A CLASS OF FIFO QUEUES
ARISING IN COMPUTER SYSTEMS

by

Edward G. Coffman* and Micha Hofri**

Technical Report #87
November 1976
REVISED VERSION

* Columbia University, Computer Science Dept. New York, U.S.A.

** Technion-Israel Institute of Technology, Computer Science Dept. Haifa, Israel.
ABSTRACT

Secondary memory devices are modeled as single server queueing systems. The non-random access to data within these devices is explicitly accounted for, as "set-up" times. Requests are typed by the location of the desired record. No distinction is made between 'read' and 'write' requests. Each request is assumed to be satisfiable from one location on the device (e.g. a single directory search may result in a number of distinct requests). Requests arrive according to a homogeneous Poisson process. The types of successive requests form a first order Markov chain, which are an approximation of reality. Alternative computational procedures and closed expressions are given for queue length, waiting times (and device utilization). Some specializations, to disks and drums are presented. Only FIFO service is considered.

C.R. Categories: 3.73, 5.5, 6.34, 8.3.

Key words: Mass Storage; secondary memory; magnetic drum; magnetic disk; single server FIFO queue; semi-Markov service.
1. **INTRODUCTION**

Important congestion points in general purpose computer systems frequently occur through interactions with secondary storage devices. In such cases the efficiency with which information is exchanged between these devices and primary (e.g. core) storage determines the system's maximum throughput or work-rate. Secondary storage units such as magnetic drums, disks, bubble memories, and tapes (whether singly or within libraries) have the characteristic feature that the total service time of a (read or write) request depends on the location addressed by the request previously served. Of course, it is precisely this property that specifies the manner in which these devices fail to be random-access, as are primary storage devices.

Our purpose in this paper is to present and analyze a mathematical model that will explicitly take into account the above characteristic of non-random access devices. Since the difficulty arises mainly from the unpredictable arrivals of requests, it is natural that a stochastic model is required for a realistic presentation of the salient features of these systems.

A specific goal will be to provide a FIFO (First-In-First-Out service) queueing analysis of secondary storage devices sufficiently general to embrace the detailed structure of a large majority of existing systems. The parameters of the mathematical model will include a stationary, discrete probability distribution describing the patterns by which requests address information on secondary storage devices (successive addresses are allowed as well to form a first order Markov chain). Such patterns normally influence system performance, and they are determined by the mechanism which allocates specific storage locations to records (units of information). Thus, in the calculation of conventional performance measures we shall also briefly consider
the essentially combinatorial problem of determining the influence of different record allocations.

The next two sections present a general model and its analysis. The remainder of the paper specializes the results to certain common secondary storage devices and discusses alternative computational methods.

2. THE MATHEMATICAL MODEL

The devices discussed in the previous section will be modeled as a single-server facility as illustrated in Fig. 1. Incoming requests are immediately inducted into service when the facility is idle. Arrivals at a busy facility enqueue for service, and there is no limit to the number of such requests that may wait at any given time. All service period contain an initial period of set-up delay, possibly of zero length. The selection from the queue for service, at the termination of a service period, is done without prior knowledge of the requested service times. During all of our analysis we consider selection procedures that provide service in the order of arrival (FIFO), but some of the results admit more general regimes.

Requests are of \( N \) types, simply called types 1 through \( N \). The probability that an arriving request is of type \( j \), given that the preceding one was of type \( i \) is \( p_{ij} \), and is otherwise independent of the state of the system and its history. These "transition probabilities" form a matrix \( P \) with an invariant probability vector we denote \( \pi \). (We are only interested in situations where \( P \) is irreducible and all its states are recurrent.) The matrix with all its rows equal to \( \pi \) will be denoted by \( \tilde{P} \). A request of type \( j \), which is immediately preceded by a request of type \( i \), requests service with duration \( S_{ij} \) drawn from a distribution \( F_{ij}(\cdot) \), independently
of the other descriptors of the state of the system. This service period is generally the sum of two components

\[ S_{ij} = T_{ij} + K_j, \]

where \( T_{ij} \), the set-up time, is the time it takes the service facility to switch over from a state of having finished the service of a type \( i \) request to the beginning of service for a type \( j \) request. The quantity \( K_j \) depends on the request type, does not depend on the state of the system and usually represents the actual transmission time of the information. In most of the applications towards which this paper is directed, the variables \( S_{ij} \) are in fact constants.

---

Figure 1: Secondary memory as a service facility with a waiting queue.

In some situations we find it expedient to distinguish the service rendered to a request that starts a busy period (i.e. it finds upon arrival on idle system). Invariably, it is the set-up time \( T_{ij} \) that is affected, and its value under these circumstances will be denoted \( T_{ij}^0 \). Associated
with \( T_{ij}^0 \) is a service duration \( S_{ij}^0 \), but the value of \( K_j \) is not changed. The arrival process of requests is assumed Poisson, with rate \( \lambda \), homogeneous in time and independent of the state of the system.

We shall be interested primarily in steady state behavior. We observe the system at the epochs of departure of requests. Since arrivals and departures happen singly, the distribution of the states of the system at these epochs is the same as at the arrival epochs, and also equal to the so-called "long term" distribution. We let \( X_n \) denote the number of requests in the system immediately following the departure of the \( n \)-th request, the one in service included. \( \eta_n \) denotes the waiting time of the \( n \)-th request, which terminates at the beginning of the \( n \)-th set-up time. We let \( S \) be the random variable denoting general service time, \( F(\cdot) \) the corresponding distribution, and \( \mathcal{E}(\cdot) \) its LST.

3. ANALYSIS OF THE MODEL

The major difference between the model we investigate here and standard queueing models is the dependence between successive services. Depending on the type of device and its operating procedures, this relationship may even extend "across" an intervening idle period. Special cases of our model can be treated as applications of Skinner's [S] model (with a loss of structure severe enough to preclude its use for most of the devices for which our model is intended). For an example, see Fuller and Baskett [FB] for approximate analyses of FIFO paging drums. A queueing model with similar structure - the main difference being that no distinction is made between a general service and a one that starts a busy period - was treated in detail by Neuts [N66].
We begin with an analysis that is independent of the order of arrivals. Then we proceed to evaluate the waiting times for a FIFO queue.

