Identification

The Traffic Controller System Process
Robert L. Rappaport, Michael J. Spier

Purpose

As explained in section BJ.6.00, a process has to be eligible
in order to qualify for running. Consequently, the mechanism
that puts a process in the eligible state is an integral
part of the Traffic Controller logic. Normally, this
mechanism should have been distributed among all processes
the way the Traffic Controller is. However, in order
to qualify for eligibility, a process has to be loaded,
and more often than not (depending upon the tc_data$max_loaded/
tc_data$max_eligible ratio) an unloaded process will be
selected as candidate for eligibility. Such a process
will have to be loaded first, and loading is a relatively
expensive operation. Therefore, the mechanisms of making
processes eligible and of loading/unloading processes
have been put into a special process, the Traffic Controller
System Process (TCSP). The advantages are as follows:

1. Efficiency. The TCSP is a level-1 process (see
    Scheduler, section BJ.5) and is therefore guaranteed
    to accomplish its tasks within the shortest-possible
    delay.

2. Consistency with the multi-level Scheduler. A
    high-priority process loading may be delayed because
    the process that was trying to load it had such
    low priority that it lost its eligibility. For
    just that reason, we would have had to bestow a
    temporary level-1 priority on processes engaged
    in loading.

3. Accounting. Loading and unloading is a system
    function that is dependent on system load. It
    should be kept separate of the user's accounting,
    especially as it seems highly unjust to make one user
    pay for loading another.

Following is a description of the TCSP functions.

Description

The Traffic Controller System Process is invoked whenever
there is an eligibility-vacancy. This is determined in
the two following steps:

1. A process gives up its eligibility. A process that blocks or stops itself, as well as a process that has suffered pre-emption or a timer-runout interrupt, makes itself ineligible. It also updates a count of the currently-eligible processes (tc_data$neligible).

2. A process that is put on the ready-list always checks to see whether or not an eligibility-vacancy exists (tc_data$neligible < tc_data$max_eligible).

In both cases, when it has been determined that an eligibility-vacancy exists, a wakeup is sent to the Traffic Controller System Process, which then proceeds through the following steps:

1. Make the comparison (tc_data$neligible : tc_data$max_eligible); in case of equality call block. Upon return from block resume step 1.

2. Scan the ready list for the first ineligible process. If none was found, call block. Upon return from block go to step 1.

3. Check to see whether or not the ineligible process is loaded (loading-state = 3). If yes, go to step 4.

4. Set the process' loading state to 2 (process being loaded/unloaded). Before attempting to load the process, check the number of loaded processes in the system. If tc_data$nloaded is equal to tc_data$max_loaded (system capacity saturated), find the oldest-low-priority-ineligible-loaded process in the loaded-list, and unload it. Load the newly-eligible process.

5. Make the process eligible by setting its load-state to 4. Go to step 1.

Certain system processes, such as the TCSP (|) or the idle processes must, for very obvious reasons, never give up their eligibility. These processes have a flag set "on" in their PDS which prevents them from giving up their eligibility.

Permanent eligibility cannot be granted to all system processes, else system saturation would be reached too soon. However, other system processes are given a level-1 priority, making it far less likely that they will be unloaded and lose their eligibility for relatively long periods of time. The point is that these other system

processes are kept in a favorable position to regain eligibility when needed, subject to the dynamic conditions of system load; note that they characteristically operate only for short bursts of time before going blocked and becoming temporarily ineligible, so that this favored treatment is not abused.