OPTIMALITY OF THE LAST-IN-FIRST-OUT (LIFO) SERVICE DISCIPLINE IN QUEUEING SYSTEMS WITH REAL-TIME CONSTRAINTS

M. H. Kallmes, D. Towsley, and C. Cassandras

COINS Technical Report 89-21

February 1989

OPTIMALITY OF THE LAST-IN-FIRST-OUT (LIFO) SERVICE DISCIPLINE IN QUEUEING SYSTEMS WITH REAL-TIME CONSTRAINTS [(1)]

Michelle Hruby Kallmes
Dept. of Electrical and
Computer Engineering

Don Towsley
Dept. of Computer and
Information Science

Christos G. Cassandras
Dept. of Electrical and
Computer Engineering

University of Massachusetts
Amherst, MA 01003

ABSTRACT

We consider a G/G/1 queueing system where the objective is to maximize the probability that a customer's system time does not exceed a given *deadline*. The deadline is defined to be the maximum amount of time the customer can spend in the system. We show that for deadlines that are i.i.d. random variables with concave cumulative distribution functions, LIFO gives the highest probability of success, and FIFO the lowest, over the class of all work-conserving non-preemptive service disciplines that are independent of service time and deadline. Extensions to policies that allow preemption in G/M/1 systems, deadlines imposed on queueing (rather than system) time, and multi-server queues are provided. Results are illustrated by the simple example of an M/M/1 queue with exponential deadlines. In addition, the case of a G/G/1 queue with constant deadlines is considered.

February 1989

28th CDC

(1) This work is supported in part by the Office of Naval Research under contracts N00014-87-K-0304 and N00014-87-K-0796 and by the Rome Air Development Center under contract F30602-88-D-0027.

1. INTRODUCTION

We consider a G/G/1 queueing system where every arriving customer is characterized by a *deadline D*. A customer's deadline is defined to be the maximum amount of time the customer can spend in the system. Deadlines are assumed to be i.i.d. random variables with a cumulative distribution function $G_D(\cdot)$. If a customer's system (sojourn) time exceeds the assigned deadline, then the customer is considered to be *lost*. (Note: a customer that exceeds his deadline is not removed from the system.) This model is particularly useful in communication systems where messages are required to reach their destination within a specified time interval, and in computer systems where tasks must be executed under real-time constraints.

Let Π_{np} denote the class of work-conserving, non-preemptive service disciplines that are independent of the service times and deadlines. Thus, when the server adopts a policy $\pi \in \Pi_{\text{np}}$, it selects at every service completion instant a customer among those presently in the queue without any information regarding service times or deadlines of the customers. Letting R_π denote the ergodic system time of a customer under $\pi \in \Pi_{\text{np}}$, the problem is to determine a policy π^* optimal in Π_{np} such that the probability $\Pr[R_\pi \leq D]$ is maximized. Our main result in this paper is that the Last-In-First-Out (LIFO) policy is optimal, as long as $G_D(t)$ is concave in t. Moreover, the First-In-First-Out (FIFO) policy provides the worst $\Pr[R_\pi \leq D]$ in this case.

The result is particularly interesting in view of the fact that for the common case of constant deadlines ($D = \tau$), no such universal claim about the optimality of a certain policy can be made without specific knowledge of the parameters of the system and deadline. For example, although LIFO outperforms FIFO in the case of *tight* deadlines (small τ), FIFO provides a higher $\Pr[R_\pi \leq D]$ than LIFO for *loose* deadlines (large τ).

With our key result as a starting point, it is possible to provide extensions to the class of policies allowing preemption, and to the problem of maximizing $\Pr[Q_\pi \leq D]$, where Q_π is the queue waiting time. In addition, we extend the results to single-server queues with bulk arrivals, and to multiple-server queues. Furthermore, the result raises a number of questions pertaining to the properties of the LIFO policy in models where deadlines play a critical role in determining system performance. For instance, can the LIFO policy still be optimal if $G_D(t)$ does not satisfy the condition above? It is also possible that LIFO outperforms FIFO even when the latter policy is used in conjunction with state-dependent admission control to the queueing system.