**System capacity** - As one usually finds in queueing systems, the highest input rate that the facility can sustain is given by \( \lambda_{\text{max}} = \frac{1}{E(S)} \), where

\[
E(S) = \sum_i \sum_j \bar{p}_{ij} E(S_{ij}) .
\]  

(1)

This statement will not be proved explicitly here. We note the occurrence of the corresponding discontinuity point in numerical calculations.

**Queue length** - We observe that \( (X_n, J_n; n = 1, 2, \ldots) \) where \( J_n \) is the type of the \( n \)-th departing request, is an aperiodic, irreducible and, for low enough input rates, recurrent Markov chain (MC). We proceed first to evaluate the probability generating function (pgf) of the steady state distribution of the number in system. This will turn out to entail most of the complexity of the analysis that we require.

We define for \( 1 \leq i \leq N, \ x \geq 0 \)

\[
p_i(x) = \lim_{n \to \infty} P(X_n = x | J_n = i) \]  

(2)

where, as usual, the vertical bar is to be read as "given that...", and

\[
G_i(z) = \sum_{x=0}^{\infty} p_i(x) z^x .
\]  

(3)

The dynamics of our MC are embodied in the matrix \( P \) and the relation

\[
X_{n+1} = X_n - U_n + Y_{n+1} ,
\]  

(4)

where \( U_n \) is 0 when \( X_n = 0 \) and is 1 otherwise, and where \( Y_{n+1} \) is the number of arrivals during the service of the \( (n+1) \)-st request. We proceed in a standard way to obtain directly from (4)
\[ P(X_{n+1} = x \mid J_{n+1} = j)P(J_{n+1} = j) = \sum_{i=1}^{N} p(j = i)P(X_n = 0 \mid J_n = i) \]

\[ P(Y_{n+1} = x, J_{n+1} = j \mid X_n = 0, J_n = 1) + \sum_{r=1}^{\infty} P(X_n = r \mid J_n = i) \]

\[ P(Y_{n+1} = x - r + 1, J_{n+1} = j \mid X_n = r, J_n = i) ; \quad x > 0, 1 \leq j \leq N. \]

The distribution of \( Y_{n+1} \) is now derived. It obviously depends on the duration of service of the \((n+1)\)-st request. As mentioned above, we distinguish between a departure followed by an idle period (with a subsequent service distributed according to \( F_{ij}^0(\cdot) \)), and a departure for which the next service commences immediately (and is distributed according to \( F_{ij}(\cdot) \)); the set-up duration may be different in the two cases.

Hence

\[ P(Y_{n+1} = x \mid X_n = 0, J_n = i, J_{n+1} = j) = \int_{s=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^x}{x!} F_{ij}^0(ds) \quad (6) \]

and

\[ P(Y_{n+1} = x \mid X_n > 0, J_n = i, J_{n+1} = j) = \int_{s=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^x}{x!} F_{ij}(ds) \quad (7) \]

We substitute (6) and (7) properly deconditioned from \( J_{n+1} \) in (5), multiply by \( z^x \) and sum over all values of \( x \); since the MC is recurrent we may drop the subscripts \( n \) and \( n+1 \) to obtain the limiting equation

\[ G_j(z) = \sum_{i=1}^{N} \pi_i P_{ij} \left\{ \sum_{x=0}^{\infty} \int_{s=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^x}{x!} F_{ij}^0(ds) \right\} + \sum_{r=1}^{\infty} z^r P(X=r \mid J=i) \sum_{x=0}^{\infty} \frac{1}{z} \int_{s=0}^{\infty} e^{-\lambda s} \frac{(\lambda s)^{x-r+1}}{(x-r+1)!} F_{ij}(ds) \}

\[ = \sum_{i=1}^{N} \{ \pi_i L_{ij}^0(a) + \frac{1}{z} (G_i(z) - \pi_i) L_{ij}(a) \} \quad 1 \leq j \leq N, \]
where $\pi_i = P(X=0|J=i)$, $a = \lambda(1-z)$, $\xi_{ij}$ (respectively $\xi^0_{ij}$) is the Laplace-Stieltjes transform of $F_{ij}$ (respectively $F^0_{ij}$) and $L_{ij}(L^0_{ij})$ is given by $\tilde{p}_i p_{ij} \xi_{ij}/\tilde{p}_j (\tilde{p}_i p^0_{ij} \xi^0_{ij}/\tilde{p}_j)$. The various changes of order of summation are allowed, since all the sums are trivially absolutely convergent.

The $N$ equations can be written in a more convenient and compact matrix form

$$A(z)\xi(z) = B(z)\pi$$

(9)

where $\pi$ and $\xi(z)$ are the obvious vectors and

$$A_{ij}(z) = z \delta_{ij} - L_{jj}(a), \quad 1 \leq i, j \leq N$$

$$B_{ij}(z) = z L^0_{jj}(a) - L_{jj}(a),$$

(10)

where $\delta_{ij}$ is 1 if $i=j$ and is 0 otherwise. Equation (9) has the formal solution

$$\xi(z) = A^{-1}(z)B(z)^\dagger \pi.$$ (11)

The unknown boundary probabilities $\pi_i$ now have to be deduced from normalization and analyticity arguments. The first is

$$\xi(1) = \pi.$$ (12)

Second, letting $C(z) = |A(z)|A^{-1}(z)$ be the adjoint matrix of $A(z)$, we must have

$$C(z)B(z)^\dagger \pi = \delta$$

(13)

at all points $\zeta$, $|\zeta| \leq 1$, which are solutions of

$$|A(z)| = 0.$$ (14)
Each of the equations in the system (13) is homogeneous, and thus (13) has to be supplemented by an equation that is inhomogeneous. Equation (12) does not give this directly, and we obtain it by noting that if \( \pi \) is the probability an incoming request finds an empty system, then balance equations yield

\[
\pi = \sum_{i=1}^{N} \tilde{p}_i \pi_i \tag{15}
\]

\[
1 - \pi = \lambda \sum_{ij} \tilde{p}_{ij} \pi_i [E(S_{ij}^0) - E(S_{ij})] + \lambda \sum_{ij} \tilde{p}_{ij} E(S_{ij}) \tag{16}
\]

Equations (15) and (16) can now be combined to yield the necessary addendum to (13)

\[
\sum_{i} \tilde{p}_i \{1 + \lambda \sum_{j} \tilde{p}_{ij} [E(S_{ij}^0) - E(S_{ij})]\} = 1 - \lambda E(S); \quad E(S) \equiv \sum \sum \tilde{p}_{ij} E(S_{ij}) \tag{17}
\]

An alternative method to compute the \( \pi_i \) is given in Section 5.

