AN EXACT ANALYSIS OF CUSTOMER LOSS UNDER MINIMUM LAXITY SCHEDULING IN DISCRETE-TIME QUEUEING SYSTEMS

James F. Kurose
Department of Computer and Information Science
University of Massachusetts
Amherst, MA 01003

COINS Technical Report 89-94

An Exact Analysis of Customer Loss Under Minimum Laxity Scheduling in Discrete-Time Queueing Systems[1]

James F. Kurose
Department of Computer and Information Science
University of Massachusetts
Amherst, MA 01003

Abstract

We consider a discrete-time queueing system in which a deadline is associated with each arriving customer. A maximum possible deadline of M is assumed and each customer must begin service before its deadline expires; otherwise the customer is considered lost and leaves the queue without receiving service. Customers in the queue are scheduled according to a non-preemptive *minimum laxity scheduling discipline* in which that customer whose deadline is closest to expiring is selected for service. The main result of this paper is a numerical algorithm which exactly computes customer loss for this queueing system with a time complexity of $O(M^4)$ for the case case of geometrically distributed service times and a bulk arrival process in which the number of customers arriving in a slot with a deadline of i slots is also geometrically distributed (for each $i, 1 \leq i \leq M$). We also demonstrate how this model can be extended to include generally distributed service times and laxities as well.

[1]This work was supported in part by the Office of Naval Research under Contract N00014-87-K-0796 and an equipment grant from the National Science Foundation, CER-DCR-8500332.

1. Introduction

In real-time computer and communication systems, temporal constraints are placed on the behaviour of the agents (e.g., processes or messages) within these systems. Typically, these constraints require that these agents initiate or complete some task (e.g., a process' computation or a message's transmission) within some *deadline*. In "hard" real-time systems, the system is engineered in such a manner that these constraints are never violated. In "soft" real-time systems, these constraints can occasionally be violated but such violations typically translate into poorer system performance; also the results of such overly-delayed computations or message transmissions are often of little (or no) use to the system. Examples of computer systems exhibiting such soft real-time behaviour include applications in distributed systems for sensor applications and autonomous manufacturing. In the communication network domain, real-time constraints often result from the need to bound delays involved in interactive human voice (and video) communication [17,22,26,5,4] or as a form of overload protection [8].

In such soft real-time systems, the performance metric of interest is no longer one of the traditional measures such as average delay or throughput, but rather the fraction of jobs which are not able to meet their specified time constraints (i.e., the fraction of jobs that are *lost*.) As a consequence, an important goal of performance models for these real-time systems is the characterization of this loss. In this paper, we consider a discrete-time queueing system in which the *deadline* associated with each customer (the amount of time from a customer's arrival until the time at which it must begin service) is bounded by some maximum possible value, M. We note that adopting a discrete-time model is particularly appropriate in the communication networks area, given the synchronous nature of many proposed high-speed network switches currently under design [25,24,6]. Customers in the queue are scheduled according to a *minimum laxity scheduling discipline* [11,20,21,2,14], in which that customer whose deadline is closest to expiring is selected for service; a customer whose deadline expires is considered lost and is removed from the queue without receiving service. We analyze the case of geometrically distributed service times and a bulk arrival process in which the number of customers arriving in a slot with a deadline of i slots is also geometrically distributed (for each $i, 1 \leq i \leq M$); we also demonstrate how this model can be extended to include generally distributed service times and laxities as well. The main result of this paper is a numerical algorithm which exactly computes customer loss for this queueing system with a time complexity of $O(M^4)$.

The analysis of the minimum laxity scheduling policy is of considerable interest since it has been shown to be optimal (in the sense of minimizing customer loss) over all non-preemptive work-conserving policies for the continuous-time $G/M/c + G$ queue (the last G here indicates that the deadlines are drawn from some general distribution) and its discrete-time analog [21]; related results have also appeared in [20,23,2]. To date, almost all performance analyses of queueing systems with impatient customers have assumed FCFS scheduling [1,8,9,10,13,18,22,26] or some other discipline which does not explicitly consider the customers' time constraints; a survey of this work can be found in [19]. *Approximate* analysis of customer loss under the minimum laxity scheduling discipline in $M/M/c + M$ systems have appeared in [27,14]. In these approximations, the customer to be served

is chosen from a group of n queued customers, and arriving customers move into the group of n on an FCFS basis. In contrast, the analysis in this paper is exact and for a more general discrete-time queueing system. One other related analysis is [11] (extended in [12]), which analyzes an earliest-due-date (EDD) scheduling rule [4,16], but in a queueing system in which late customers receive service (i.e., there is no loss).

The remainder of this paper is structured as follows. Section 2 describes the queueing model considered in this paper and discusses the structure of the minimal laxity sample path to be exploited in the analysis. Sections 3.1 and 3.2 derive recursive relationships among various random variables associated with the minimum laxity sample path. These relationships form the basis of the $O(M^4)$ algorithm for computing customer loss, as described in sections 3.3 through 3.5. The derivations in section 3 assume geometrically distributed service times and geometrically distributed bulk arrivals; section 4 discusses extensions to this basic model. Finally, section 5 summarizes this paper.

2. Model

We consider a time-slotted queueing system in which each arriving customer (job) has an associated deadline. If a customer does not *enter service* before this deadline expires, it is considered lost and leaves the queue without receiving service. We refer to the *laxity* of a queued customer as the amount of remaining time it can spend in the queue before being lost. Note that a customer's laxity decreases at the rate of one (slot per slot), as it waits in the queue.

In the basic model, we assume a single server and adopt the following assumptions about the arrival/laxity process, service time distribution, and scheduling discipline; extensions to the basic model are discussed in section 5. The unit of time throughout the remainder of the paper is the unit slot length.

- **Arrival/Laxity Process:** We define M as the maximum possible initial laxity and model the number of customers that arrive in a slot with an initial laxity of i slots ($i < M$) as a geometrically distributed random variable with mean λ_i. All arrivals are assumed to occur at the very beginning of a slot and an arrival at the beginning of a slot can potentially receive a full slot's worth of service during that slot.

  The number of arrivals in a slot with a laxity of i is also assumed to be independent from one slot to the next and also independent of the number of arrivals in a slot with a laxity of j, $j \neq i$.

- **Service Time Distribution:** The amount of service required by a customer is geometrically distributed with mean $1/\mu$. We assume that a job's service time is not known when it enters service.

- **Scheduling Discipline:** When the server selects a customer for service, that customer with the smallest laxity is chosen. Scheduling is non-preemptive. If several customers have the same

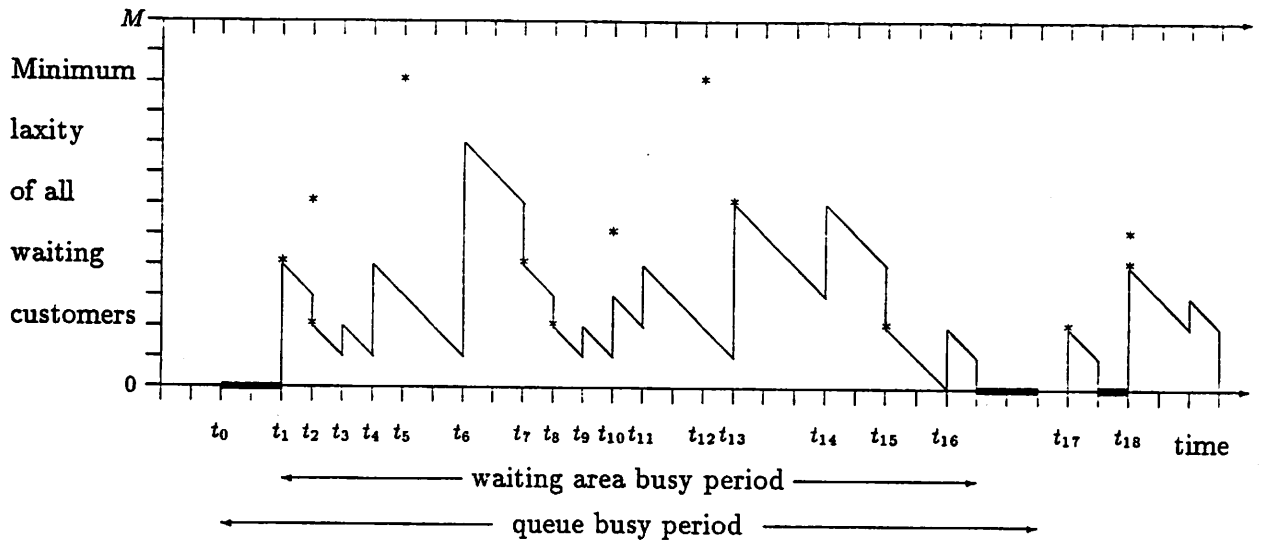

Figure 1: Minimum laxity sample path

minimal laxity, a minimal laxity customer that has been in the queue the longest is chosen (although the results presented below hold for other policies for breaking ties as well). This scheduling policy is variously known as a *minimum laxity,* shortest deadline first, or shortest time to extinction policy [20,23,2].

