

ON QUEUES WITH INTERARRIVAL TIMES PROPORTIONAL TO SERVICE TIMES

ISRAEL CIDON*†

*Department of Electrical Engineering
Technion, Haifa 32000, Israel*

ROCH GUÉRIN

*IBM T. J. Watson Research Center
Yorktown Heights, New York 10598*

ASAD KHAMISY‡ AND MOSHE SIDI‡

*Department of Electrical Engineering
Technion, Haifa 32000, Israel*

We analyze a family of queueing systems where the interarrival time I_{n+1} between customers n and $n + 1$ depends on the service time B_n of customer n . Specifically, we consider cases where the dependency between I_{n+1} and B_n is a *proportionality* relation and B_n is an exponentially distributed random variable. Such dependencies arise in the context of packet-switched networks that use rate policing functions to regulate the amount of data that can arrive to a link within any given time interval. These controls result in significant dependencies between the amount of work brought in by customers/packets and the time between successive customers. The models developed in the paper and the associated solutions are, however, of independent interest and are potentially applicable to other environments.

Several scenarios that consist of adding an independent random variable to the interarrival time, allowing the proportionality to be random and the combination of the two are considered. In all cases, we provide expressions for the Laplace–Stieltjes Transform of the waiting time of a customer in the system. Numerical results are provided and compared to those of an equivalent system without dependencies.

*This work was done while at the IBM T. J. Watson Research Center.

‡Part of this work was done while visiting the IBM T. J. Watson Research Center.

1. INTRODUCTION

This paper deals with the analysis of a family of queueing systems, where the interarrival time between two customers depends on the service time of the first customer. The motivation for studying these queueing systems originated in the context of high-speed packet-switched networks. The issue of interest was the effect of input policing functions that have been proposed to control the flow of packets in such networks. The basic goal of these policing functions is to ensure adequate network performance by regulating the amount of data that can arrive to a link within any given time interval. These controls result in significant dependencies between the amount of work brought in by packets and the time between the arrival of successive packets. Such dependencies have a significant impact on system performance [2,5,7,10,14], and the reader is referred to Fendick, Saksena, and Whitt [10] and Cidon, Guérin, Khamisy, and Sidi [4] for discussions on this topic and reviews of related works.

Whereas numerous previous papers have studied the effect of many different dependencies in queueing systems, very little work seems to have been done on the type of dependencies that are the focus of this paper. In particular, the case where the service time of a customer depends on the time since the *previous* arrival has been thoroughly studied [4–8,13,14,19,20]. In contrast, this paper addresses the converse problem where the time to the *next* arrival depends on the service time of the arriving customer. To the best of our knowledge, previous work on this problem has been essentially limited to the study of general conditions for either stability [21] or finite moments of the busy period [11].

The main contribution of this paper is, therefore, to analyze the waiting time of a customer in a single-server first in–first out (FIFO) queue, when the interarrival time I_{n+1} between customers n and $n + 1$ depends on the service time B_n of customer n . Specifically, we consider cases where the dependency between I_{n+1} and B_n is a *proportionality* relation, and B_n is an exponentially distributed random variable. The solution method is based on the application of the standard Wiener–Hopf method (e.g., see Kleinrock [17]). However, we exploit the particular structure of the system to provide a computationally much more efficient solution than available by applying standard techniques. This enables us to study a number of configurations that provide useful insight on the effect of new control mechanisms being used in high-speed networks.

Specifically, some policing functions now used in packet-switched networks introduce the type proportional dependency already mentioned. For instance, let us illustrate how a simple *spacer* controller that is used to limit the peak rate at which a source can generate data into a network (see Bala, Cidon, and Sohraby [1], Cidon et al. [3], and Elwalid and Mitra [9]) introduces such dependencies. The enforcement of a maximum rate is achieved by requiring that after sending a packet of size B a *space* of duration B/R be inserted before the next packet can be sent. The rate R is then the maximum allowable rate for the source. (Note that the existence of a maximum network packet size is assumed

here.) This rate R is typically equal to the source peak rate but can be set to a lower value when low-speed links are present in the connection's path [1,3] or if the network traffic has to be smoothed [12].