We present here an important result concerning those points at which \( |A(z)| \) vanishes.

**Theorem 1** The determinant of \( A(z) \) vanishes at \( z=1 \) and at precisely \( N-1 \) points that satisfy \( |z| < 1 \).

**Proof.** The first claim is immediate by substitution and using \( \tilde{p}p = \tilde{p} \).

The second claim can be proved as follows:

Let \( a_i(z) \) denote the \( N \) eigenvalues of the matrix \( L(a) \), \( |z| < 1 \). They need not necessarily be distinct, but in such a case we "perturb" the matrix \( P \) to separate them and invoke continuity arguments to assure that the number of roots of (14) stays the same.* In the sequel we assume

* What is not necessarily preserved is the strict inequality \( |\zeta| < 1 \). It may happen that a root \( \zeta \neq 1 \) will have \( |\zeta| = 1 \).
They are distinct. Then Eq. (14) can be rewritten as

\[
\prod_{i=1}^{N} (z - a_i(z)) = 0. \tag{18}
\]

Since the matrix \( L^T(a) \) is term by term strictly smaller (for \( z \neq 1 \)) in absolute value than \( L^T(0) \), a stochastic matrix (with spectral radius 1), all of \( a_i(z) \) satisfy (see [G], vol. II, p. 57) \(|a_i(z)| < 1\). Rouché's theorem can be now applied to each of the factors of Eq. (18), to the effect that it has a single root in the open unit disk \(|z| < 1\) (except for that factor where \( a_i(z) = 1\)).

We note a phenomenon that is interesting for its numerical implications: when successive request types are independent (i.e. \( p_{ij} = p_j = \tilde{p}_j \)) and \( \lambda = 0 \) we have \(|A(z)| = (z-1)z^{N-1}\), and thus it has one simple zero at \( z=1 \), and one of multiplicity \( N-1 \) at \( z=0 \). When \( \lambda \) increases continuously from zero to its operational value, the roots of the equation (14) (which consists of continuous functions only) also move continuously in the \( z \)-plane. Writing \( L_{ij} \) as a power series in \( \lambda \),

\[
L_{ij}(a) = \sum_{k=0}^{r} b_{ijk} \lambda^k + o(\lambda^r)
\]

we obtain \(|A(z)|\) as a polynomial in \( z \) of degree \( N \), with a simple zero at \( z = 1 \) and a zero at \( z=0 \) of multiplicity \( N-r-1 \) (\( r < N-1 \)). This is obtained by using \( O(\lambda) \) as an approximation for the other roots. The Lévy-Desplanques theorem assures us that for \(|z| = 1\), only \( z=1 \) is a
root of the determinant*; since the roots departed continuously, they perforce are somewhere in the unit disk. We may expect then for small values of $\lambda$ to have very close together roots, which are are hard to separate and accurately evaluate. The method described in Section 5 is superior in such circumstances.

To obtain the expected number in system $e_i (1 \leq i \leq N)$, we differentiate (11) at $z=1$. We obtain the set of relations after some cancellations:

$$e_i = \frac{d}{dz} G_i(z) \bigg|_{z=1} = \frac{1}{\delta_1^2} \left( (\delta_2 C_0 B_1 + \delta_1 C_1 B_1 + \delta_1 C_0 B_2) \right)_{i} , \quad (19)$$

where the following derivatives are all with respect to $z$, and evaluated at $z=1$: $\delta_1 = |A(z)|$, $\delta_2 = \frac{1}{2} |A(z)|''$, $C_0 = C(1)$, $C_1 = -C'(z)$, $B_1 = -B'(z)$ and $B_2 = \frac{1}{2} B''(z)$. The values of these quantities, in terms of the model parameters are given in the appendix. The overall mean queue size is then given by $\sum_{i=1}^{N} \rho_i e_i$.

Waiting time - We consider now the waiting time in a linear (FIFO) queue. Let $\eta_n$ denote the waiting time of the $n$-th request. As in any single server linear queue,

$$\eta_{n+1} = (\eta_n + S_n - t_n)^+ , \quad n = 0, 1, \ldots \quad (20)$$

except that here the requests types have to be incorporated in the

* The theorem states, in one of its versions, that if a matrix $C$ satisfies the condition $|C_{ii}| > \sum_{j \neq i} |C_{ij}|$, for all $i$, then it is non-singular.

For the matrix $A(z)$, where $|z| = 1$, this condition reduces, using the inequality $|a-b| \geq |a| - |b|$, to the requirement $|L_{ij}[\lambda(1-z)]| < 1$, which is true when $\lambda > 0$ and $\text{Re}(1-z) > 0$ [MM].
calculation. $S_n$ is the service duration of the $n$-th request, and $t_n$ is the time between its arrival and that of the $n+1$-st. $t_n \sim \text{exp}(\lambda)$, independently of the other variables.