We note that the above model permits the performance of a wide range of real-time slotted systems to be modeled, including the important case of multiple classes of traffic, each with different (fixed) deadlines.

Figure 1 shows a sample path of the minimum laxity of all *queued* customers (i.e., customers waiting to enter service) as a function of time. Customer arrivals are shown as asterisks and intervals of time during which there is one customer in the queue (one in service, but none waiting) are indicated by bold lines along the x-axis. In the following we will define the *waiting area busy period* as an interval of time during which there is one or more queued (waiting for service) customers during a slot. The busy period of the queue as a whole is defined in the standard manner; the queue busy period may thus contain zero, one, or two or more waiting area busy periods.

The jumps in the minimum laxity at the beginning of a slot in Figure 1 result from either customer arrivals with a new minimum laxity (in which case the minimum laxity decreases) or customer departures (in which case the customer with the lowest laxity enters service and hence the customer with the next smallest laxity becomes the minimum laxity customer, thus resulting in a non-negative increase in minimum laxity). For example, the customer arriving t_1 with an initial laxity of 4 finds the server busy (with the customer that entered service at t_0) and becomes the

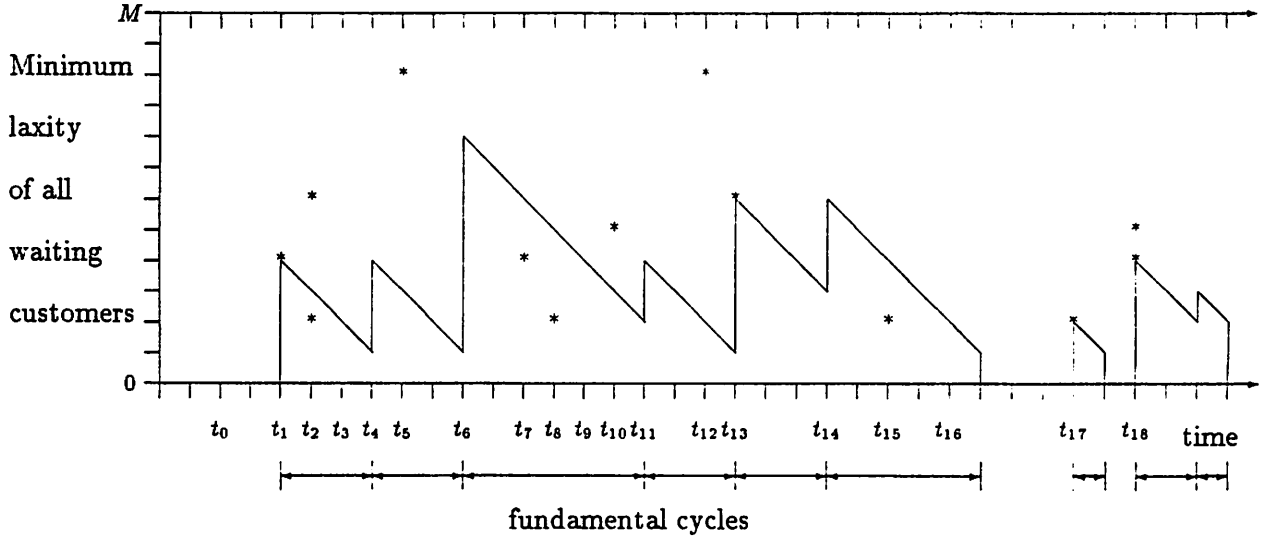

Figure 2: The laxity of customers generating fundamental cycles

minimum laxity customer. At the beginning of slot t_2 this customer's laxity has decreased to 3, but another customer has arrived with a laxity of 2. Note that the second customer arriving at t_2 (with a laxity of 6) does not affect the minimum laxity value at t_2. At t_3 the customer in service leaves the queue and the queued minimum laxity customer (the one that arrived at t_2 with the smaller laxity) enters service. At this point, the new minimum laxity customer is again the customer that arrived at t_1; note that its laxity has decreased to 2 by the beginning of slot t_3. At t_4 the customer that entered service at t_3 completes, the customer that arrived at t_1 enters service, and the customer that arrived with the larger laxity at t_2 becomes the minimum laxity customer - with a laxity of 4 at t_4. Note that at t_{16}, the customer that arrived at t_{15} is lost since its laxity becomes zero (assuming it did not enter service at t_{16}).

Given such system evolution, we are faced with the problem of determining the fraction of customers lost due to the fact that they do not enter service before their laxity becomes zero. We conjecture that any modeling approach based on simply tracking the minimum laxity from one slot to the next is not viable, since the future evolution of the system will be based on the individual laxities of *all* queued customers, not just the minimum laxity customer (and the laxities of all queued customers can not be recovered from knowing simply the laxity of the minimum laxity customers). An exact approach based on explicitly tracking the laxity of all queued customers has a prohibitively large state space of size M^M. In [14], the state space is truncated by assuming that the customer to be served is chosen from a group of n queued customers, and arriving customers move into the group of n on an FCFS basis.

Our approach is based on the fact that the system does nonetheless possess a Markovian structure which can exploited. This structure is based on what we will term *fundamental cycles* in the minimum laxity sample path. We define a fundamental cycle as follows. Consider a customer which upon arrival finds that either *(a)* there are no other customers waiting but there is a job in service or *(b)* there are other customers waiting with a smaller laxity. A fundamental cycle is defined as the time from which such a customer first becomes the minimum laxity customer until it either enters service or leaves the queue due to an expired deadline. Figure 2 shows the fundamental cycles associated with the minimum laxity sample path of Figure 1; also plotted is the laxity associated with the customers generating the fundamental cycles. The key property of a fundamental cycle to be exploited is the fact that conditioned on it beginning more than M slots after the start of a waiting area busy period, the statistical properties of a fundamental cycle starting at height i are *independent* of the behavior of all preceding fundamental cycles. For example, the statistics of the fundamental cycle beginning at t_{11} in Figures 1 and 2 are independent of the statistics of the fundamental cycle beginning at t_6.

The analysis in the following section is based on first finding the expected length of a fundamental cycle starting a height i, computing the expected lengths of the waiting area busy period and queue busy period, and then computing customer loss.

3. An Algorithm for Computing Customer Loss

3.1 Distinguished Customers

We begin by examining the amount of time that a special type of customer, which we refer to as a distinguished customer, spends waiting in the queue until either it enters service or its deadline expires. A *distinguished customer* is defined as a customer which, when it arrives to the system, finds the server busy and also finds that it has the (strictly) minimum laxity of all queued customers. If several customers arrive in a bulk with the same minimal laxity, only one of these customers is considered a distinguished arrival.

Figure 3 shows a sample path of the minimum laxity of all *queued* customers (i.e., customers waiting to enter service) as a function of time; customer arrivals are again shown as asterisks. Customer 0 arrives at time t_0 with a laxity of l_0, finds the server busy, and is queued. Since there are no other queued customers at time t_0, this customer is distinguished. In the following slot there are neither arrivals nor service completions, and hence the minimum laxity of the queued jobs (in this case the single job) decreases by 1. Customer 1, arriving at time t_1 with a laxity of l_1, is not distinguished, since its laxity is greater than that of customer 0. At t_2 customer 2 arrives with a laxity of l_2; it is a distinguished arrival since its laxity at time t_2 is less than that of customers 0 and 1. At time t_3, the customer that was already in service at t_0 completes service and the customer with the minimum laxity (customer 2) enters service. Note that the minimum laxity (over all queued customers) at time t_3 is l'_0, the laxity of the customer 0. At time t_4, customer 2 completes service, customer 0 enters service, and the minimum laxity is now l'_1 - that of customer 1. Customer 1 finally enters service at t_5.

