

Draft:

Identification

Quality of Response

L. J. Lambert, K. J. Martin

Purpose

The Transactor provides the Multics user with the mechanism for changing console response. Associated with any change in response, a user should also expect a change in the charge for the use of the system. A user may request a faster response with associated increase in cost or a slower response and thus a reduced cost.

Discussion

The response that an individual user desires at his console is relative depending on the nature of the problem being worked. In some cases, 5 minutes between interactions is sufficient while at other times 10 seconds is annoying. The response that the system is capable of providing is also relative, depending on the number of users presently using the system. Each Multics installation has an upper limit of logged-in users beyond which response is so poor that very few, if any, users are serviced satisfactorily. This upper limit is a function of hardware configuration, present load on the system and personal desires of system subscribers and is determined by system administrators.

A user with the proper authority and resources is provided the capability of improving (or degrading) his console response. The response change is associated with an increased or decreased cost of service. A notion of response levels is employed, each level being associated with a response time and cost of service. Higher levels indicate better response and increased cost.

To advance to a given level requires that the user requesting such a change have the authority to do so. This authority is granted through administrative procedures. Authority of a user is not the only criteria to consider when a level change is indicated. System efficiency and the response of other users must also be considered. For this reason each response level has a maximum number of users acceptable at this level. The acceptable number of users at the lowest response level is equal to the number of users allowed to use the system at any one time. This always gives the user wishing to degrade his response a place to go. The maximum number of users allowed at any given time depends on the current load. As more users demand better service, the total number allowed is decreased. Conversely, as users choose to degrade their service, the total number of users allowed is increased.

Since response time is a function of load, it cannot be said that each level is associated with a constant fixed response. There is a minimum response

guaranteed for each level which comes about when all levels and systems load are at capacity. It is true that each level is associated with a fixed cost for service. Figure 1 illustrates a response level table, showing five levels of response, corresponding user maximums, cost and associated schedules.

The improvement of console response depends on many things. Three of the most important ones are as follows:

- Processor service - the frequency and duration that the processes of a user are run by processors.
- File System response - the treatment that the file system gives the user's processes' requests for pages.
- I/O Requests - The handling of the user's I/O requests by the I/O system.

The initial implementation of this Transactor facility considers only that improvement or degradation that can be realized by changing the processor service for a user. Changes in this area indirectly effect the other areas in that processor time can be made available more frequently and/or for longer periods to process page and I/O requests. Better processor service, on the other hand, does not affect the length of time to process I/O requests or page requests spend in the various queues. Changes in file system and I/O response can only be brought about by changes to the queuing strategies used in these areas for the process in question. The initial implementation is designed to facilitate adding the handling of file system response and I/O system response.

In return for improved or degraded service, a user pays accordingly. This increase or decreased cost is charged in terms of a rate change for processor time used.

Usage

A user desiring to improve or degrade his response presents his request in terms of command to the Transactor. The form of the command allows him to specify the level to which he desires to go or simply that he wishes to move up or down one level. It is also possible to inquire as to what level he is presently at and the rate associated with this level.

When a change in response is requested, the user's authority to proceed to this level is checked in his user profile and if allowed, the request is processed. If the authority is not consistent with the request, the user is informed that his request was denied because of his lack of authority. A user's response authority is in terms of the highest response level to which he may go. If the number of users already at the requested level is at a maximum, the user is informed that this request cannot be granted. If the level change is possible, the change is made and the new rate imposed on the user. Once the user is at the new level, he may decide if the realized improvement is worth the fee paid. The user remains at this level until he himself decides to change levels, or his processes terminate.

Implementation

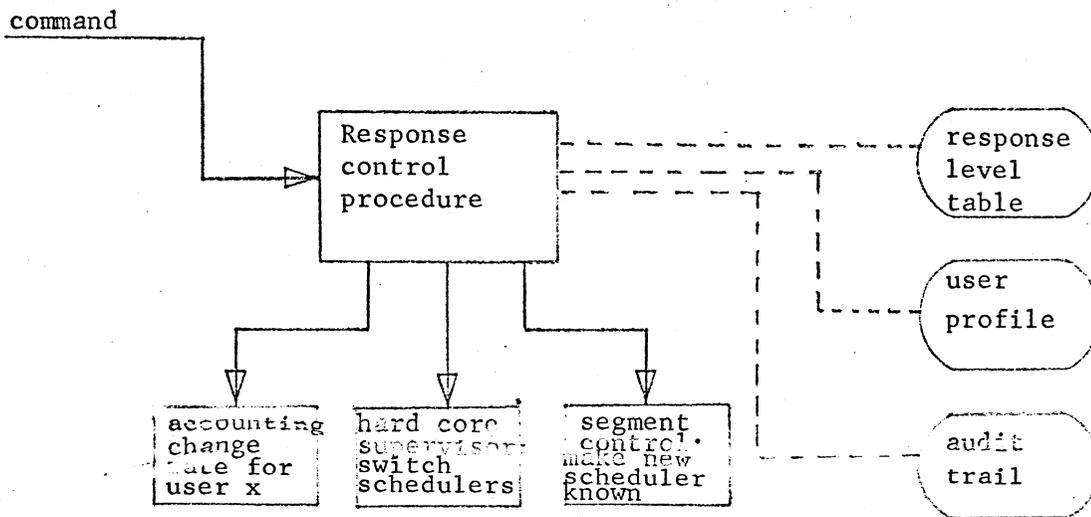
As stated earlier, the initial implementation is concerned only with changing processor service when a request to change console response has been received. Figure 2 is a block diagram of the response control procedure. A request causes the response control procedure to be called. After checking the users authority and the number of users at the desired level, this procedure determines from the response level table the Traffic Controller scheduler associated with the level requested. The new scheduler is brought into play by making it known to the process and then calling into the hard core supervisor indicating that this segment is now to be used as the scheduler for the process. This is repeated for every process in the user's family of processes. The response level indicator in the user profile is changed to reflect the new level.

Accounting is called and informed of the new rate to be used when changing the processes of this user, for their use of processor time. The data base giving the total number of users allowed is changed to reflect the response change for this user. Since a commitment of funds is involved over and above those originally contracted for, an audit trail is prepared which reflects the users request and acceptance of level changes. These records are kept for a finite length of time. Each Multics installation determines the retention period. *how?*

Response level	User limit	Cost	Traffic Controller Scheduler
0	200	1 cost unit/user	A
1	50	2 cost units/user	B
2	20	3 cost units/user	C
3	5	4 cost units/user	D
4	1	5 cost units/user	E

logged in user limit - 200

figure 1 - Response level table



—— call
 - - - - access

figure 2 - response command