Define

$$w_{ij}^n(x) = P(n \leq x | J_0 = i, J_n = j)$$

$$w_{ij}^n = w_{ij}(0) \quad (21)$$

$$\tilde{w}_{ij}^n(s) = \int_0^\infty e^{-sx} d\tilde{w}_{ij}^n(x) = w_{ij}^n + \int_0^\infty e^{-sx} d\tilde{w}_{ij}^n(x).$$

Using the dependence structure of $\eta_n$ and $S_n$ we obtain from Eq. (20), with some manipulations,

$$(\lambda-s)\tilde{w}_{ij}^{n+1}(s) = -s\tilde{w}_{ij}^{n+1} + \lambda \sum_{k=1}^N \tilde{w}_{ik}^n(s) [L_{kj}(s) - L_{ik}^0(s) - L_{kj}^0(s)]. \quad (22)$$

Taking the limit $n \to \infty$, and assuming stationarity, as before, (22) goes over to

$$C(s)\tilde{w}(s) = D(s)\tilde{w}, \quad (23)$$

where

$$C_{ij}(s) = (\lambda-s)\delta_{ij} - \lambda L_{ij}(s) \quad \text{or} \quad C(s) = (\lambda-s)I - \lambda L^T(s)$$

$$D_{ij}(s) = -s\delta_{ij} + \lambda [L_{ij}^0(s) - L_{ij}(s)], \quad \text{or} \quad D(s) = -sI + \lambda [L^0(s) - L^T(s)]$$

and $w_j$, the limit of $w_{ij}^n$ is independent of $i$ and equal to $\pi_j$.

Note that $L^T(0)$ is a stochastic matrix with the invariant vector $\pi$.

Equation (23) is of interest to us mainly as the starting point for the evaluation of the expected conditional waiting times

$$v_i = E(n|J=i) = -\frac{d}{ds} \tilde{w}_i(s) \Big|_{s=0}. \quad (25)$$
We present a method of calculation. Higher moments can be calculated by continuation of the procedure below.

Differentiating (23) at $s=0$ yields

\[
\lambda (I - L^T(0)) \hat{v} = [-I + L^{OT'}(0) - \lambda L^T(0)] \hat{e} + [\lambda L^T(0) + I] \hat{e}
\]  

(26)

where $\hat{e}$ is the all ones $N$-vector.

Equation (26) is a singular set of equations for $\hat{v}$. We obtain a more convenient form by adding $\lambda \tilde{p} \hat{v}$, which can be written as $\lambda (\hat{v} \cdot \tilde{p}) \hat{e}$, to both sides and get

\[
\lambda \hat{v} = (I - L^T(0) + \tilde{p})^{-1} ([\lambda \sigma - \lambda \sigma^0 - I] \hat{e} + [I - \lambda \sigma] \hat{e} + \lambda (\hat{v} \cdot \tilde{p}) \hat{e}), \sigma = -L^T(0); \sigma^0 = -L^{OT}(0).
\]  

(27)

Note that $(I - L^T(0) + \tilde{p})^{-1} \hat{e} = \hat{e}$, $\tilde{p} (I - L^T(0) + \tilde{p})^{-1} = \tilde{p}$.

The RHs of Eq. (27) is known, except the last term, $\hat{v} \cdot \tilde{p} \hat{v}$, which we now determine.

To this avail we need the Frobenius' eigenvalue of $L^T(s)$, $\alpha(s)$, and its right and left eigenvectors, $\hat{\alpha}(s)$ and $\hat{\beta}(s)$, respectively.

Thus

\[
[L^T(s) - \alpha(s)I] \hat{\alpha}(s) = \hat{\beta}(s)[L^T(s) - \alpha(s)I] = 0
\]  

(28)

and we may also stipulate, in addition

\[
\hat{\alpha}(s) \cdot \hat{\beta}(s) = \hat{\beta}(s) \cdot \hat{e} = 1.
\]  

(29)

From these and (28) we immediately get

\[
\hat{\alpha}(0) = \hat{e}, \quad \hat{\beta}(0) = \hat{\tilde{p}}, \quad \alpha(0) = 1.
\]  

(30)

Similar to Neuts [N76] we obtain

\[
a'(0) = -\tilde{p} / \lambda, \quad \tilde{p}'(0) = \tilde{p} (\tilde{p} / \lambda - \tilde{p})(I - L^T(0) + \tilde{p})^{-1}
\]

\[
a''(0) = \tilde{p} L^{T'}(0) \hat{e} - 2 \tilde{p}^2 / \lambda + \tilde{p} \tilde{p} (I - L^T(0) + \tilde{p})^{-1} \sigma \hat{e}.
\]  

(31)
Now, multiplying (23) on the left by $\beta(s)$ yields

$$
\beta(s) \cdot \dot{W}(s) = \beta(s)[\lambda L^0 T(s)-(s+\lambda a(s))I] \dot{\lambda} / (\lambda-s-\lambda a(s))
$$

(32)

Differentiating by $s$, and letting $s \to 0$ results in

$$
\ddot{\lambda} = \{[2(1-\rho)\beta'_0(0)-\lambda a''(0)\dot{\rho}][\lambda a-\lambda a' I]-\lambda \beta[L f''(0)-L^0 T''(0)]\dot{\rho} / 2(1-\rho)^2.
$$

(33)

Thus, the expected waiting times can be directly calculated.

4. SPECIALIZATIONS

In this section we examine specific secondary storage devices, applying the results of previous sections. A specialization consists of specifying $T_{ij}$, $T^0_{ij}$, and $K_i$ and their relations with the device parameters.

Drum-like Devices - We consider a drum that comprises $N$ logical sectors. The number of tracks is left unspecified. The time required for the $i$-th sector to pass under the read heads is a constant $\delta_i$; thus the set-up time, called rotational latency here, is given by

$$
t_{ij} = \begin{cases} 
\sum_{k=1}^{j-1} \delta_k & j > i \quad 1 \leq i, j \leq N \\
R - \sum_{k=1}^{j} \delta_k & j \leq i \quad R = \sum_{k=1}^{N} \delta_k 
\end{cases}
$$

(34)

We assume that the physical motion of the drum is the only element that creates delays; i.e. electronic switching times are entirely neglected. A similar device, with a slightly simpler distance structure, is magnetic bubble loop memory [M]. Our drum is midway, in terms of record structure, between a "paging drum" and a "file drum" [FB]. In our discussion
of the drum we also specialize the input process, by assuming that the
types of successive requests are independent (and drawn from a distribu-
tion \( \{p_i\} \)). This is a reasonable assumption for a drum, which is
normally the shared device *par excellence* in a system.