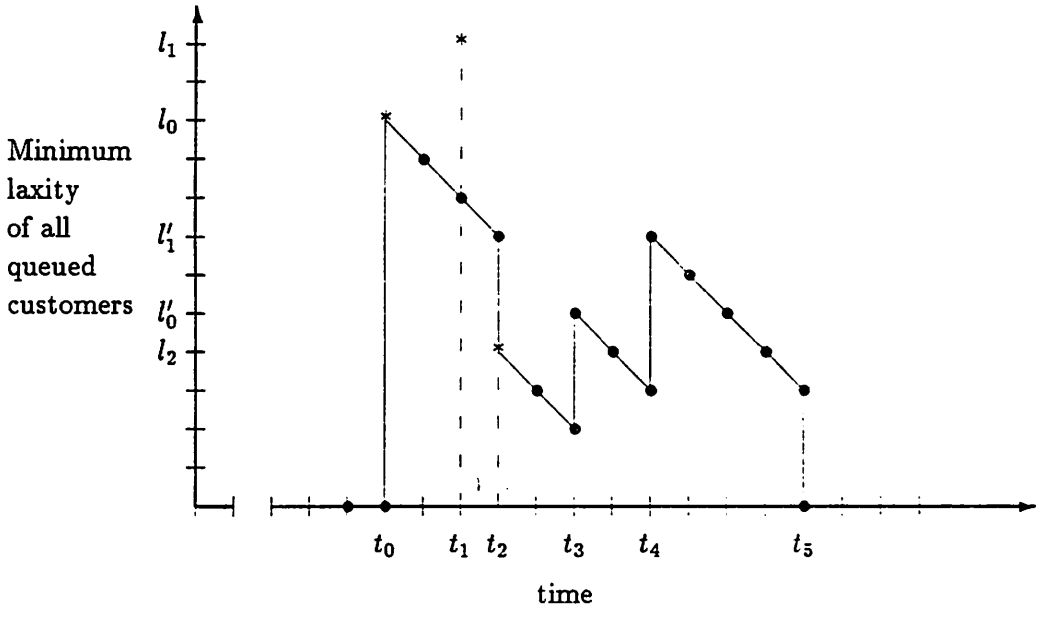

Figure 3: The laxity of queued customers as a function of time

Let us define:

X_i as a random variable indicating the number of slots between the arrival of a distinguished customer with an initial laxity of i, and the time at which this distinguished customer either enters service or leaves the system due to an expired deadline. Note that X_i consists of slots needed to complete the job which is either currently in service or enters service (see below) when the distinguished customer is first queued, plus any slots needed to serve future arrivals (occurring before the distinguished customer enters service) which have an initial laxity less than the current laxity of the distinguished customer.

$X_i(k)$ as the probability of the event $X_i = k$.

μ as the probability that a customer in service completes service at the end of the current slot. Given geometric service times, the mean service time is $1/\mu$.

$\pi_{l,j}$ as the probability of j (new) arrivals with laxity l in a slot; for each l, $\sum_{j=0}^{\infty} \pi_{l,j} = 1$.

$\alpha_0, \alpha_1, \alpha_{2+}$ as the probabilities of zero, one or two or more arrivals in a slot (regardless of laxity), respectively.

p_l^0 as the probability of no new arrivals with a laxity less than or equal to l in a slot. $p_l^0 = \prod_{l'=1}^{l} \pi_{l',0}$ for $l > 0$ and $p_0^0 = 1$ by definition.

p_l^1 as the probability of exactly one new arrival with a laxity less than or equal to l in a slot. $p_l^1 = \sum_{j=1}^{l} \pi_{j,1} \prod_{l'=1 \neq j}^{l} \pi_{l',0}$ for $l \geq 1$.

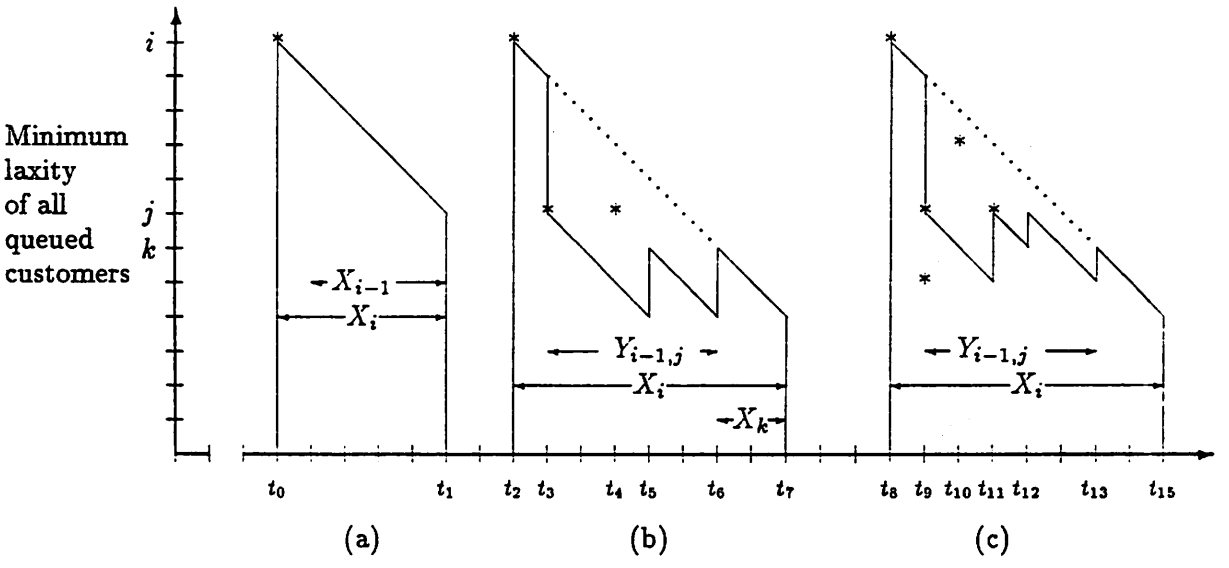

Figure 4: Case analysis for $X_i(k)$

σ_l^1 as the probability of one or more customers arriving in a slot, with the minimum laxity over all arriving customers being l. $\sigma_l = (1 - \pi_{l,0})p_{l-1}^0$, $1 \leq l \leq M$.

σ_l^2 as the probability of two or more customers arriving in a slot, with the second smallest laxity over all arriving customers being l. For $1 \leq l \leq M$,

$$\sigma_l^2 = p_{l-1}^0(1 - \pi_{l,0} - \pi_{l,1}) + (1 - \delta_{l,1})(1 - \pi_{l,0}) \sum_{j=1}^{l-1} \pi_{j,1} \prod_{k=1 \neq j}^{l-1} \pi_{k,0}.$$

Given the above definitions, an expression for $X_i(k)$ can be easily obtained by conditioning on the events that can occur at the end of the slot following the slot in which the distinguished customer arrives:

$$\begin{aligned} X_1(k) &= \delta_{k,1} \qquad \forall k \\ X_i(0) &= 0 \qquad \forall i \\ X_i(k) &= \mu p_{i-2}^0 \delta_{k,1} + \left((1-\mu)p_{i-2}^0 + \mu p_{i-2}^1\right) X_{i-1}(k-1) \\ &\quad + \textstyle\sum_{m=1}^{i-2}((1-\mu)\sigma_m^1 + \mu\sigma_m^2) \sum_{n=1}^{i-2} Y_{i-1,m}(n) X_{i-1-n}(k-1-n) \\ &\quad 1 < i \leq M; \quad 1 \leq k \leq i \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{1}$$

The first term on the right hand side of equation 1 for $X_i(k)$ results from the case in which the customer in service completes service at the end of the slot and there are no new distinguished

arrivals which begin service. In this case, the distinguished customer enters service and the value of X_i will be 1.

The second case is when the distinguished customer is again the minimum laxity customer at the beginning of the following slot and does not enter service; this is shown in figure 4a. This event can occur if either the customer in service does not complete and there are no new distinguished arrivals, or if the customer in service does complete and there is exactly one new distinguished arrival (which itself enters service); these events occur with probabilities $(1-\mu)p^0_{i-2}$ and μp^1_{i-2}, respectively. In these cases, since the distinguished customer has already waited one slot, the probability that it leaves the waiting area k slots after its arrival is the same as if it had arrived in the subsequent slot, with an initial laxity of $i-1$, and left the waiting area after $k-1$ slots.

In the third case, a new minimum laxity customer is present at the beginning of the next slot. This can happen if the job in service does not complete but a new distinguished customer with a laxity of $m < i-1$ arrives (this occurs with probability $(1-\mu)\sigma_m$). In this case the original distinguished customer will not leave the waiting area until the job in service, the new distinguished customer, and all future arrivals which have an initial laxity smaller than the laxity of the original distinguished customer all leave the queue. This is shown in figure 4b. In this example, the distinguished customer arriving at t_3 goes into service at t_5 (at which point the undistinguished customer arriving at t_4 becomes the minimum laxity customer), the arrival at t_4 goes into service at t_6 (at which point the original distinguished customer is again the minimum laxity customer) and the original distinguished customer goes into service at t_7.

