Whatever hard data from the past we may use as a base for projections, whatever theoretical propositions we may employ to lend reason to our conjectures, still the future is too important to leave solely to futurologists, experts, politicians or professionals. Librarians are human beings first--so are we all. Our primary task, in investigating future alternatives, is to discover, or invent, those which appear more conducive to a last gasp effort to redefine and reclaim our humanity. If libraries, or any other mode for the organization and dissemination of information, do not have some offer to make to that need, speculations about the future of libraries and their social domain become mere exercises in technical intellectuality of which I want no part--and to which I can make no contribution.

SOME FUTURE ISSUES FOR POST-INDUSTRIAL AMERICA

The following list of emerging post-industrial themes and problems is by no means exhaustive or unbiased. Still, it is a first cut at outlining a future agenda which all major institutions in our society must address, particularly those whose functions include determining the uses of, the control over, and the access to knowledge and information about these themes and problems. We begin, then, with a portrayal of social conditions of such pervasive dimension that to avoid their consideration is tantamount to suggesting that the institution of the library will have relevance to nothing important to mankind.

(1) One issue will be how work will be redefined. I see a continuing erosion of the Protestant work ethic to the extent to which it is conventionally identified with job and occupational career. Work is a profoundly human activity whose original meaning included such notions as discipline, craft, and a loving fashioning of the environment which became an extension of self. Alienation from work means that the job no longer facilitates this human activity. Indeed, the formal, economic organization of industrial society has clearly eliminated most opportunities for participation in this original sense of work. This issue is directly related to the continuing education of all Americans. It forces us to question how and whether libraries can assist their audience to participate actively in the redefinition of what constitutes both work and economic security.

(2) A second issue will be how the criteria for social selection will be redefined. I see a breakdown of the current structure and meaning of educational credentials. It is no longer certain that their acquisition will continue to certify the selection and distribution of youth into socially-acceptable roles. It is difficult to forecast what alternative devices for social selection will emerge until we are much clearer about the major themes of future society, for these themes define the content of social roles and their status hierarchy. If economic growth as conventionally defined is superseded by a new relationship between man and his natural resources, which is non-manipulatory and non-exploitative, then we might expect rather radical redefinitions of both the distributive structure and substantive standards of social selection criteria. This issue

is related to the continuing question of the relationship between the provision of information and knowledge and the formal credentials which certify access to those sources.

(3) A third issue will be engendered by the consequences of an increasing gap between affluent populations and impoverished groups, both on the international and domestic scene, in which the latter are unable to achieve a status comparable to the Western middle-class, and increasingly reject production-consumption indicators of self-worth. One consequence will be heightened socio-economic conflict in both the transnational and national context. A second consequence for affluent societies may be the increasing disassociation of job from social provisions for minimal standards of economic security.

(4) Another set of very complex issues arises from the continuing erosion of traditional social meanings ascribed to the primary institutions of family, community, and religion. These issues can even now be appreciated, if not fully understood, by recognizing the emergence of alternative life-experiences through which individuals locate their place in society--i.e., establish the basis for psychologically satisfactory definitions of self-worth, social competence and human effectiveness. I believe we see a variety of experimental behavior in some youth in developing alternative life-experiences, but I claim no prescience about how this present behavior will effect future society.

(5) The rise of technicist, cybernetic social control and governance institutions in direct conflict with more traditional notions of individual autonomy and participatory democracy raise crucial issues about the future of the public and the republic. These are political issues in the broadest sense, and force us to ask whether and under what conditions libraries will serve the open information needs of a revitalized public--or, help in its demise.

(6) The continuing erosion of traditional face-to-face human relationships and their replacement by electronic communications and information management devices which intensify the man-machine symbiosis characteristic of industrial societies. Whether these devices will lead to increased feelings of alienation as now seen predominant in Western society or will lead to a new sense of self-control and involvement is not known. To some extent, this depends on whether new forms of electronic communication technology--such as cable-TV, holographic image-projection, videotape cassettes, regional and national computerized information storage and retrieval systems, and home/business/school-located audio-visual consoles--are utilized to promote self-directed human engagement in economic, social, political and educational change.