An interesting question in the design, and hence in the analysis,
of such devices is the dependence of capacity, or delays, on the pattern
of use. For the drum as modelled here, it is well known that when re-
quests are processed continuously, which would be the case in our model
when the system is overloaded, the average rotational latency, \( T \), depends
on the distribution \( \{p_i\} \) of relative frequencies, but not on the manner
in which the corresponding records are arranged around the circumference.
Indeed, from (34) we easily have

\[
t_{ij} + t_{ji} = \begin{cases} R - \delta_i - \delta_j & j \neq i \\ 2R - 2\delta_i & j = i \end{cases}
\]

Hence

\[
E(T_{\text{sat}}) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_ip_j t_{ij} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} p_ip_j (t_{ij} + t_{ji})
\]

\[
= R(1 + \sum_{i=1}^{N} p_i^2) / 2 - \sum_{i=1}^{N} p_i^2 \delta_i
\]

(36)

for which the claimed invariance manifestly holds.

We shall show that this property is not retained when idle periods
intervene. We distinguish as in Section 3 between a set-up within a
busy period \( T \) and one that follows an idle period \( T^0 \). From (36)
we have

\[
E(T) = \frac{R}{2} \left( 1 + \sum_{i=1}^{N} p_i^2 \right) - \sum_{i=1}^{N} p_i \delta_i
\]

(37)
To compute $E(T^0)$ consider the following sequence of events, on which we condition our calculation. A request for sector $j$ is completed (and "departs"); no other request is queued for service; a request for sector $i$ arrives and finds the head over sector $M$, at a distance $D$ from its termination. Thus

$$T^0_{j,M,i,D} = D + t_{Mi}.$$  \hfill (38)

The duration $\tau$ between the departure of the request for the $j$-th sector and the arrival of the new one, is distributed exponentially with parameter $\lambda$. We may write

$$P(M = m, D = x) = \sum_{k=0}^{\infty} P(\tau = kR + t_{jm} + \delta_m - x) dx$$

where $k$ is the number of complete revolutions the drum made between the departure and arrival. Using the distribution of $\tau$ we readily obtain

$$P(M = m, D = x) = \lambda e^{-\lambda(t_{jm} + \delta_m - x)} / (1 - e^{-R\lambda}) dx, \quad 0 \leq x < \delta_m; \quad 1 \leq m \leq N. \hfill (39)$$

Using this result in (38) we have

$$E(T^0) = \sum_i \sum_j p_i p_j \sum_m \int_{x=0}^{\delta_m} e^{-\lambda(t_{jm} + \delta_m - x)} \lambda(x+t_{jm}) e^{-\lambda t_{jm}} dx,$$

after integration and rather massive cancellations one obtains

$$E(T^0) = R(1 + \sum_i p_i^2) / 2 - 1 / \lambda - \sum_i p_i \delta_i + R \sum_i \sum_j p_i p_j e^{-\lambda t_{ji}} / (1 - e^{-R\lambda}). \hfill (40)$$

Looking at the last term of (40) we see that $E(T^0)$ clearly depends, as claimed, on the relative arrangement of the records.
Remarks: 1. We assumed here that sector lengths $\delta_i$ may differ. In paging systems this is not necessarily the case. Nevertheless, the dependence expressed in (40) is maintained then as well.

2. Although we did not address ourselves to the problem of finding the optimal arrangement of the records on the drum (i.e. the relative placement of $\delta_i$), which minimizes $E(T^0)$, this problem is of some theoretical interest. We digress here briefly to present a variation on the last model, where the nature of the problem is more evident.

In this variation all sectors have equal length, time is discrete, and arrivals may occur only at those evenly spaced epochs when an inter-sector boundary arrives at a read head. On this time scale, inter-arrival times are distributed geometrically, with a parameter which we denote by $\alpha$; this is the discrete analogue of the exponential distribution. $E(T)$ is still given by (36), with no change, but when we come to evaluate $E(T^0)$ and examine (38), we see that $D$ has no counterpart, because of the discretization of arrival times.

If the calculation of $E(T^0)$ is carried to conclusion, we obtain instead of (40)

$$E(T^0, \text{discrete}) = \frac{(1-\alpha)/(1-\alpha^N)}{\sum_i \sum_j p_i p_j \sum_m t_{mi} a^{tm}}.$$  \hspace{1cm} (41)

Consider the sum in (41). This is a polynomial of degree $N-1$ in $\alpha$. The coefficient of $\alpha^r$ contains various terms that do not depend on the relative order of the records, and the term $\beta_r = N \sum_i \sum_j p_i p_j$, which does depend (the indices $j$ are calculated modulo $N$). Thus, the minimization of $E(T^0)$ here requires solution of $\min_{r=0}^{N-1} \sum \beta_r \alpha^r$. 
3. Obviously, when the traffic intensity increases, idle times become rarer, and the relative arrangement is thus least important just when capacity is most critical. We note that this result does not justify randomly placing records on a drum, since this policy would affect the values of \( p_i \) as well (through the aggregating effects of tracks). As is apparent from (37), (40) and (41), the \( p_i \) do have considerable influence on \( E(T) \), not just \( E(T^0) \).

The maximum traffic intensity that the drum can handle under this regime is immediately given by (37)

\[
\lambda_{\text{max}} < 1/E(S) = 1/[E(T) + \sum_i p_i \delta_i] = 2/R(1 + \sum_i p_i^2).
\]  

(42)

This result shows the way to obtain the distribution \( \{p_i\} \) that results in efficient operation of the system (remember that the \( p_i \) are determined by the records that are placed in each sector, and normally some choice can be exercised in this respect). To this end we only have to find the vector \( \{p_i\} \) that minimizes \( f = \sum_i p_i^2 \), subject to \( \sum p_i = 1 \).