This paper is organized as follows. Section 2 defines a convex ordering and states some existing results. The main result is derived in section 3; extensions to this result are provided in section 4. In section 5, we present examples of queues with (a) exponential deadlines, and (b) constant deadlines.

2. DEFINITIONS AND PRELIMINARY RESULTS

Our main result is based on the idea of establishing a certain type of partial ordering between two random variables X and Y defined over all non-negative real numbers. In particular, we shall use the following definition of a *convex* ordering [1]:

Definition: *X is smaller than Y in the convex ordering, denoted by $X \leq_c Y$, if and only if:*

$$E[g(X)] \leq E[g(Y)]$$

for any convex function $g: [0,\infty)\rightarrow\Re$.

The following Theorem, due to Shanthikumar and Sumita [1], establishes a convex ordering for the ergodic system time in a G/G/1 queueing system. Recall that Π_{np} is the class of work-conserving non-preemptive service disciplines that are independent of the service times and deadlines. Clearly, the FIFO and LIFO policies belong to this class.

Theorem 1a: *For any $\pi\in\Pi_{np}$, let R_π be the ergodic system time in a G/G/1 queueing system under π. Then:*

$$R_{FIFO} \leq_c R_\pi \leq_c R_{LIFO}.$$

Shanthikumar and Sumita [1] derive a similar result for the G/M/1 queue and the more general class Π of work-conserving service disciplines that are independent of the service times and deadlines. In this case, preemption is allowed, and we use LIFO-P to represent the LIFO preemptive/resume policy, which belongs to Π.

Theorem 1b: *For any $\pi \in \Pi$, let R_π be the ergodic system time in a G/M/1 queueing system under π. Then:*

$$R_{\text{FIFO}} \leq_c R_\pi \leq_c R_{\text{LIFO-P}}.$$

The proofs of these Theorems in [1] are based on interchange arguments applied to sample paths of the queueing system under consideration. It can be shown, for instance, that if some policy $\pi \in \Pi_{\text{np}}$ other than FIFO is used in a G/G/1 system, then interchanging the first customer not served under FIFO with the customer satisfying the FIFO policy results in a lower value of the ergodic system time in the convex ordering sense.

3. NON-PREEMPTIVE G/G/1 QUEUES

The following theorem establishes the optimality of the LIFO service discipline; furthermore, it shows FIFO to be the worst.

Theorem 2: *Let deadlines be i.i.d. random variables with a cumulative distribution function $G_D(\cdot)$. For a G/G/1 queue and any service discipline $\pi \in \Pi_{\text{np}}$, if $G_D(t)$ is a concave function of t, then*

$$\Pr[R_{\text{LIFO}} \leq D] \;\geq\; \Pr[R_\pi \leq D] \;\geq\; \Pr[R_{\text{FIFO}} \leq D]. \tag{1}$$

Proof: Define the following function:

$$h(r) \;=\; \Pr[D \geq r] \;=\; 1 - G_D(r).$$

But $h(r)$ is convex since $G_D(r)$ is concave; therefore $h(R_\pi)$ is convex and it follows by Theorem 1a that

$$\mathrm{E}[h(R_{\text{LIFO}})] \;\geq\; \mathrm{E}[h(R_\pi)] \;\geq\; \mathrm{E}[h(R_{\text{FIFO}})].$$

Let $F_\pi(\cdot)$ be the cumulative distribution function of R_π under policy π. Observing that

$$\mathrm{E}[h(R_\pi)] \;=\; \int_0^\infty \Pr[D \geq R_\pi \,|\, R_\pi = r]\, \mathrm{d}F_\pi(r) \;=\; \Pr[D \geq R_\pi],$$

the result (1) follows. **QED**

Remark: Constraining the cumulative distribution function to be concave is equivalent to requiring the probability density function to be non-increasing for $t > 0$. For example, a deadline that is exponentially distributed would satisfy this condition, as would one that is distributed uniformly on the interval $[0,b]$, $b > 0$.