(7) The alternative uses and meanings of affluence. The issue here has something to do with the erosion of personal satisfactions coincident with the increasing material abundance of goods and services. The quest for material affluence has clearly been accompanied, historically, with a widespread sense of social alienation, and personal discontent and anxiety. Affluence may become redefined as the opportunity for personal rediscovery and renewal, for developing a sense of personal effectiveness and authenticity outside of the framework of "getting and spending." Still, there would remain a critical issue about the basis for distributing these opportunities among different socio-economic groups . . . perhaps a question of the redistribution of work and leisure.

(8) The emergence of a science and technology of human behavior based upon exponentially increased power for genetic manipulation, for control of the autonomic nervous system, and physical and chemical intervention into the historically sacrosanct domain of "mind," "will," "intentionality," and "personality."

(9) The redefinition of minimum levels and contents of functional literacy in a society in which the capacity to use knowledge and manage information (including its discard, rejection, production and retrieval) become the major requirements for cognitive skill training.

THE CENTRAL IMPLICATION

Let us assume, for the moment, that this set of post-industrial themes and issues portrays a future which should concern people in all of the multiplicity of institutions and roles of a contemporary society in transition. Still, from that general human concern, what more direct consequences emerge for the library function?

The key problem, it seems to me, is the question of access to those contents and modes of knowledge and information which might enable the population to develop more effective means for resolving the social, economic, political and human problems suggested above.

Access is the idea which confronts prevailing structures of the production, control and use of knowledge. Access, like control, is always a legal--indeed, a constitutional--issue in our society, but it is more than that. It is a complicated social issue, because habits of mind and social custom lead to taking advantage of expanding opportunities for continuing engagement with knowledge--or withdrawing from those opportunities. This is a traditional challenge to libraries, as it has been to education, but the challenge magnifies in the future.

Access is not, however, a passive provision of opportunity. It must also become an active function promoted by libraries which consider it their responsibility to effectively participate in the on-going education of the citizenry. That educative effort cannot be--and never has been--reposited solely in our schools and colleges. Indeed, libraries have played a sometimes emancipated and always important role in the continuing learning of citizens.

Given the kinds of issues I believe confront our society, that role must expand--or diminish. Neutrality in the face of to what uses we shall put our knowledge in our tasks of self-emancipation and social reconstruction is no longer possible, if it ever was. Certainly, for the future, neutrality will be immoral. We seek, then, for an emerging understanding of how and in what ways libraries may contribute to a more effective participation by the members of this society in determining the future social conditions in which we and our children shall live out our lives. The how and the what are beyond the task of this brief inquiry and exhortation. The why, I believe, needs no further argument.

LIBRARY ECONOMICS—THE FUTURE

Frederick G. Kilgour
Ohio College Library Center
Columbus, Ohio, U.S.A.

Abstract

Unit-expenditure growth figures for libraries range from 2 to 9 percent depending on the unit and type of library selected. In general, growth rates in unit operating costs are about double analogous rates in the economy as a whole. This extravagant rate of increase stems from a lack of a library technology that continuously increases productivity per manhour. In manufacturing and agriculture the recent rate of increase in productivity has been 3 percent per year which, together with an inflationary rise in prices of 2.5 percent, produces a 5.5 percent yearly increase in wages. A 5.5 percent increase in library salaries, in the absence of an increase in productivity, forces library costs to climb at a 3 percent compounded rate relative to costs in the general economy.

This acceleration in the rise of library costs has generated a crisis in funding which can be met either by funding agencies accepting the inevitability of the increasing rate or by utilizing library technology that can continuously increase productivity.

Libraries have taken advantage of some aspects of economies of scale and of computerization, but as yet there has been no regular increase in productivity within libraries. For the future, however, it is the computer that presents the greatest source of hope for an increase in productivity per manhour. This can come about both by enhancing economies of scale and by reducing the need for large staffs. The computer will provide libraries with a productive technology that can decelerate the rate of increase of library costs and bring those costs into line with costs in the economy as a whole.

o

Conference Committees

EXECUTIVE COMMITTEE

FOUNDER
Dr. [illegible]
[illegible]
[illegible], Connecticut, U.S.A.