Since \( f \) and the constraint are convex, and \( p_i = 1/N \) is an extremum point, that point must be a global minimum of \( f \). Thus, \( \lambda_{\text{max}} = 2N/R(N+1) \) represents the best performance the system can exhibit. We remark that this optimization problem is "hard" (in fact, NP-complete \([K]\)), as verified in \([CC]\) and \([CW]\). Thus, one must expect an essentially enumerative search for that partition of the set of records such that \( f \) is minimized. (See also \([CC]\) and \([CW]\) for analyses of a simple but very efficient heuristic.)

Finally, we look at the system of equations in (9) and their interpretation in the geometry and dynamics of the drum. We note first that
\[ L_{ji}^{0}(a) = \frac{p_j \exp\{-a(t_{ji} + \delta_i)\}}{z(1-e^{-\lambda R})} \sum_{m=1}^{N} \left\{ e^{-\lambda t_{jm} - a(\delta_j + t_{mi})} - e^{-\lambda t_{jm} - \lambda \delta_j - at_{mi}} \right\} \]

since the service time is constant. The transform of \( S^{0}_{ji} \) is calculated in a way similar to that producing (40), and we obtain at some labor

\[ L_{ji}^{0}(a) \]

Thus, we have from (10)

\[ A_{ij}(z) = z \delta_{ij} - p_j e^{-a(t_{ji} + \delta_i)} \]

and

\[ B_{ij}(z) = p_j e^{-(aR-1)} e^{-a \delta_i - \lambda t_{ji}} \]

In order to use (13) the roots of the equation \(|A(z)| = 0\) are required. The following result is instrumental in obtaining an efficient solution.

**Theorem 2** The determinant of \( A(z) \) (in (45)) can be expressed as

\[ |A(z)| = (b-z^N)/(q-1) \]

where

\[ q = e^{-aR}, \quad b = q \prod_{j=1}^{N} [z-p_j(q-1)]. \]

**Proof.** Consider \(|A(z)|\) as an \( N^{th} \) degree polynomial expression in the term linear in \( z \), with the exponentials regarded as coefficients. At the \( N \) values of \( z \) given by \( z_j = p_j(q-1) \) (when we treat \( q \) as an
explicit coefficient of $z$ rather than display its functional dependence),
the determinant can be easily evaluated, and we obtain $-z_j^N/(q-1)$. The
right-hand side of (47) is a polynomial expression of degree $N$ that
correctly interpolates $|A(z)|$ at the $N+1$ points $z = z_j$ and $z = 1^*$,
and is therefore the unique interpolating polynomial expression of degree $N$.

This is perhaps a somewhat curious result, since the roots of this
equation turn out not to depend to any extent on lengths of individual
records (sectors) but merely on their frequency of use. We have no
intuitive explanation for this phenomenon.

The equation $|A(z)| = 0$ can now be easily solved numerically and
our experience with a straightforward Newton-Raphson iteration procedure
demonstrated very fast convergence and good resolution between the roots
(we only looked inside the unit disk).

**Disk-like Devices** - We consider now the characteristics of a disk pack
(or cartridge), with $N$ cylinders (tracks), and a single arm carrying
the read heads. For the purposes of our analysis this is functionally
identical with any device where the setup time $T_{ij}$ is merely a function
of $|i-j|$, such as magnetic bubble or shift register storage devices.
The following will be in disk terminology.

It is customary to consider the set-up time in disks as composed of
two parts: Seek time, the duration required for the arm to move between
cylinders, and a rational latency similar to the drum.

---

* Obviously $|A(1)| = 0$, since $e_p = 1$; a single application of L'Hospital's
  rule establishes that $z = 1$ is a zero of the right-hand side of (47).
As we consider here a primitive request-queue management technique we also limit all explicit calculations concerning disks in two ways:

- Rotational latency is eliminated by the method of reading, which is to transmit one whole track per request (the portion of the disk passing under a read head during one full revolution); the desired record is subsequently located in memory, and perhaps pieced together from two portions. The latter situation occurs when the requested record was under the read head when the seek terminated and transmission started.

- In these devices (in contrast to the situation in scanning disks) the arm does not react "on the fly" to changes of destination, but rather maintains a "busy" status until a desired seek is terminated and the arm is stopped; only then can a new seek be initiated. Comprehensive discussions of these delays can be found in [W] and [F].

Although the set-up times $T_{ij}$ of bi-directional tapes conform with the above characterization, we exclude them from this discussion on both "practical" and analytical grounds. Firstly, rotational latency is of course absent here; also the tape system can usually handle changes of destination "on the fly" in a much simpler way than in a disk system. (Thus, FIFO is a less natural operating technique for tapes than it is for disks. Even there, however, low processor speeds may require a FIFO regime). Analytically, we find the dependence structure between successive services even more involved than the model presented in Section 2: $T_{ij}$ depends on the boundary of record $i$ where its reading terminated, and this, in turn, involves the even earlier record. We note, though, that if the idle-period policy were of the type denoted by (a) in the following, one
randomization on the identity of that preceding record is enough to properly define the necessary variables.

Unlike the drum, the behavior of the system when no requests are pending may have different modes. The more common ones are (in disk terminology)

a. The arm remains in place, at the cylinder last used.

b. The arm is directed to move to a predetermined "rest place", cylinder r.