Assuming that the preceding spacer is saturated by a source of rate R whose traffic is fed to a link of speed C , interarrival and service times at the link are then proportional with $I_{n+1} = \alpha B_n$ and $\alpha = C/R$. As shown in Figure 1, which plots the evolution of the workload at the network link for the extreme case $\alpha = 1$, that is, equal source and link rates (stability requires $\alpha \geq 1$), the analysis of this simple case is of little interest. However, there exist a number of extensions to this basic model that make it nontrivial, although still tractable, and, more important, useful in modeling actual systems.

The simplest extension consists of adding an independent random variable to the interarrival time. The presence of such a random component in the interarrival time allows us to model more accurately how the traffic generated by a spacer controller arrives at an internal network link. First, such a model can capture the effect of interactions between packets from a given source and other traffic streams inside the network. In particular, the gaps that the spacer initially imposes between packets are modified according to the different delays that consecutive packets observe through the network. The arrival process at a link can then be modeled as consisting of a deterministic component (the spacing imposed by the spacer at the network access), to which a random network jitter has been added. Second, the addition of a random component also allows us to relax the assumption of a saturated spacer queue, because it can be used to model the time between packets that arrive to the spacer. Finally, another useful application is when the spacer itself randomizes the gaps between successive packets. This randomization in the spacer may be useful to avoid correlation between traffic streams of distinct sources. In particular, it helps to prevent (malicious) sources from harming network performance by cooperating to generate a large burst of data into the network.

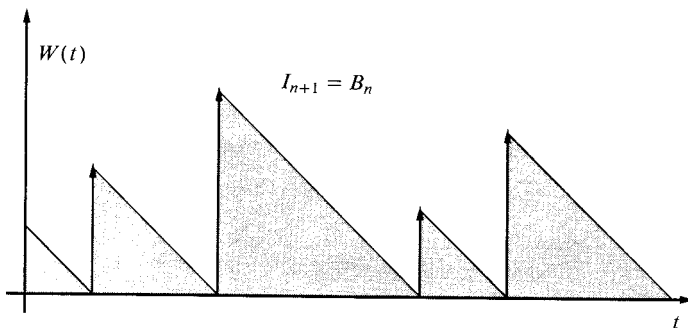


FIGURE 1. Workload in queue with interarrival time proportional to service time.

Another extension of the basic model is to allow the proportionality constant to be itself a random variable. Specifically, we consider cases where the proportionality factor is randomly chosen from a finite set of values. This allows the modeling of a generalized spacer, where the factor used to compute the enforced spacing is allowed to vary. For example, the arrival of a high-priority packet that is sensitive to access delay could be handled by allowing earlier transmission. The interarrival times at the network link would then depend on both packet priorities and the size of the previous packet. This model has also applications to other environments, such as manufacturing or repair stations systems.

The organization of the paper is as follows. In Section 2, we introduce the notations and models studied in the paper. In Section 3, we derive the Laplace-Stieltjes Transform (LST) $\mathcal{W}(s)$ of the waiting time in steady state for the simple case of deterministic proportional dependencies with an additional, positive, exponentially distributed and independent jitter. The Appendix shows how this can be extended to allow for both positive and negative jitters. The case where the proportionality factor is allowed to be itself a random variable is treated in Section 4. The analysis of this system requires the use of the spectral analysis method typical to G/G/1 queues. We provide some numerical examples that compare the performance of the system with dependencies with an equivalent system without dependencies. Section 5 covers the addition of a random delay similar to that of Section 3 to the scenario of Section 4. Finally, Section 6 summarizes the findings of this paper.