GENERAL CONFERENCE CHAIRMAN
Dr. Maurice Karnaugh
IBM Thomas J. Watson Research Center
Yorktown Heights, New York, U.S.A.

CHAIRMAN [illegible]
Dr. [illegible]
[illegible]
[illegible]

EXECUTIVE COMMITTEE CHAIRMAN
Philip H. Enslow, Lieutenant Colonel, U.S. Army
Office of Telecommunications Policy
Executive Office of the President
Washington, D.C., U.S.A.

SECRETARY
Floyd [illegible], Director, [illegible]
Systems Evaluation and [illegible]
Computer Sciences Corporation
Falls Church, Virginia, U.S.A.

LIBRARY ECONOMICS—THE FUTURE

Frederick G. Kilgour
Ohio College Library Center
Columbus, Ohio, U.S.A.

ABSTRACT

[illegible] rates [illegible] for libraries range from 2 to 3 percent depending on the cost and type of library activity. In general, growth rates in unit operating costs are [illegible] than analogous rates in the economy as a whole. This unfavorable rate of increase arises from a lack of a library technology that can enhance increase productivity per worker. In manu[illegible] and agriculture, the annual rate of in[illegible] in productivity has been 3 percent per [illegible], together with an inflationary rate [illegible] of 4.5 percent, produces a 1.5 percent [illegible] increase in wages. A 4.5 percent increase [illegible] library salaries, in the absence of a [illegible] in productivity, causes library costs to rise at a 3 percent compounded rate relative to costs in the general economy.

This acceleration in the rise of library costs has generated a crisis in funding which can be met either by funding agencies accepting the inevitability of the increasing rate or by [illegible] library technology that can continuously increase productivity.

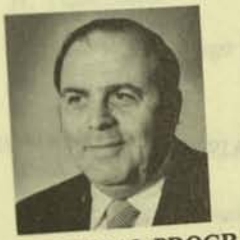
Libraries have taken advantage of some economies of scale and of specialization, but as yet there has been no regular increase in productivity within libraries. For the future, however, it is the computer that presents the greatest source of hope for an increase in productivity per employee. This new tool which works by extending economies of scale and by reducing the need for large staffs. The computer will provide libraries with a productive technology that can decelerate the rate of increase of library costs and bring these costs into line with costs in the economy as a whole.

Conference Committees

EXECUTIVE COMMITTEE

FOUNDER
Dr. Reg A. Kaenel
AMF Incorporated
Stamford, Connecticut, U.S.A.

GENERAL CONFERENCE CHAIRMAN
Dr. Maurice Karnaugh
IBM Thomas J. Watson Research Center
Yorktown Heights, New York, U.S.A.



GENERAL PROGRAM CHAIRMAN
Dr. Stanley Winkler, Senior Scientist
IBM Systems Development Division
Gaithersburg, Maryland, U.S.A.

EXECUTIVE COMMITTEE CHAIRMAN
Philip H. Enslow, Lieutenant Colonel, U.S. Army
Office of Telecommunications Policy
Executive Office of the President,
Washington, D.C., U.S.A.

SECRETARY
Floyd H. Jean, Director,
Systems Evaluation and Simulation Operation,
Computer Sciences Corporation
Falls Church, Virginia, U.S.A.

Executive Committee (continued)

TREASURER & COMPUTER SOCIETY LIAISON
Harry Hayman
Silver Spring, Maryland, U.S.A.

SPECIAL PROJECT
Dr. Lawrence G. Roberts, Director for Information
Processing Techniques, ARPA,
Department of Defense,
Arlington, Virginia, U.S.A.

COMMUNICATIONS SOCIETY LIAISON
Charles Fisher, Director,
Switching Systems Engineering,
Data Transmission Company,
Vienna, Virginia, U.S.A.

LOCAL ARRANGEMENTS & REGISTRATIONS
Col. William T. Lincoln
Defense Communications Agency
Systems Engineering Facility
Reston, Virginia, U.S.A.

ACM LIAISON
Dr. Peter E. Jackson
Bell Telephone Laboratories
Holmdel, New Jersey, U.S.A.