4. EXTENSIONS

The following results are easily derived as extensions of Theorem 2. We omit the proofs.

G/M/1 Queues with Preemption Allowed. Consider the larger set of policies, Π, where preemption is allowed. If we restrict ourselves to exponential servers, then with application of Theorem 1b, the following ordering result can be obtained for concave $G_D(\cdot)$:

$$\Pr[R_{\text{LIFO-P}} \le D] \ \ge\ \Pr[R_\pi \le D] \ \ge\ \Pr[R_{\text{FIFO}} \le D]. \tag{2}$$

Queues with Deadlines on Queueing Time. We can extend the results of of the previous sections to systems where the deadline applies to queueing time rather than system time. That is, we can replace R_π by Q_π where Q_π is the stationary queueing time, and the results will still hold. The proof relies on certain properties of convex ordering that can be found in [2, p.272].

Queues with Bulk Arrivals. Here we show that the results presented in the previous sections can be extended to bulk arrival queues. For single-server queues where the bulk size of the nth arrival is random, Theorems 1a and 1b still hold [1]. Therefore, it can be shown that the inequalities in equations (1) and (2) hold for queues with bulk arrivals assuming that the conditions on the deadline distributions are met.

Multiple-Server Queues. We can show that the results of Theorem 2 can be extended to G/G/c queues. The proof entails first extending the convex ordering result of Shanthikumar and Sumita [1] to apply to multiple-server queues.

5. EXAMPLES

We will look at the two special cases of an M/M/1 queue with exponential deadlines, and a G/G/1 queue with constant deadlines. In the case of exponential deadlines, the required property is satisfied for Theorem 2 to hold, and we give an example to show the magnitude of the difference between the FIFO, LIFO, and LIFO-P policies. In addition, we will show that in the case of a constant deadline, whose cumulative distribution function does not satisfy the concavity property, no policy is always optimum over all choices of the deadline. In fact, we can show that there exists a critical deadline, τ_{crit}, above which one policy outperforms the other, and below which the converse is true.

5.1. M/M/1 Queue with Exponential Deadlines

Consider an M/M/1 queueing system with all deadlines drawn from an exponential distribution. Let the arrival rate be λ; the service rate μ; and the mean deadline $\frac{1}{\theta}$. Using standard queueing theoretic techniques, we can show that the analytical expressions for the probabilities of interest are:

$$\Pr[R_{\text{LIFO-P}} \le D] = \frac{1}{2\lambda}\left(\theta+\lambda+\mu-\sqrt{(\theta+\lambda+\mu)-4\lambda\mu}\,\right), \tag{3}$$

$$\Pr[R_{\text{LIFO}} \le D] = \begin{cases} \frac{1}{2(\theta+\mu)}\left(\theta-\lambda+3\mu-\sqrt{(\theta+\lambda+\mu)-4\lambda\mu}\,\right), & \lambda<\mu, \\ \frac{\mu}{2\lambda(\theta+\mu)}\left(\theta+\lambda+\mu-\sqrt{(\theta+\lambda+\mu)-4\lambda\mu}\,\right), & \lambda\ge\mu, \text{ and} \end{cases} \tag{4}$$

$$\Pr[R_{\text{FIFO}} \le D] = \begin{cases} \frac{\mu-\lambda}{\theta+\mu-\lambda}, & \lambda<\mu, \\ 0, & \lambda\ge\mu. \end{cases} \tag{5}$$

Through algebraic manipulation of equations (3)-(5), one can show that, indeed, the following ordering is satisfied for any choice of λ, μ, and θ:

$$\Pr[R_{\text{LIFO-P}} \le D] \ge \Pr[R_{\text{LIFO}} \le D] \ge \Pr[R_{\text{FIFO}} \le D].$$

To illustrate the difference in performance under the different policies, we plot $\Pr[R_\pi \le D]$ for $\pi =$ LIFO-P, LIFO, and FIFO as a function of λ for the case where $\mu = 1$ and $\theta = 1/2$ (see Fig. 1).