These modes determine the distributions of the respective $S_{ij}^0$. In case (a) it is clear that $S_{ij}^0 = S_{ij}$. In case (b), let $f(i,j)$ be the time taken to travel from cylinder i to cylinder j when no intervening cylinders are read. Normally this is the same as the set-up time $T_{ij}$ and we assume so in the following. The set-up time $T_{ij}^0$ succeeding an idle period is then given by

$$T_{ij}^0 = \begin{cases} f(r,j) & s \geq f(i,r) \\ f(i,r) - s + f(r,j) & s < f(i,r) \end{cases}$$

(49)

where $s$ is the length of the idle period. Since $s$ is exponentially distributed, we immediately find

$$E_s(T_{ij}^0) = f(i,r) + f(r,j) - [1-e^{-\lambda f(i,r)}]/\lambda.$$ 

(50)

The expected duration of this delay is calculated as follows. Note first that

$$P_i^0 = P(\text{cylinder i was just read|an idle period just started})$$

$$= \pi_i p_i / \sum_k \pi_k p_k$$

(51)
where \( \pi_i \), as defined earlier, is the probability that the request queue is empty following the completion of service from cylinder \( i \). Thus we obtain
\[
E(T^0) = \sum_i \sum_j p_i^0 p_j^0 E(T^0_{ij})
\]
\[
= \sum_j p_j f(r,j) + \sum_i \pi_i p_i f(i,r) e^{-\lambda f(i,r)}/(\lambda \sum_k \pi_k p_k) - 1/\lambda .
\] (52)

The value of \( E(T^0) \) does not influence the overall service capacity of the system. It is a factor in its response when not fully loaded. In fact, it becomes more important as the load becomes lighter.

Unlike the drum which is a constant speed system, we have here important acceleration and deceleration effects. An approximation that holds for a rather large subset of available disks is
\[
T_{ij} = f(i,j) = \begin{cases} 
0 & i = j \\
A + B|i-j| & i \neq j,
\end{cases}
\] (53)

where \( A \) "summarizes" the effects of the changes of speed of motion of the arm and \( B \) corresponds to movement in constant speed. (The approximation is not very good for short distances and quite acceptable when a sizeable portion of the disk radius has to be traversed.) This completes the specification, so that the procedures of Section 3 can be applied. (Nothing comparable to Theorem 2 was found here, however.)
5. ALTERNATIVE PROCEDURE TO CALCULATE BOUNDARY PROBABILITIES

In this section we describe a general method to evaluate the boundary probabilities $\pi_j$ defined below equation (8). The analytical method given in Section 3, that depends on Theorem 1 is devoid of probabilistic-physical content. This makes any numerical idiosyncracies occurring in its implementation hard to interpret, and thus instabilities are not easy to remove, even for moderate values of $N$. We present an approach parallel to the one in [N76]*; here all the steps and interim results have intuitive meaning, and thus error control is materially simplified. Excepting values of $\lambda$ 'close to $\lambda_{\text{max}}$ this method would also be cheaper.

We call upon a familiar result: in a recurrent Markov chain, the invariant probability of a state (its steady-state probability) is equal to the inverse of its recurrence time [H, p.195].

Consider then the embedded chain, formed of the $M$ states $(0,j)$, obtained at departure epochs. Its steady-state probabilities were called $p(0,j) = \tilde{p}_j \pi_j$, and its recurrence times are given by

$$
\sum_{i=1}^{N} \ell_i \mu_i^*/\ell_j,
$$

where $\tilde{\pi}$ is the invariant probability vector of the matrix $\ell$, defined as follows:

$$
\ell_{ij}(k,x) = \text{Prob (a busy period that starts at $(0,i)$ terminates at $(0,j)$, following $k$ services and requiring up to $x$)}
$$

The only essential difference between our model and the problem treated in [N76], is that we must ascribe an extraordinary distribution to the service duration that initiates a busy cycle. We note that in [N76] batch arrivals are treated, though.
and

\[ L = \sum_{x=0}^{\infty} \sum_{k=1}^{\infty} dL(k,x). \]  

(56)

The quantity \( \mu_i^* \) is the expected number of service completions in a 
busy period that started in state \((0,i)\). If \( L(z,s) \) is the LST-pgf 
of \( L(k,x) \), we have

\[ \frac{\partial L(z,s)}{\partial z} \bigg|_{z=1} = e^i \cdot (1, \ldots, 1). \]

(57)

We proceed to derive \( \tilde{c} \) and \( \hat{\mu}^* \).

Define now the matrices \( \tilde{c}(\cdot) \) and \( \tilde{c}^0(\cdot) \)

\[
\tilde{c}_{ij}(x) = p_{ij}F_{ij}(x) \\
\tilde{c}^0_{ij}(x) = p_{ij}F^0_{ij}(x)
\]

(58)

and the first-passage measure

\[ \tilde{G}_{ij}(k,x) = \text{Prob}(\text{A first transition of the system from a state} \) 
\((n,i) \text{ to a state where } X = n-1, \text{ will be to the state} \) 
\((n-1,j), \text{ will involve } k \text{ services and terminate} \) 
\text{within } x). \]

(59)

The interpretations of \( \tilde{c}(\cdot) \) and \( \tilde{c}^0(\cdot) \) are obvious. \( \tilde{G}(\cdot,\cdot) \) is also called 
"down level-crossing" distribution.

We further define the matrices

\[
\tilde{c}_v(x) = \int_{t=0}^{\infty} p(v,t)d\tilde{c}(t) \quad v \geq 0 \\
\tilde{c}^0_v(x) = \int_{t=0}^{\infty} p(v,t)d\tilde{c}^0(t) \quad v \geq 0
\]

(60)

where \( p(v,t) \) is the probability of exactly \( v \) arrivals within \( t \).

Forming the LST's of \( \tilde{c}_v(\cdot), \tilde{c}^0_v(\cdot) \) and \( \tilde{G}(\cdot,\cdot) \) (the latter is also
a pgf), denoted respectively by $c_v(s), c_0^v(s)$, and $G(z,s)$ we have, following [N76] from renewal considerations,

$$G(z,s) = \sum_{v=0}^{\infty} zc_v(s)G^v(z,s) \quad |z| < 1, \ Re \ s \geq 0. \quad (61)$$

For $z = 1$, $s = 0$, $G(1,0) \equiv G$ is a stochastic matrix, with invariant probability vector $\vec{g}$, from which we form $\vec{G}$, as $\vec{P}$ was defined.