PRINTING & PUBLICATIONS
Maurice Raffensperger, Technical Director
Defense Communications Agency
Systems Engineering Facility
Reston, Virginia, U.S.A.

EXECUTIVE ADVISORS:

Dr. Lewis Billig, Technical Director
Communications Division, The Mitre Corporation
Bedford, Massachusetts, U.S.A.

Fred Israel, Counsel
Israel and Maness
Washington, D.C., U.S.A.

Prof. Manley R. Irwin
Whittemore School of Business and Economics
University of New Hampshire, Durham, New Hampshire

Philip J. Musgrave, Public Relations
Vienna, Virginia, U.S.A.

Sidney Metzger, Assistant Vice President and
Chief Engineer
Communications Satellite Corporation
Washington, D.C., U.S.A.

Prof. Anthony G. Oettinger
The Aiken Computation Laboratory
Harvard University, Massachusetts, U.S.A.

Dr. Addison Richmond
Department of State
Washington, D.C., U.S.A.

REGIONAL DIRECTORS

CANADA:
Gordon B. Thompson
Manager, Communications Studies
Bell-Northern Research
Ottawa, Ontario, Canada

CENTRAL EUROPE:
Dr. M. Clayton Andrews
Director of Data Communications
IBM World Trade Corporation
IBM Research Laboratory,
Zurich, Switzerland

NORTHERN EUROPE:
Dr. Edward G. Aghib
ITT Europe Inc.
Brussels, Belgium

FRANCE:
Dr. A. Profit
Centre National d' Étude des Télécommunication
Issy-Les-Moulineaux, France

GREAT BRITAIN:
Philip Hughes (1972-)
Managing Director
Logica, Ltd.
London, England

F. M. Trapnell, Jr. (1971-1972)
Director of Systems
Wavetec Data Communications
San Diego, California, U.S.A.
(formerly with T.C. Hudson Associates Ltd., London, England)

INDIA:
Prof. M. G. K. Menon
Secretary, Department of Electronics
New Delhi, India

Prof. K. Venkitaraman
Chairman, Electronics and Telecommunication
Engineering Division
The Institution of Engineers (India)
Sivaram Nivas, Kevala, India

JAPAN:
H. Yoneguchi
Vice President
Nippon Univac Sogo Kenkyusho, Inc.
Tokyo, Japan

ROMANIA:
Prof. Al. Spataru
President
Romanian Commission for Space Activity
Bucarest, Romania

SOUTH AMERICA:
Boris I. Plazas
Chief, Departmento Nacional de Planeacion
División de Commuicaciones
Bogota, Columbia

SWITZERLAND:
Dr. Hans Diggelmann
Hasler A. G., Bern
Bern, Switzerland

U.S.A.:
Dr. Reg A. Kaenel, Chairman
AMF Incorporated
Stamford, Connecticut, U.S.A.

USSR:
Dr. S. I. Samoylenko
Vice-Chairman
Council for Cybernetics, USSR Academy
of Sciences
Moscow, USSR

LIAISON REPRESENTATIVES

American Bar Association:
Roy N. Freed
Widett and Widett
Boston, Massachusetts, U.S.A.

Association for Computing Machinery:
Dr. George Capsis
Singer Electronics Corporation
New York, New York, U.S.A.

American Federation of Information Processing Societies:
Dr. Richard I. Tanaka
California Computer Products, Inc.
Anaheim, California, U.S.A.

Association of Data Processing Service Organizations:
Robert F. Guise, Jr.
Comshare, Inc.
Ann Arbor, Michigan, U.S.A.

Business Equipment Manufacturers Association:
Charles A. Philips
Vice President,
Business Equipment Manufact. Association
Washington, D.C., U.S.A.

Comité Consultatif International des Téléphonique et Télégraphique:
Vern C. MacDonald
Department of Communication
Ottawa, Ontario, Canada

Data Processing Management Association:
John W. Larmer
Tax Computer Association
Falls Church, Virginia, U.S.A.

Federal Communications Bar Association:
Edward J. Grenier, Jr.
Sutherland, Asbil and Brennan
Washington, D.C., U.S.A.