If the job in service *does* complete at the end of the slot and two or more new customers arrive with an initial laxity less than the laxity of the original distinguished customer (this occurs with probability $\mu\sigma_{m,i}$), a new minimal laxity customer will again be present during the slot. The arriving customer with the smallest laxity enters service, and the newly arriving customer with the second smallest laxity behaves as a distinguished arrival. The original distinguished customer will not leave the waiting area until the new distinguished customer as well as all future arrivals which have a smaller laxity leave the waiting area. This is shown in figure 4c. The distinguished customer arriving at t_8 goes into service at t_{15}, the arrival at t_9 with the minimum laxity enters service immediately and the distinguished arrival at t_9 (with an initial laxity of j) goes into service at t_{11}.

The $Y_{i-1,m}(n)$ term in equation 1 and the random variable $Y_{i,j}$ in figure 4 are defined as follows.

$Y_{i,j}$ Suppose that at the beginning of a slot (before new arrivals are considered) customer w is the minimum laxity customer, with a laxity of i. Suppose further that when new arrivals in this slot are considered, the minimum laxity decreases from i to j (i.e., there is a distinguished arrival with a laxity of j, e.g., as at times t_3 and t_9 in figure 4). We define $Y_{i,j}$ as the number of slots that elapse until customer w again becomes the minimum laxity customer.

$Y_{i,j}(k)$ is defined as the probability that $Y_{i,j} = k$.

The structure of equation 1 suggests a recursive procedure to compute $X_i(*)$ in terms of $X_{i-j}(*)$, $j \geq$

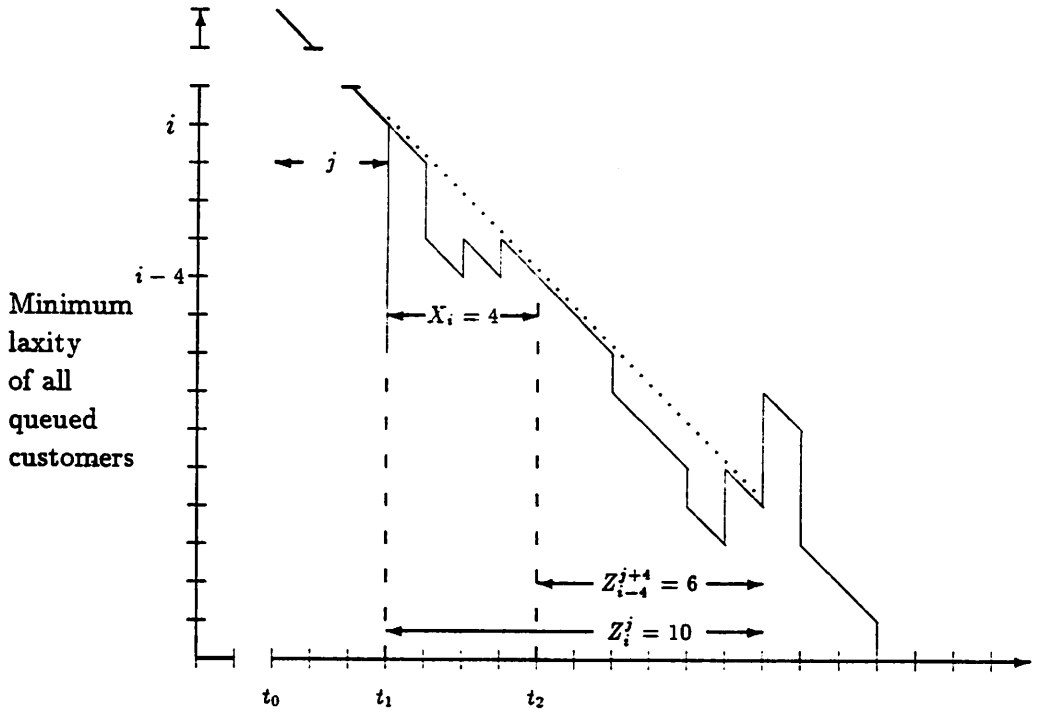

Figure 5: A sample path for Z_i^j

1 and $Y_{i-1,*}(*)$. In the following section, we discuss the calculation of $Y_{i-1,*}(*)$. We note that as a result of the double summation in equation 1, the complexity of computing $X_i(k)$ for a given i and k is $O(M^2)$, where M is the maximum possible laxity. Hence the overall complexity of computing $X_i(k)$ for all i and k is $O(M^4)$.

3.2 Computing $Y_{i,j}(k)$

Before deriving an expression for $Y_{i,j}(k)$, we will need to to derive an expression for one other related quantity. Suppose that at the beginning of some slot t_1, all customers with a current laxity of i (if any) would be the minimum laxity customers and that these customers may have arrived in any of the j slots preceding t_1, where j includes the current slot beginning at t_1 (i.e., $j = 1$ is the case in which only arrivals at the beginning of the current slot are considered). As shown in figure 5, these customers must have arrived at the beginning of a slot with an initial laxity falling along the boldface line of slope -1 in the closed interval $[t_0, t_1]$. We define:

Z_i^j Given that the minimum laxity customers have a laxity of i and given that these customers may have arrived in any of the preceding j slots (where the current slot is counted in j), Z_i^j is the number of future *consecutive* slots for which the minimum laxity over all queued customers (including future arrivals) is (a) non-zero during the slot and (b) at the beginning of the wth slot in the future, this minimum laxity is less than or equal to $i - w$. In graphical terms (see

figure 5), the minimum laxity must be bounded above by a diagonal line of slope -1 beginning at height i. For example, the dotted line beginning at height i at slot t_1 (and ending at t_{13}) is the diagonal line associated with Z_i^j at time t_1.

$Z_i^j(k)$ The probability that $Z_i^j = k$.

Given these definitions, we have for $i \geq 1$:

$$Z_i^j(k) = \begin{cases} \pi_{i+j-1,0} Z_i^{j-1}(k) + (1 - \pi_{i+j-1,0}) \sum_{n=1}^{k} X_i(n) Z_{i-n}^{n+j}(k-n) & 1 \leq j \leq M-i+1; \\ & 0 \leq k \leq i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

and $Z_0^j(0) = 1$, $Z_j^0(0) = 1$, and $Z_j^0(k|k \neq 0) = 0$ by definition.

Equation 2 may be interpreted as follows. With probability $\pi_{i+j-1,0}$ none of the currently queued customers with a (minimal) laxity of i arrived j slots in the past. Conditioned on this event, the probability that $Z_i^j = k$ is simply $Z_i^{j-1}(k)$, the probability that $Z_i^{j-1} = k$. If there is a queued customer with a (minimal) laxity of i which arrived j slots in the past, this customer leaves the waiting area n slots in the future (i.e., $X_i = n$) with probability $X_i(n)$. At this future point in time, we must have the event $Z_{i-n}^{j+n} = k - n$ in order for the original event $Z_i^j = k$ to occur. Figure 5 shows the case that $X_i = 4$ (i.e., the oldest minimal laxity customer goes into service at t_2 and $Z_{i-4}^{j+4} = 6$, yielding the event $Z_i^j = 10$).

For a given i, assuming that $X_i(k)$ are known for all k, equation 2 provides a recursive formulation for computing $Z_i^j(k)$. Due to the single summation, the complexity of computing $Z_i^j(k)$ for all i, j and k is $O(M^4)$.

We are now finally in the position to derive an expression for $Y_{i,j}(k)$. Recall that $Y_{i,j}(k)$ is the probability that $Y_{i,j} = k$, where $Y_{i,j}$ is the number of slots that elapse until a customer that has a minimal laxity of i (before new arrivals in a slot are considered) again becomes the minimum laxity customer, given that when new arrivals in the slot are considered, the customer is no longer the minimal laxity customer (i.e., there is a distinguished arrival with laxity $j < i$). Informally, $Y_{i,j}$ corresponds to the number of slots strictly below the dotted lines in figures 4a, 4b, and 6.

