FILE MEMO - SUPERVISORY PROGRAM

SUBJECT: Multiprogram Control - CPU Attention Switching

DATE: May 8, 1958

When several problem programs (PP) are being operated on a multiprogrammed basis in a computer which has a single (built-in) instruction counter, the CPU can at any instant attend to only one PP at a time. Switching the attention of the CPU from one PP to another requires in a machine like Stretch a certain amount of internal dumping, restoring and bookkeeping. Because of this penalty, the attention of the CPU should not be switched from one PP to another without good cause.

Some "good causes" are described below. The list is intended to be complete, but it is quite possible that some circumstances which justify switching the attention of the CPU have been overlooked.

1. END OF PP

When a problem program is terminated either normally or abnormally, the attention of the CPU should be switched to some other PP. If, in addition, the PP is to be taken off the machine, the supervisory program (SP) will execute an appropriate unloading procedure and look for additional work load before the attention switch is completed.

2. OPERATOR INTERVENTION

Successful solution of a problem may depend upon scheduled or unscheduled human intervention. Scheduled intervention implies that the intervention is requested by the program at various stages of its execution. Unscheduled intervention implies that the operator is intervening on his own initiative. In either case, CPU activity on the PP in question is stalled and some other PP should be serviced instead.

3. WAIT

Execution of a PP has reached a point beyond which no further progress can be made until one or more of the I/O activities for this PP have been completed. The PP issues a pseudo-operation of the WAIT type to the SP (see memo on Multiprogram Control dated November 15, 1957.), and control is passed by the SP to some other PP.

4. RESUME HIGHER PRIORITY PP

This situation will be illustrated by taking an example in which two problem programs P,Q are being operated on a multiprogrammed basis. P is of higher priority than Q. The CPU attention was at some previous time switched to Q due to, say, a WAIT operation being issued by P. Now, the I/O operation for which P was waiting has been completed. Observing the higher priority of P, the supervisory program switches the attention of the CPU back to P.

Priorities may be assigned to programs in order to meet demands (or deadlines) which are external to the computing system or in order to seek the best utilization of the system itself. A simple example will illustrate this "best utilization" objective. Suppose an on-line card-to-tape program A is being multiprogrammed with a program B which requires a great deal of CPU activity compared to its I/O activity. By assigning higher priority to A than B, A will be able to get the attention of the CPU whenever it needs and, as a consequence, effective progress will be made with both programs, the I/O units assigned to A will be kept in operation, and only a slight delay will be experienced in the execution of B.

5. I/O REFUSAL

Assume that the supervisory program handles the issuing of all I/O instructions but has either a limited ability or none at all to handle queues of such instructions. The case may arise in which a PP requests an I/O activity but the SP is unable to handle it immediately: that is, it cannot issue the I/O instruction or remember it or remember its location. In such a circumstance the PP in question must be stalled and the attention of the CPU switched to some other PP.

6. OVERDUE

The supervisory program may utilize the internal clock to detect a PP which is requiring an excessive amount of CPU time to reach a certain check point. If the time interval allotted by the SP (or specified by the PP) for the PP to reach its next check point has expired, the SP will send an OVERDUE signal to the operator and switch the CPU to some other PP.

7. ON-LINE (PRIORITY SPECIFYING) INPUT

In an IDP application (for example, airlines reservations) the incoming messages normally contain some indication of the priority with which they are to be processed. The priority of a particular message may be such that the SP must switch the CPU from the program previously being serviced to some other program.

E. F. Codd
Product Planning Representative
Project 7000

EFC/jcv