2. THE MODEL

2.1. General

We consider queueing systems in which the interarrival time between the n th and the $n + 1$ st customers (packets) depends on the service time of the n th customer. We focus on *proportional* dependency, which is very natural in packet-switched networks. Here the interarrival time between two consecutive packets arriving over a communication link is proportional to the size (in bits) of the first packet and, consequently, to the time it will take to forward this packet over the next link. As was previously discussed, this model (interarrival proportional to previous service time) captures the effect of a rate control mechanism that “spaces” packets apart as a function of their size. In addition to the proportional dependency, it is very helpful to allow the addition of an independent random component (to the interarrival time) that models a variable delay (jitter) caused by a travel through the network or additional delays between packets imposed at the source.

In what follows, we describe the three scenarios that are considered in this paper. The first and the simplest one assumes deterministic proportional dependency with an additive interarrival random delay. The second scenario explores the effect of random proportional dependency without additive delay. The third

is a combination of the first and the second scenarios, namely, random proportional dependency with additive interarrival random delay.

Our system consists of a single-server queue with an infinite buffer. The service time of the n th packet ($n \geq 1$) is denoted by B_n and is assumed to be an independent exponentially distributed random variable (r.v.) with parameter μ . The interarrival time between the n th and the $n + 1$ st packets is denoted by I_{n+1} . (Without loss of generality, we assume that the first arrival occurs at time 0.) The workload in the system just *before* the n th packet arrival is a r.v. denoted by W_n with a probability density function (p.d.f.) $f_{W_n}(w)$ and LST $\mathcal{Q}_{W_n}(s) = E[e^{-sW_n}]$. We also define $\mathcal{Q}W(s) = \lim_{n \rightarrow \infty} \mathcal{Q}_{W_n}(s)$ when the limit exists. The evolution of the workload at arrival epochs is given by

$$W_{n+1} = (W_n + B_n - I_{n+1})^+, \quad n \geq 1, \quad (1)$$

where $X^+ = \max(0, X)$ and W_1 is usually assumed to be 0. Obviously, W_n is also the waiting time of the n th packet when packets are served according to the FIFO order.

The evolution described in Eq. (1) is typical to G/G/1 queues and is very well known [17]. In particular, when $B_n, I_{n+1}, n \geq 1$, are independent r.v.'s and B_n is exponentially distributed, Eq. (1) describes the evolution of the workload in a GI/M/1 system. The feature that characterizes the queuing systems considered in this paper is that I_{n+1} is allowed to depend on B_n .

2.2. Deterministic Proportional Dependency with Additive Delay

We first characterize the most basic system of deterministic proportional dependency without additive delay. In this system, the interarrival time I_{n+1} is proportional to the previous service time of the n th packet B_n with a (fixed) proportional dependency parameter α , that is, $I_{n+1} = \alpha B_n$. Therefore, the workload in the system just *before* the arrival epochs of packets evolves according to

$$W_{n+1} = (W_n + (1 - \alpha)B_n)^+. \quad (2)$$

The solution of Eq. (2) when $n \rightarrow \infty$ is trivial. We have

$$\lim_{n \rightarrow \infty} W_n = \begin{cases} 0 & \text{if } \alpha > 1 \\ W_1 & \text{if } \alpha = 1 \\ \infty & \text{if } \alpha < 1 \end{cases} \quad (3)$$

because the term $(1 - \alpha)B_n$ is always negative, zero, or positive, respectively. For example, the evolution of a system where $\alpha = 1$ and $W_1 = 0$ is depicted in Figure 1. Therefore, we introduce the simplest nontrivial system by adding an independent component to the interarrival time I_{n+1} , namely, $I_{n+1} = \alpha B_n + J_n$ when J_n is an independent r.v. Therefore,

$$W_{n+1} = (W_n + (1 - \alpha)B_n - J_n)^+. \quad (4)$$

It should be clear now that if J_n is positive then it must be that $\alpha < 1$ to avoid a trivial zero solution in steady state. It is also clear that the system is stable if $E[(1 - \alpha)B_n - J_n] < 0$. Consequently, if the expected value of J_n is $1/\delta$, then the system is stable if and only if $\alpha + \mu/\delta > 1$.