Consider first Figure 6a. In this case, the laxity of the new minimum laxity arrived customer at t_0 is one slot less than the previous minimum laxity and

$$\begin{aligned} Y_{i,i-1}(k) &= \begin{cases} \sum_{n=1}^{k} X_{i-1}(n) Z_{i-1-n}^{1+n}(k-n) & 3 \leq i \leq M; \ 1 \leq k \leq l-1 \\ 0 & \text{otherwise} \end{cases} \\ Y_{2,1}(k) &= \delta_{k,1} \end{aligned} \quad (3)$$

Consider next figure 6b, in which the initial drop in laxity is greater than 1, from a laxity of i to a laxity of m. In this case, the number of consecutive slots for which the minimum laxity is below the upper dotted diagonal line is the sum of the number of consecutive slots ***strictly*** below the lower

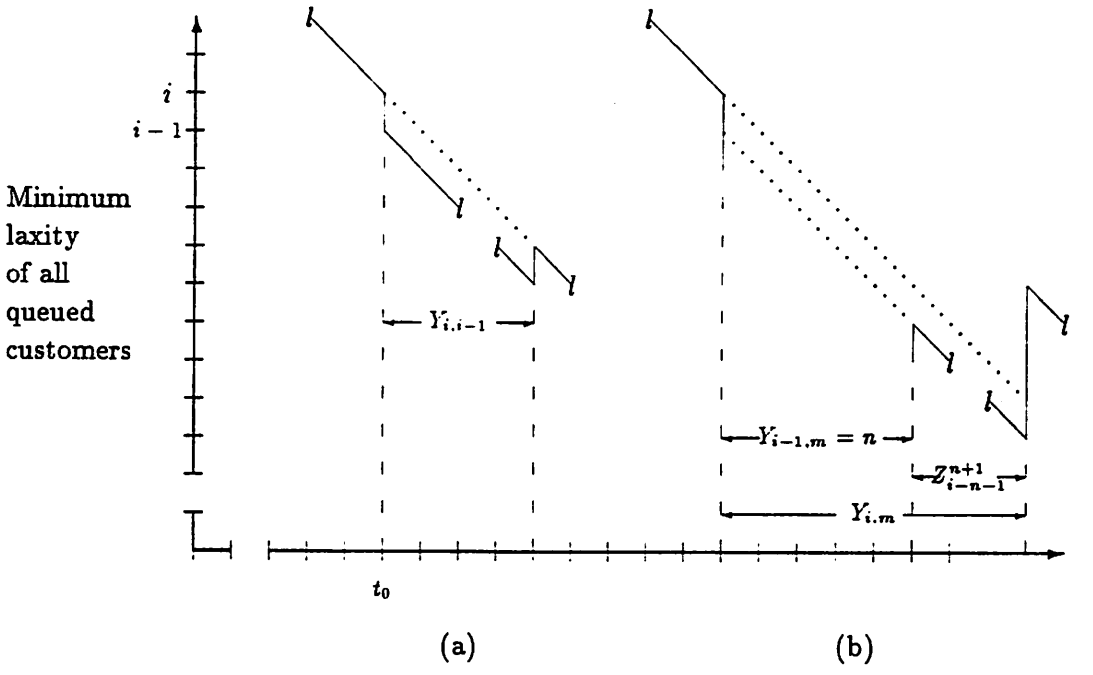

Figure 6: A sample path showing $Y_{l,m}$

dotted diagonal line (which is given by $Y_{i-1,m}$), plus the number of subsequent consecutive slots that the laxity remains on or below the lower diagonal line. This latter quantity is given by Z^{n+1}_{i-n-1}, where n is value of $Y_{i-1,m}$. We thus have:

$$Y_{i,m}(k) = \begin{cases} \sum_{n=1}^{k} Y_{i-1,m}(n) Z^{n+1}_{i-n-1}(k-n) & 3 \leq i \leq M;\ 1 \leq m \leq i-2;\ 1 \leq k \leq i-1 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Equations 3 and 4 provide a recursive formulation for computing $Y_{i,m}(k)$. Due to the single summation, the complexity of computing $Y_{i,m}(k)$ for all i, m and k is $O(M^4)$.

3.3 Computing the queueing time of a Distinguished Customer

Thus far, we have derived expressions for $X_i(k)$, $Z_i^j(k)$ and $Y_{l,m}(k)$. We now summarize how these quantities can be computed in a simple recursive manner.

- **Initialization.** Compute $X_1(j)\ \ \forall j$ using equation 1. Compute $Z_1^j(0)$ and $Z_1^j(1)\ \forall j$ using equation 2.
- **For each $i \leq M$.**
  - Compute $X_i(k)\ \ \forall k \leq i$ using $X_j(*)$ and $Y_{j,*}(*), j < i$ via equation 1.
  - Compute $Y_{i,m}(k)\ \ m = i-1,\ k \leq i-1$ using $Z^1_{i-1}(k)$ via equation 3.

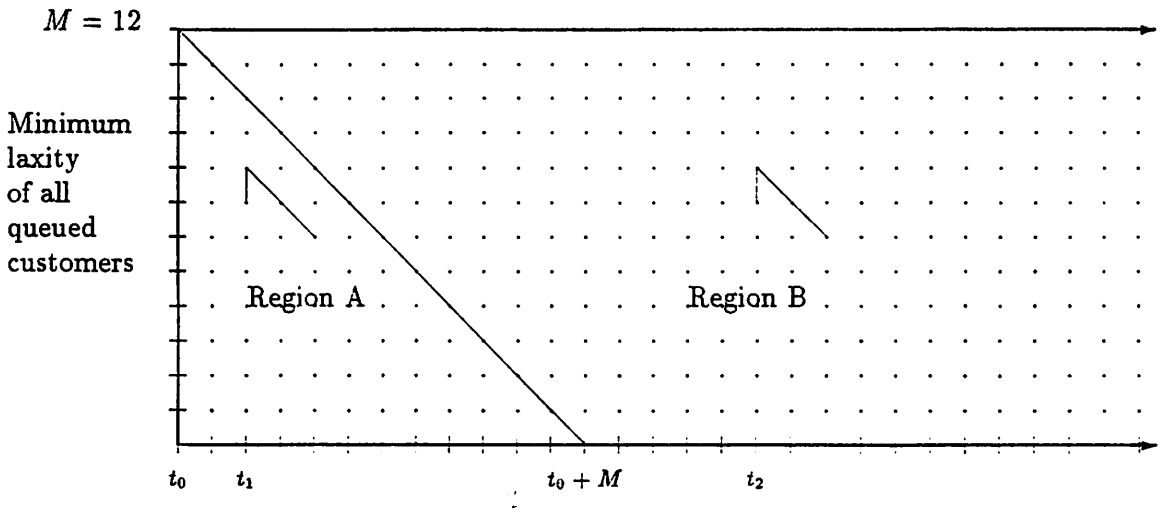

Figure 7: Two regions with different fundamental cycle statistics

- Compute $Y_{i,m}(k)$, $m < i-1, k \leq i-1$ using $Z_j^{\cdot}(*)$ and $Y_{j,\cdot}(*)$, $j \leq i-1$ via equation 4.
- Compute $Z_i^j(k)$ $\forall k \leq i, j \leq M-i+1$ using $Z_i^{j-1}(*), X_i(*)$, and $Z_{i'}^{\cdot}(*), i' < i$ via equation 2.

The computed quantities $Z_i^j(k)$ will be used in the following section to compute the expected length of the busy period.

3.4 The Expected Length of the Waiting Area Busy Period.

In computing the expected length of the waiting area busy period it will be helpful to consider Figure 7. Assume that a waiting area busy period begins at time t_0 and recall our definition of a fundamental cycle from section 2. We note that any fundamental cycle starting at height i in region B of this figure (i.e., to the right of the diagonal line from $(0, M)$ to $(M, 0)$) was started by a customer which may have arrived in any of the previous $M-i+1$ slots (where, as before, the current slot in which the fundamental cycle begins is included in the $M-i+1$ slots). For example, a fundamental cycle beginning at $(t_2, 8)$ could have resulted from an arrival at t_2 with a laxity of 8, at t_2-1 with a laxity of 9, at t_2-2 with a laxity of 10, at t_2-3 with a laxity of 11, or at t_2-4 with a laxity of 12. On the other hand, a fundamental cycle beginning in region A at some time t_1 at height i could only have been started due to an arrival in the previous $(t_1-t_0)+1$ slots, since by definition the waiting area busy period began at t_0 and thus all arrivals before t_0 have already left the waiting area. For example, the fundamental cycle beginning at $(t_1, 8)$ could only have resulted from an arrival at t_1 with a laxity of 8, at t_1-1 with a laxity of 9, or t_0 with a laxity of 10.

