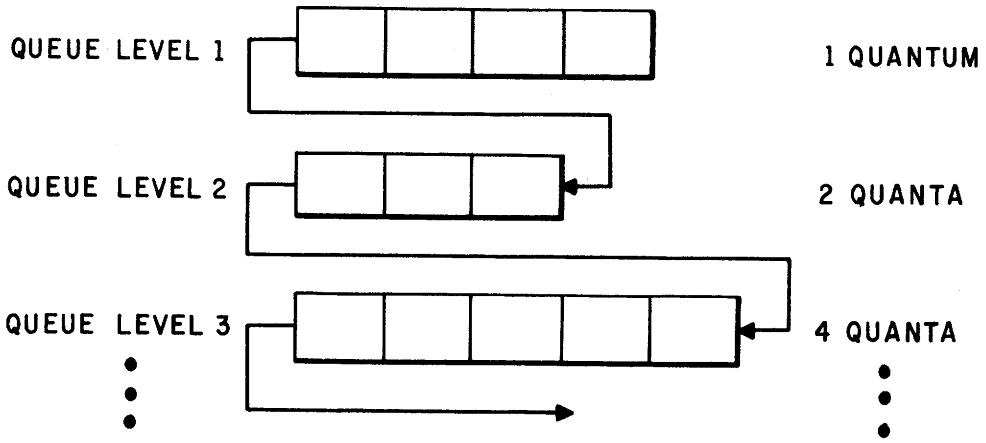

FIGURE 3. Multilevel queue.

user's tolerance to system delay in answering a question is something like an exponential function of the actual computing time needed to answer the question—i.e., that the user prefers a system that delays answering complicated questions in return for quick response to simple questions. Thus, a queueing algorithm which was cognizant of the user programs' past history would be able to favor the small computer-time requests (whose users were less tolerant of slow response) at the expense of penalizing those user programs which demanded a great deal of computer processing time.

One answer to some of the objections to the round-robin method is a *multilevel queue*, which has been implemented on this system. In this system, there are 12 queues into which a user can be placed. A user in the first (highest) queue is given one quantum of computing time; the time allotted to a user in any other queue is twice the time allotted to the user in the next higher queue. Thus, a user in queue n will receive 2^{n-1} quanta of time. If a program in queue n uses all the time allowed in that queue, then the program is placed in queue n+1 (see Fig. 3). On the other