5.2. Constant Deadlines

Consider a G/G/1 queue where deadlines are constant and equal to τ. We can show that for two policies $\pi_1, \pi_2 \in \Pi_{\text{np}}$ such that $R_{\pi_1} \le_c R_{\pi_2}$, there exists a critical deadline, τ_{crit}, such that

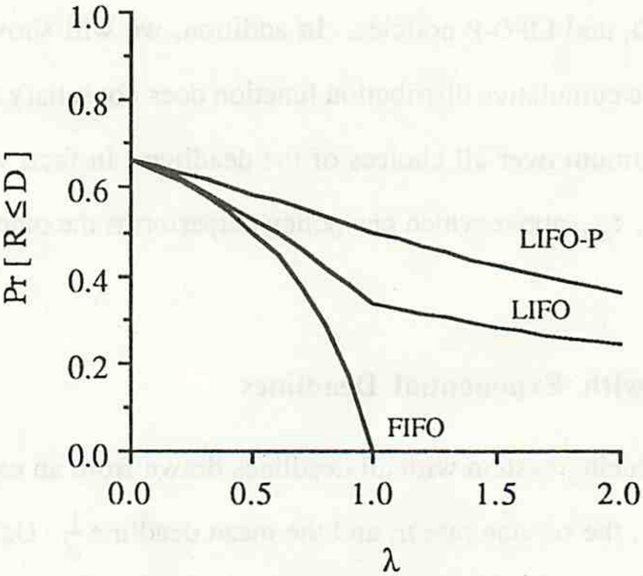

Figure 1: *Comparison of LIFO-P, LIFO, and FIFO Service Disciplines for an M/M/1 Queue with Exponential Deadlines*

$$\Pr[R_{\pi_1} \le \tau] \begin{cases} < \Pr[R_{\pi_2} \le \tau], & \tau > \tau_{\text{crit}}, \\ = \Pr[R_{\pi_2} \le \tau], & \tau = \tau_{\text{crit}}, \\ > \Pr[R_{\pi_2} \le \tau], & \tau < \tau_{\text{crit}}. \end{cases} \tag{6}$$

The proof comes from the fact that $\Pr[R_\pi \le \tau]$ is just the cumulative distribution function (cdf) of R_π and a result in [3, p.107] stating that if two random variables with the same mean have a convex ordering, then their cdf's cross exactly once and in a direction consistent with (6).

Remark: The critical deadline, τ_{crit}, depends on the parameters of the interarrival and service-time distributions. Therefore, to compare two policies, one needs exact values for (a) the deadline, and (b) the system parameters.

6. CONCLUSIONS

The LIFO service discipline in G/G/1 systems is often regarded as undesirable, since it results in a system time distribution with the same mean but higher variance compared to FIFO. In the context of the problem we have considered, however, this feature can be exploited. Our key result is that for G/G/1 queues where customers have deadlines drawn from a concave cumulative distribution function, a LIFO service discipline is optimal over the set of work-conserving non-preemptive policies that are independent of service demands and deadlines. We also show FIFO to be the worst policy. In addition, we extend the basic result to policies that allow preemption and to more general systems.

Our results suggest several interesting questions pertaining to the desirable properties of the LIFO service discipline. For example, we can ask how a LIFO policy compares to a FIFO policy with state-dependent admission control. There is, in fact, evidence that LIFO can do better even in this case. Another idea is to make the same comparisons in a system where customers are removed from the queue as soon as they are detected to have exceeded their deadline. Further, we can question whether LIFO is optimal for any deadline distributions that do not satisfy the concavity constraints.

REFERENCES

[1] J. G. Shanthikumar and U. Sumita, "Convex Ordering of Sojourn Times in Single-Server Queues: Extremal Properties of FIFO and LIFO Service Disciplines," *J. Appl. Prob.*, vol. 24, pp. 737-748, 1987.

[2] S. M. Ross, *Stochastic Processes*, Wiley (New York), 1983.

[3] R. Barlow and F. Proschan, *Statistical Theory of Reliability and Life Testing*, To Begin With (Silver Spring, MD), 1981.