Let us define

B_i^j as the expected length of the remainder of a waiting area busy period, measured from the

beginning of a fundamental cycle of starting height i that begins j slots after the beginning of a waiting area busy period. *Note that for all $i \geq M$, B_i^j is independent of j; we shall refer thus to these values simply by B_i^M.*

$\nu_{l,m}$ as follows. Suppose a fundamental cycle ends at some height $n < l$ at the end of a slot. Then $\nu_{l,m}$ is the probability that there are no queued customers at the beginning of the following slot with a current laxity of l which arrived within the previous m slots, where the current slot is included in m.

$$\nu_{l,m} = \begin{cases} \prod_{j=l}^{l+m-1} \pi_{j,0} & 1 \leq l \leq M, \ \ 1 \leq m \leq M-l+1 \\ 0 & \text{otherwise} \end{cases}$$

$\eta_{i,j,k}$ as the probability that a fundamental cycle which ends with a minimum laxity of i is followed by another fundamental cycle which begins with a laxity of j, $j > i$, given that all customers which are still in the waiting area must have arrived in the previous k slots. The current slot is included in k.

$$\eta_{i,j,k} = \begin{cases} (1-\nu_{j,k}) \prod_{l=i+1}^{j-1} \nu_{l,k} & 0 \leq i \leq M-1, \ \ i+1 \leq j \leq M, \ \ 1 \leq k \leq M-j+1 \\ 0 & \text{otherwise} \end{cases}$$

$W_i^j(k)$ as the conditional value of $Z_i^j(k)$ given that the length of Z_i^j is non-zero. From equation 2 we have $W_i^j(k) = \sum_{n=1}^{k} X_i(n) Z_{i-n}^{n+j}(k-n)$. Note that all values of $W_i^j(k)$ can be computed in a total of $O(M^4)$ time.

Given the above, the following relationship holds for $\{B_i^M\} \forall i$.

$$B_i^M = \sum_{k=1}^{i} W_i^{M-i+1}(k)(k + \sum_{j=1}^{M} \eta_{i-k,j,M-j+1} B_j^M) \tag{5}$$

Equation 5 results from the fact that the remainder of a waiting area busy period begins first with a fundamental cycle (starting at height i) of some length k. Once the fundamental cycle is over, the waiting area busy period will either be over, or another fundamental cycle of some starting height j will begin. $\eta_{i-k,j,M-j+1}$ is simply the probability that a subsequent fundamental cycle begins at height j. Equation 5 can be rewritten in the more convenient form:

$$B_i^M = \sum_{k=1}^{i} W_i^{M-i+1}(k)k + \sum_{j=1}^{M} B_j^M \left(\sum_{k=1}^{i} \eta_{i-k,j,M-j+1} W_i^{M-i+1}(k) \right) \tag{6}$$

Equation 6 is M simultaneous linear equations and hence can be solved in $O(M^3)$ time using a method such as Gauss elimination [7].

For a fundamental cycle beginning in the region $[t_0, t_0 + M]$, we have $B_i^j = B_i^M$ for all i and $j \geq M - i - 1$. Using similar arguments as above, for a fundamental cycle which might itself be

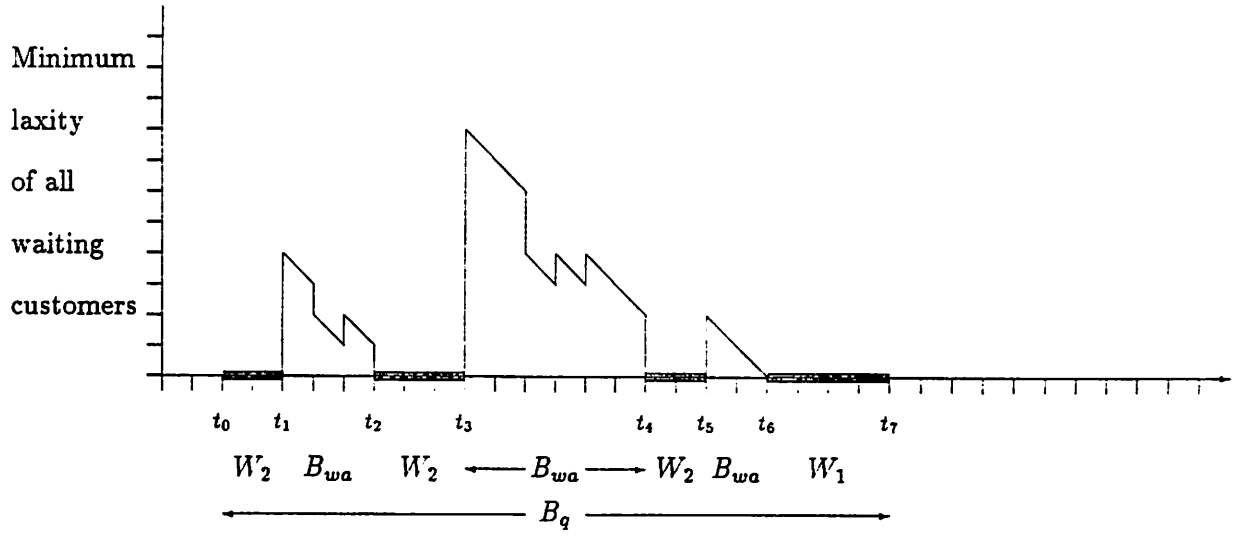

Figure 8: Waiting area busy periods (B_{wa}) and the queue busy period

followed by a fundamental cycle beginning below the diagonal line in Figure 1, (i.e., for all $i < M$ and $j < M - i - 1$ we have:

$$B_i^j = \sum_{k=1}^{i} W_i^{j+1}(k)(k + \sum_{l=i-k+1}^{M} \eta_{i-k,l,min(j-k+1,M-l-1)} B_l^{min(j+k,M)}) \tag{7}$$

Due to the recursive structure of equation 7, it can be solved for all i and j by backwards substitution in $O(M^4)$ time.

3.5 The Expected Length of the Queue Busy Period

Given the results of the previous section, we are now in a position to describe the complete procedure for computing customer loss. Recall that as shown in Figure 8, each busy period of the queue as a whole, B_q, consists of zero or more "waiting area busy periods", B_{wa}, connected by intervals of time in which there is but one customer in the queue; these latter periods of time are labeled W_2 and W_1 in Figure 8. In this section, we show how the expected lengths of W_1, W_2 and B_{wa} can be computed and, in turn, how $E[B_q]$ can be computed. $E[B_q]$ will be used to compute the expected customer loss.

Given the assumptions in section 2, each queue busy period contains one interval of type W_1 (see Figure 8), which ends the queue's busy period. Each busy period of the queue will also contain zero or more pairs of intervals consisting of W_2 and B_{wa}, where the W_2 is an interval of time during which there is exactly one customer in the queue (the job in service), followed by a waiting area busy

Event name	Event	Outcome	Probability
E_1	job finishes, 0 arrivals	W ends, idle period follows	$\mu\alpha_0$
E_2	job finishes, 1 arrival	W continues	$\mu\alpha_1$
E_3	job finishes, 2^+ arrivals	B_{wa} starts	$\mu\alpha_{2+}$
E_4	job continues, 0 arrivals	W continues	$(1-\mu)\alpha_0$
E_5	job continues, 1 arrival	B_{wa} starts	$(1-\mu)\alpha_1$
E_6	job continues, 2^+ arrivals	B_{wa} starts	$(1-\mu)\alpha_{2+}$

Table 1: Outcome of subsequent slot during a interval of queue length 1 (W_i)

period. Due to the independence of the service and arrival processes, the termination of a W interval by either the start of a new waiting area busy period (in the case of W_2) or the start of a queue idle period is independent from one W interval to the next. Hence, the number of such pairs of intervals in a queue busy period is geometrically distributed. *Thus, each queue busy period consists of one interval of length W_1 plus a geometrically distributed number of intervals of length $W_2 + B_{wa}$* and the expected length of the busy period is given by:

$$E[B_q] = E[W_1] + \sum_{i=0}^{\infty} b(1-b)^i (E[W_2] + E[B_{wa}])i = E[W_1] + \frac{1-b}{b}(E[W_2] + E[B_{wa}]) \qquad (8)$$

where b is the to-be-determined parameter of the modified geometric distribution for the number of W_2, B_{wa} pairs. We now proceed to calculate $E[W_1], E[W_2]$, b and $E[B_{wa}]$.