Institute of Electrical and Electronics Engineers:
Dr. Richard Emberson
IEEE
New York, New York, U.S.A.

International Federation for Information Processing:
Dr. Heinz Zemanek
Laboratory Manager
IBM Austria
Vienna, Austria

National Academy of Engineering:
Dr. James Mulligan
Executive Secretary,
National Academy of Engineering
Washington, D.C., U.S.A.

National Academy of Sciences:
Dr. Warren House
National Academy of Sciences
Washington, D.C. 20418

National Institutes of Health:
Dr. Arnold W. Pratt
Director Div. Computing Research and Technology
National Institute of Health
Bethesda, Maryland, U.S.A.

Society for the Application of Learning Technology:
Raymond G. Fox
IBM Federal Systems Division
Gaithersburg, Maryland, U.S.A.

United Nations Educational, Scientific and Cultural Organization:
Prof. Fazlollah M. Reza, Ambassador
Permanent Delegate of Iran to Unesco
Paris, France

PROGRAM COMMITTEE

GENERAL PROGRAM CHAIRMAN
& EDITOR, PROCEEDINGS
Dr. Stanley Winkler, Senior Scientist
IBM Systems Development Division
Gaithersburg, Maryland, U.S.A.

EXECUTIVE VICE-CHAIRMAN
Edward Fuchs
Bell Telephone Laboratories
Holmdel, New Jersey, U.S.A.

VICE-CHAIRMAN, PROCEEDINGS
Dr. Raymond L. Pickholtz
George Washington University
Washington, D.C., U.S.A.

CONSULTANT
Dr. Nathaniel Macon,
Department of Mathematics
and Computer Science,
American University,
Washington, D.C., U.S.A.

VICE-CHAIRMAN, PROGRAM
Dr. Carl Hammer, Director, Computer Sciences
UNIVAC,
Washington, D.C., U.S.A.

VICE-CHAIRMAN, SPECIAL PROJECT
Dr. Robert E. Kahn
Bolt, Beranek and Newman
Cambridge, Massachusetts, U.S.A.

CONSULTANT
Gordon B. Thompson, Manager,
Communications Studies,
Bell-Northern Research,
Ottawa, Ontario, Canada

Program Committee (continued)

Jorge E. Ardila
IBM Systems Development Division
Gaithersburg, Maryland, U.S.A.

Norman A. Heck
IBM Federal Systems Division
Gaithersburg, Maryland, U.S.A.

Philip H. Enslow, Lieutenant Colonel, U.S. Army
Office of Telecommunications Policy
Executive Office of the President,
Washington, D.C., U.S.A.

Lynn Hopewell, Vice President
Network Analysis Corporation
Glen Cove, New York, U.S.A.

Louis Feldner, Common Carrier Bureau
Federal Communications Commission
Washington, D.C., U.S.A.

Dr. Peter E. Jackson
Bell Telephone Laboratories
Holmdel, New Jersey, U.S.A.

Charles Fisher, Director,
Switching Systems Engineering,
Data Transmission Company,
Vienna, Virginia, U.S.A.

Stuart L. Mathison
Arthur D. Little, Inc.
Cambridge, Massachusetts, U.S.A.

David Sibley
Economics Department, Yale University
New Haven, Connecticut, U.S.A.

Virginius N. Vaughn, Jr.
American Telephone and Telegraph Company
New York, New York, U.S.A.

Dr. Wayne B. Swift
Policyholder Service Corporation
Falls Church, Virginia, U.S.A.

Philip M. Walker
Georgetown University Law Center
Washington, D.C., U.S.A.

Prof. David L. Cohn,
Information and Control Sciences Center,
Southern Methodist University,
Dallas, Texas, U.S.A.

Stephen D. Crocker, ARPA,
Department of Defense,
Arlington, Virginia, U.S.A.

Elaine M. Kokiko
Moshman Associates
Washington, D.C., U.S.A.

Prof. Mischa Schwartz
Polytechnic Institute of Brooklyn
Brooklyn, New York, U.S.A.

Prof. Peter Weiner
Department of Computer Science,
Yale University,
New Haven, Connecticut, U.S.A.