The same consideration that led to (61) can be applied to $L(z,s)$, and we obtain

$$L(z,s) = \sum_{v=0}^{\infty} zc_0^v(s)G^v(z,s) \quad |z| < 1, \ Re \ s > 0. \quad (62)$$

The equation (61), at $z = 1$, $s = 0$ can be iteratively solved* whence $L$ and $\vec{L}$ are obtained quite painlessly. $L$ is the probability transition matrix of our embedded chain.

We need yet $\vec{\mu}$. To this avail we note the following result, given in [N76]:

$$\vec{\mu} \equiv \frac{\partial G}{\partial z} \bigg|_{z=1} \vec{e} = (I-G+\vec{G})(I-P+\vec{G}+\lambda \text{diag}(\vec{G})\vec{G})^{-1}\vec{e}$$

$$= (I-G+\vec{G})\vec{a}, \quad (63)$$

where $\sigma_i = \sum_j p_{ij}E(S_{ij})$. Thus,

* e.g. by the sequence

$$G_1 = (I-c_1)^{-1}$$

$$G_{k+1} = (I-c_1)^{-1}\{c_0 + \sum_{v=2}^{\infty} c_v G^v_k\} \quad k \geq 1,$$

which converges quite well, normally.
\[ \mu^* = \frac{\partial L}{\partial z} \bigg|_{z=1}^{s=0} e = [\sum_{v=0}^{\infty} C_v^0 G^v + \sum_{v=0}^{\infty} \sum_{i=0}^{v-1} G^i M Z G^{v-i-1}]e \]  \hspace{2cm} (64)

where \( M_z = \frac{\partial G(z,s)}{\partial z} \), \( z=1, s=0 \);

since \( G \) is stochastic we get

\[ \mu^* = 1 + \sum_{v=0}^{\infty} C_v^0 \sum_{i=0}^{v-1} G^i \mu = 1 + \sum_{v=0}^{\infty} \sum_{i=0}^{v-1} G^i (I - G + G) \mu \]

\[ = 1 + [P - L + \lambda \text{diag}(G^0) G] (I - G + G)^{-1} \mu, \]  \hspace{2cm} (65)

which can be readily calculated.

Reference [N76] contains further results that are of interest, and can be applied - mutatis mutandis - to our model. Expressions for mean queue lengths were derived, to be used as a check on equation (19). As they are rather involved we do not present them here. We note, however, that the two procedures pose numerical problems of entirely different nature and the investigation of their respective behavior modes, particularly in extreme situations (-very light or very heavy traffic, large \( N \), etc.) is of great interest.

6. DISCUSSION

We have shown in the preceding sections a method of analyzing system models that although they are simple to describe in queueing-theoretical terminology, they yet display features that render standard methods ineffective in tackling them. The factor that particularly exacerbates the work is the dependence between successive services; put another way - the time required to service a set of requests depends on the way we order them. Rarely, if ever, will FIFO prove the most efficient
service method, though we can very well imagine situations where its simplicity of implementation would outweigh other considerations.

In contrast with the foregoing analysis we mention a rather prevalent approach to the same situation which is often found in the literature (cf. [Wi] for a recent example). The approach we refer to consists of evaluating a distribution function for the duration required to service a request by averaging over request types [essentially, writing \( P(S \leq t) = \sum_{i,j} p_i p_{ij} P(S_{ij} \leq t) \)], and substituting the result within formulas derived in the standard analysis of an M/G/1 queueing system, which explicitly assumes (and utilizes) independence between successive services. This can often lead to gross misestimation of the evaluated quantities.
APPENDIX

Defining $\sigma_{ij} = E(S_{ij})$, $\sigma_{ij}^{(2)} = E(S_{ij})^{2}$, similarity with superscript 0, and letting $\tilde{M}_{ij}$ be $\tilde{p}_{i}p_{j}M_{ij}/\tilde{p}_{j}$ where $M$ is any of $\sigma$, $\sigma_{0}$, $\sigma^{(2)}$, $\sigma_{0}^{(2)}$, we find

$$B_{1}^{T} = \lambda \tilde{\sigma} - \lambda \sigma_{0} - I,$$
$$B_{2}^{T} = \frac{1}{2}[\lambda^{2} \sigma_{0}^{(2)} + 2\lambda \sigma_{0} - \lambda^{2} \sigma^{(2)}].$$

The quantities $\delta_{1}$, $\delta_{2}$ and the matrices $C_{0}$ and $C_{1}$ have generally to be directly evaluated by differentiating $|A(z)|$ and the relation

$$|A(z)| = A(z)C(z).$$

For the special case $P = P$ (independent references) closed expressions can be readily found:

$$\delta_{1} = 1 - \rho,$$
$$\delta_{2} = N - 1 - \lambda \sum_{j} p_{j} \sigma_{jj} - (N-2)\rho - \frac{1}{2} \lambda^{2} E(s^{2}) + \lambda^{2} \sum_{j<k} p_{j}p_{k}(\sigma_{jj}\sigma_{kk} - \sigma_{jk}\sigma_{kj})$$
$$- \lambda^{2} \sum_{j<k} \sum_{l<k} p_{j}p_{k}p_{l}(jkl)^{\sigma},$$

where

$$\sigma(jkl)^{\sigma} = \sigma_{jj}\sigma_{kk} + \sigma_{jj}\sigma_{ll} + \sigma_{kk}\sigma_{ll} + \sigma_{kj}\sigma_{kl} + \sigma_{kj}\sigma_{jl} + \sigma_{lk}\sigma_{jl} + \sigma_{jl}\sigma_{lk} + \sigma_{jl}\sigma_{kj} + \sigma_{kl}\sigma_{lj} + \sigma_{kl}\sigma_{jk} + \sigma_{jk}\sigma_{kl} - \sigma_{jk}\sigma_{kj} - \sigma_{kl}\sigma_{lj} - \sigma_{kl}\sigma_{jk} - \sigma_{kj}\sigma_{kl} - \sigma_{kl}\sigma_{lj} - \sigma_{kl}\sigma_{kj} - \sigma_{kj}\sigma_{kl} - \sigma_{kl}\sigma_{lj}.$$
REFERENCES