Computing $E[W_1]$

Table 1 enumerates the probability of the various possible states of the queue during a slot, given that the queue was in a W_1 or W_2 interval at the end of the previous slot. Given that the current slot occurs in a W_1 interval, the only possible outcomes for the subsequent slot are E_1, E_2 and E_4, with event E_1 terminating the W_1 interval. Hence, the probability that W_1 terminates at the end of the current slot is given by:

$$w_1 = \frac{\mu\alpha_0}{\mu\alpha_0 + \mu\alpha_1 + (1-\mu)\alpha_0} = \frac{\mu\alpha_0}{\alpha_0 + \mu\alpha_1}$$

and $W_1(i)$, the probability that W_1 is of length i, is geometrically distributed with parameter w_1:

$$W_1(i) = w_1(1-w_1)^{i-1} \qquad (9)$$

and hence

$$E[W_1] = \frac{1}{w_1} \qquad (10)$$

Computing $E[W_2]$

Similarly, given that the current slot occurs in a W_2 interval, the possible outcomes for the subsequent slot are E_2, E_3, E_4, E_5 and E_6, with events E_3, E_5 or E_6 terminating the W_2 interval. Hence, the probability that W_2 terminates at the end of the current slot is given by:

$$w_2 = \frac{\mu\alpha_{2+} + (1-\mu)(1-\alpha_0)}{1-\mu\alpha_0}$$

and $W_2(i)$, the probability that W_2 is of length i, is geometrically distributed with parameter w_2:

$$W_2(i) = w_2(1-w_2)^{i-1} \tag{11}$$

and hence

$$E[W_2] = \frac{1}{w_2} \tag{12}$$

Computing b

Consider Table 1 again. Given that the queue is in an interval of time in which the queue length is one, this interval is followed by a waiting area busy period if events E_3, E_5 or E_6 terminate the interval. If event E_1 terminates the interval, an idle period will follow. Thus, the value of b (the parameter of the geometrically distributed number of W_2, B_{wa} interval pairs occurring during a queue busy cycle) is given by

$$b = \frac{\mu\alpha_0}{\mu\alpha_0 + \mu\alpha_{2+} + (1-\mu)(1-\alpha_0)} \tag{13}$$

Computing $E[B_{wa}]$

Finally, in order to calculate $E[B_{wa}]$, we condition on the laxity of the minimum laxity customer that begins the waiting area busy period and use the values of B_i^0 from the previous section. Let us define h as the starting height (in laxity) of the waiting area busy period. For $E[B_{wa}]$, we have

$$E[B_{wa}] = \sum_{i=1}^{M} B_i^0 P(h=i). \tag{14}$$

Note that events E_3, E_5 and E_6 in Table 1 result in the start of a waiting area busy period. Hence,

$$\begin{aligned} P(h=i) \;=\; & P(h=i|E_3)\tfrac{\mu\alpha_{2+}}{\mu\alpha_{2+}+(1-\mu)(1-\alpha_0)} + \\ & P(h=i|E_5 \text{ or } E_6)(1-\tfrac{\mu\alpha_{2+}}{\mu\alpha_{2+}+(1-\mu)(1-\alpha_0)}). \end{aligned} \tag{15}$$

From first principles,

$$P(h=i|E_3) = \frac{(1-\pi_{i,0}-\pi_{i,1})}{1-\alpha_0-\alpha_1}\prod_{j=1}^{i-1}\pi_{j,0} + \frac{(1-\delta(M,i)\pi_{i,1}}{1-\alpha_0-\alpha_1}\prod_{j=1}^{i-1}\pi_{j,0}(1-\prod_{j=i+1}^{M}\pi_{j,0}) \tag{16}$$

$$P(h = i | E_5 \text{ or } E_6) = \frac{1}{1-\alpha_0}(1-\pi_{i,0}) \prod_{j=1}^{i-1} \pi_{j,0} \tag{17}$$

The expected length of the busy period of the queue can now be computed via equation 8, using equations 10 through 17.

3.6 From the Busy Period to Customer Loss

Given $E[B_q]$, we can now exploit the regenerative structure of the queueing system to compute the probability that the server is busy. The probability that the idle period of the queue is of length i is given by $I(i) = (\alpha_0)^{i-1}(1-\alpha_0)$ and hence the expected length of the idle period, $E[I]$ is given by $E[I] = 1/(1-\alpha_0)$. The probability that the server is idle, $P0$ is thus given by

$$PO = \frac{E[I]}{E[I] + E[B_q]}. \tag{18}$$

The probability that a job is lost can now be computed using flow conservation arguments. Recall that λ is the mean arrival rate to the queue (including both lost and successfully served customers) and that μ is the service rate of the queue. The rate at which customers are successfully accepted to the server is given by $\lambda(1 - P_{loss})$ and the rate at which customers leave the queue having received service is $\mu(1 - P0)$. Hence, by flow conservation, the probability of customer loss is:

$$P_{loss} = 1 - \frac{\mu}{\lambda}(1 - P0) \tag{19}$$

4. Extensions to the Basic Model

In this section we briefly sketch how the basic algorithm of section 3 can be extended to the cases of general service times, a generally distributed number of arrivals with a laxity of i ($1 \leq i \leq M$) in a slot, and multiple-server queues.

General Service Times

In order to account for general service time distributions, the basic recursions for $X_i(k)$, $Z_i^j(k)$ and $Y_{i,m}(k)$ (equations 1, 2 and 4, respectively) could be modified to explicitly track the amount of service already received by the customer in service. This adds yet another "dimension" to each of the associated random variables and hence increases the complexity of computing the distributions of the random variables. If the maximum service time is bounded by some value, S, then the complexity of computing $Z_i^{j,s}$, where s now indicates the amount of service already received by the customer in service, increases to $O(SM^4)$. If s is potentially infinite, the range of s could be truncated at some appropriate value.

Even with this increase in complexity, however, an exact calculation of the loss does not seem possible. This is due to the fact that in order to compute the expected length of the W_i intervals in Figure 8, the amount of service time already received by the customer in service at the beginning of a W_i interval (and hence at the end of a B_{wa} period) must be known. This is often easy to compute; note that for waiting area busy periods which end with the minimum laxity customer entering service (e.g., the waiting area busy periods ending at t_2 and t_4 in Figure 8), the amount of service time accrued by the customer in service is known to be zero at the start of a W_i interval. However, in cases where the waiting area busy period ends as the result of the loss of the single queued minimum laxity customer (e.g., the waiting area busy period ending at t_6), the amount of accrued service time for the job in service is not known. One possible *approximation* would be to assume such events are rare and that each W_i period thus always begins with a new customer entering service. In this case, the remainder of the analysis would proceed along the same lines as in section 3.5.

Generally Distributed Number of Bulk Arrivals

In the case of a generally distributed number of arrivals in a slot with an initial laxity of i, the basic recursions for $X_i(k)$, (equation 1) remain unchanged. The recursion for $Z_i^j(k)$, equation 2, would have to be modified to explicitly track b, the number of customers that arrived in the same bulk (i.e., in the same slot and with the same initial laxity) as the current minimum laxity customer and that have already received service. In this case, equation 2 becomes

$$Z_i^{j,b}(k) = \begin{cases} \pi^b_{i+j-1,0} Z_i^{j-1,0}(k) + (1 - \pi^b_{i+j-1,0}) \sum_{n=1}^{k} X_i(n) Z_{i-n}^{n+j,b+1}(k-n) & 1 \le j \le M-i+1; \\ & 0 \le k \le i \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

and $Z_0^{j,b}(0) = 1$, $Z_j^{0,b}(0) = 1$, and $Z_j^{0,b}(k|k \neq 0) = 0$ for all b by definition. Here $\pi^b_{i+j-1,0}$ is the probability that there are no additional arrivals in the same bulk as the current minimum laxity customer, given that b customers arriving in this same bulk have already been served.

The basic recursion (equation 4) for $Y_{i,m}(k)$ then becomes:

$$Y_{i,m}(k) = \begin{cases} \sum_{n=1}^{k} Y_{i-1,m}(n) Z_{i-n-1}^{n+1,1}(k-n) & 3 \le i \le M;\ 1 \le m \le i-2;\ 1 \le k \le i-1 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

Note that when there are more than M arrivals in the same bulk, at most M of these customers will ever receive service, and hence the range of b values can be truncated at M. Once the values of $Z_i^{j,b}(k)$ have been computed, the analysis in sections 3.4 and 3.5 proceeds as before (using only different values of $\nu_{l,m}$). Note that while the introduction of b does not require any additional summations besides those already in equations 2 and 4, values of $Z_i^{j,b}(k)$ must be computed for all values of $b, 0 \le b \le M$. The complexity of computing the loss thus increases to $O(M^5)$.