LOCAL ARRANGEMENTS & REGISTRATIONS COMMITTEE

CHAIRMAN
Col. William T. Lincoln
Defense Communications Agency Systems Engineering Facility
Reston, Virginia, U.S.A.

VICE-CHAIRMAN, ARRANGEMENTS
Daniel M. Garnett, Cartographer
National Ocean Survey
Silver Spring, Maryland, U.S.A.

VICE-CHAIRMAN,
REGISTRATION & SOCIAL ACTIVITIES
Edward E. Boyar, Staff Industrial Engineer
Defense Communications Agency
Systems Engineering Facility
Reston, Virginia, U.S.A.

VICE-CHAIRMAN, PRINTING
Edward A. Havens, Chief,
Technical Information Office
Defense Communications Agency
Systems Engineering Facility
Reston, Virginia, U.S.A.

Mrs. Dolores B. Heck
IBM Federal Systems Division
Gaithersburg, Maryland, U.S.A.

Mrs. Linn Karnaugh
Yorktown Heights, New York, U.S.A.

Mrs. Laurel A. Sibley
New Haven, Connecticut, U.S.A.

Mrs. Linda Sweeney
UNIVAC
Washington, D.C., U.S.A.

Mrs. Vera M. Thompson
Ottawa, Ontario, Canada

Mrs. Carolyn J. Winkler
Rockville, Maryland, U.S.A.

Mrs. Rosetta L. Winkler
Bethesda, Maryland, U.S.A.

Mrs. Mimi Boyar
Potomac, Maryland, U.S.A.

INDEX OF PAPERS BY TITLE

INDEX OF PARTICIPANTS

9 a.m. 10 a.m. 11 a.m. 12 1 p.m. 2 p.m. 3 p.m. 4 p.m. 5 p.m. 6 p.m. 7 p.m. 8 p.m. 9 p.m.

TUESDAY

1
KEYNOTE SESSION
COMPUTER COMMUNICATIONS
IN INDUSTRIALLY
ADVANCED NATIONS
IBR Center

2 [illegible]
Lincoln West

3 TELECONFERENCING
Lincoln East

4 COMPUTERS AND A
TELEPHONE COMMUNICATION
SYSTEM OF THE FUTURE
Monroe

5 COMPUTER COMMUNICATIONS–
AN EMERGING TOOL IN RESPONSE
TO SOCIETY'S NEEDS
IBR East

CONFERENCE RECEPTION
IBR East

WEDNESDAY

6 IMPACTS I
IBR East

7 NETWORKS–
RECENT DEVELOPMENTS
Lincoln East

8 DATA NETWORK DESIGN
PROBLEMS I
Lincoln West

9 THE WIRED CITY
Monroe

CONFERENCE LUNCHEON
Speaker:
Dr. Hans J. von Baeyer

"CONFLICT
IN COMPUTER
COMMUNICATION"
IBR Center

10 IMPACTS II
IBR East

11 NETWORKS–
HIGHER EDUCATION
Lincoln East

12 DATA NETWORK DESIGN
PROBLEMS II
Lincoln West

13 THE ROLE OF COMPETITION
Monroe

14
COMPUTER
COMMUNICATION–
THE QUIET REVOLUTION
IBR Center

THURSDAY

15 DATA BANKS AND
INDIVIDUAL PRIVACY
IBR Center

16 TELEPROCESSING–A
COMPUTER UTILITY
Lincoln East

17 COMPUTERS, COMMUNICATION &
DISTRIBUTED HEALTH CARE SYSTEMS
Lincoln West

18 SOFTWARE ASPECTS IN
COMPUTER COMMUNICATIONS
Monroe

19 SOCIAL CONCERNS
Lincoln West

20 NETWORKS–PUBLIC DATA
Monroe

21 COMPUTERS AND THE
LIBRARIES OF THE FUTURE
Jefferson

9 a.m. 10 a.m. 11 a.m. 12 1 p.m. 2 p.m. 3 p.m. 4 p.m. 5 p.m. 6 p.m. 7 p.m. 8 p.m. 9 p.m.

conference at a glance