Multiple Server Queues

In the case of an N-server queue, the basic recursions for $X_i(k)$, $Z_i^j(k)$ and $Y_{i,m}(k)$ (equations 1, 2 and 4, respectively) remain unchanged except for the replacement of μ by $N\mu$ (assuming geometrically distributed service times); hence $E[B_{wa}]$ can be computed in $O(M^4)$ time. Note that during a B_{wa} period (see Figure 8), there are always N customers in service and jobs are thus leaving the server (after having successfully met their deadline constraint) at a rate of $N\mu$.

The calculation of the loss is not straightforward, however, and we only outline the procedure here. The primary complications arise from the fact that, in the case of multiple servers, the steady state distribution of the number of customers queued during the W_i periods of the queue busy period must be known if we are to use the flow conservation approach towards computing loss. Conditioned on the system being in a W_i period, these steady state probabilities are the same as the occupancy probabilities for the $GI/Geom/N/N$, a system analyzed in [3] and pp. 271 in [15]. Note that the assumption of a geometrically distributed number of arrivals with the same laxity in a slot (for all possible laxities) generally gives rise to a non-geometric number of overall arrivals in a slot and hence the $GI/Geom/N/N$ queue must be analyzed.

Once the expected length of the busy period of the $GI/Geom/N/N$ queue and the expected amount of time the queue spends in a waiting area busy period during a busy period have been computed, the steady occupancy probabilities during a queue busy period, p_i, $i = 1, \ldots, N, N^+$ can be calculated (where P_{N^+} is the probability that there are more than N customers in the queue). The rate at which customers leave the queue during a queue busy period after having successfully been served is given by $\xi = Np_{N^+} + \sum_{i=1}^{N} ip_i$. The unconditional rate at which customers successfully leave the queue is then ξP_0, where P_0 is given by equation 18. The customer loss can then be computed by flow conservation arguments, as in section 3.5.

5. Conclusion

In this paper we have presented an algorithm which exactly computes customer loss for a discrete-time queueing system having customers with real-time constraints. The time complexity of the algorithm is $O(M^4)$ (where M is the largest possible laxity of a customer) for the case of geometrically distributed service times and a bulk arrival process in which the number of customers arriving in a slot with a deadline of i slots is also geometrically distributed. We also discussed how this model might be extended to include multiple servers and generally distributed service times and laxities as well.

A number of important related problems still remain. Perhaps most importantly, the approach presented in this paper can not be used to compute customer loss when deadlines are to the *end* of service (except for the case of constant service times) since the service time distribution of successfully served customers is no longer the same as that of arriving customers (and hence the flow conservation approach of section 3.5 can not be used). Two other (even more challenging) open problems are that

of computing the waiting time distribution of successfully served customers and the laxity distribution of queued customers.

REFERENCES

[1] F. Baccelli, P. Boyer, G. Hebuterne, "Single-Server Queues with Impatient Customers", *Adv. Appl. Probability*, Vol. 16 (1984), pp. 887-905.

[2] P. Bhattacharya and A. Ephremides, "Optimal Scheduling with Strict Deadlines," *IEEE Trans. on Automatic Control*, Vol. 34, No. 7 (July 1989), pp. 721-728.

[3] W. Chan and D. Maa, "The GI/Geom/N Queue in Discrete Time," *INFOR - The Canadian Journal of Oper. Res. Inform. Process.*, Vol. 16, pp. 232-252.

[4] T. Chen, J. Walrand and D. Messerschmidtt, "Dynamic Priority Protocols for Packet Voice," *IEEE J. on Selected Areas in Communications*, Vol. 7, No. 5 (June 1989), pp. 632-643.

[5] R. Chipalkatti, J.F. Kurose, and D. Towsley, "Scheduling Policies for Real-time and Non-real Time Traffic in a Statistical Multiplexer," *Proc. IEEE Infocom '89 Conference*, (April 1989, Ottawa Canada), pp. 774-783.

[6] J.P. Coudreuse and M. Servel, "Prelude: An Asynchronous Time-Division Switched Network," *Proc 1987 IEEE Int. Conf. on Commun.*, (June 1987, Seattle, WA), pp. 22.2.1-22.2.5.

[7] G. Dahlquist and A. Bjork, Numerical Methods, Prentice Hall (Englewood Cliff, NJ, 1974).

[8] B. Doshi and H. Heffes, "Overload Performance of Several Processor Queueing Disciplines for the M/M/1 Queue", *IEEE Trans. Communications*, Vol. COM-34, No. 6 (June 1986), pp. 538-546.

[9] B. Gavish, P. Schweitzer, "The Markovian Queue with Bounded Waiting Time", *Management Science*, Vol. 23, No. 12 (Aug. 1977), pp. 1349-1357.

[10] B. Gnedenko and I. Kovalenko, Introduction to Queueing Theory, Israel Program for Scientific Translation, Jerusalem, 1968.

[11] H. Goldberg, "Analysis of the Earliest Due Date Scheduling Rule in Queueing Systems," *Mathematics of Operations Research*, Vol. 2, No. 2 (May 1977), pp 145-154.

[12] H. Goldberg, "Jackson's Conjecture on Earliest Due Date Scheduling," *Mathematics of Operations Research*, Vol. 5, No. 3 (Aug. 1980), pp. 460-466.

[13] R. Haugen and E. Skogan, "Queueing Systems with Stochastic Timeout," *IEEE Trans. on Communications*, Vol. 28, No. 2 (Dec. 1980), pp. 1984-1989.

[14] J. Hong, X. Tan and D. Towsley, "A Performance Analysis of Minimum Laxity and Earliest Deadline Scheduling in a Real-Time System," to appear in *IEEE Transactions on Computers*, Dec. 1989.

[15] J. Hunter, Mathematical Techniques of Applied Probability (Vol 2, Discrete Time Models: Techniques and Applications), Academic Press (New York, 1983).

[16] J.R. Jackson, "Scheduling a Production Line to Minimize Maximize Lateness," Research Report 43, Management Science Research Report, UCLA, 1955.

[17] J.F. Kurose, M. Schwartz and Y. Yemini, "Multiple Access Protocols and Real-Time Communication", *Computing Surveys*, Vol. 16, No. 1, (March 1984), pp. 43-70. Also reprinted in Hard Real-Time Systems, J. Stankovic and K. Ramamritham, IEEE Computer Society Press, Washington, DC, 1988.

[18] J.F. Kurose and R. Chipalkatti, "Load Sharing in Real-Time Distributed Computer Systems," *IEEE Transactions on Computers*, Vol. 36, No. 8 (August 1987), pp. 993-1000.

[19] M. Ouaily and I. Rubin, "A Single Server Queueing System Under Various Time Constraints," Technical Report UCLA-ENG-88-10, Department of Electrical Engineering, UCLA, Los Angeles, CA.

[20] S. Panwar, D. Towsley and J. Wolf, "Optimal Scheduling Policies for a Class of Queues with Customer Deadlines to the Beginning of Service," *J. of the ACM*, Vol. 35, No. 4 (Oct. 1988), pp. 832-844.

[21] S. Panwar and D. Towsley, "Optimal Scheduling "On the Optimality of the STE Rule for Multiple Server Queues That Serve Customers with Deadlines," COINS Technical Report 88-81, University of Massachusetts, Amherst MA, August, 1988.

[22] I. Rubin and M. Ouaily, "Performance of Communication and Queueing Processors Under Message Delay Limits," *Proc. 1988 IEEE Globecom Conference*, (Miami, Dec. 1988), pp. 501-505.

[23] H. Saito, "Optimal Queueing Discipline for Real-Time Traffic at ATM Switching Nodes," *Proceedings of the IEICE of Japan*, Sept. 1988, pp. 49-54.

[24] T. Takegughi and T. Yamaguchi, "Synchronous Composite Packet Switching for ISDN Switching System Architecture," *Proc. 1984 Int. Switching Symp.*, (May, 1984, Florence Italy).

[25] Y. Yeh, M. Hluchyj, A. Acampora, "The Knockout Switch: A Simple, Modular Architecture for High Performance Packet Switching," *IEEE J. on Selected Areas in Commun.*, Vol. SAC-5, No. 8 (Aug. 1987), pp. 1274-1282.

[26] C. Yuan and J. Silvester, "Queueing Analysis of Delay Constrained Voice Traffic in a Packet Switching System," *IEEE J. on Selected Areas in Communications*, Vol. 7, No. 5 (June 1989), pp. 729-738.

[27] W. Zhao and J. Stankovic, "Performance Analysis of FCFS and Improved FCFS Scheduling Algorithms for Dynamic Real-Time Computer Systems," to appear in *Proc Tenth Real-Time System Symposium*, Dec. 1989.