In Section 3, we analyze the steady-state behavior of the system assuming that J_n is an independent exponentially distributed r.v. with parameter δ . In the Appendix, we extend the results of Section 3 to the case where J_n can also take negative values.

It is also of interest to explicitly identify the resulting arrival process to this system. When $I_{n+1} = \alpha B_n + J_n$ is a sum of two independent exponentially distributed r.v.'s with parameters μ/α and δ , respectively, its LST can be expressed as $[(\mu/\alpha)/(\mu/\alpha + s)][\delta/(\delta + s)]$ (independent of n). Therefore, the p.d.f. of I_n is $f_{I_n}(t) = [\delta\mu/(\alpha\delta - \mu)][e^{-(\mu/\alpha)t} - e^{-\delta t}]$ when $\mu \neq \alpha\delta$ and $\delta^2 t e^{-\delta t}$ otherwise.

2.3. Random Proportional Dependency

Let us reexamine the system without additive delay as described by Eq. (2). To avoid a trivial solution in the form of Eq. (3), we will assume that the proportionality parameter itself is a r.v. different from a constant, namely,

$$W_{n+1} = (W_n + (1 - \Omega_n)B_n)^+. \quad (5)$$

Therefore, $I_{n+1} = \Omega_n B_n$, where Ω_n is a r.v. with a finite support, independent of any other r.v. in the system. Specifically, we consider the case where $\Omega_n = \alpha_i$ with probability a_i for $1 \leq i \leq N + M$ for some integers $N, M \geq 1$. Clearly, $\sum_{i=1}^{N+M} a_i = 1$. Furthermore, it is clear that in order to avoid trivial solutions not all the α_i are greater than or equal to 1 and also not all the α_i are equal to or smaller than 1. Without loss of generality, assume that $1 < \alpha_1 < \alpha_2 < \dots < \alpha_N$ and $\alpha_{N+1} < \alpha_{N+2} < \dots < \alpha_{N+M} \leq 1$. The stability condition for the system is $E[(1 - \Omega_n)B_n] < 0$ and, therefore, the system is stable if and only if $\sum_{i=1}^{N+M} a_i \alpha_i > 1$. In Section 4, we analyze the steady-state behavior of this system.

The interarrival time I_{n+1} is again independent of n and with probability a_i is an exponentially distributed r.v. with parameter μ/α_i . Therefore, its LST is $\sum_{i=1}^{N+M} a_i [(\mu/\alpha_i)/(\mu/\alpha_i + s)]$ and its p.d.f. is $\sum_{i=1}^{N+M} a_i (\mu/\alpha_i) e^{-(\mu/\alpha_i)t}$.

Another possible extension of this model is to assume that the average service time is also random—for example, with probability a_{ij} $\Omega_n = \alpha_i$ and $\mu = \mu_j$ —making μ itself a r.v. It turns out, however, that this is equivalent to enlarging the support of the r.v. Ω_n , as shown in Section 4.

2.4. Random Proportional Dependency with Additive Delay

The third model we consider is simply a combination of the two preceding scenarios. Here we have $I_{n+1} = \Omega_n B_n + J_n$, where Ω_n is a (nonconstant) r.v. and J_n is an additive random jitter to the interarrival period. Hence,

$$W_{n+1} = (W_n + (1 - \Omega_n)B_n - J_n)^+, \quad (6)$$

where $\Omega_n = \alpha_i$ with probability a_i and J_n is exponentially distributed with parameter δ . We will assume the ordering of the various α_i as defined in the previous model. The stability condition here is $\sum_{i=1}^{N+M} a_i \alpha_i + \mu/\delta > 1$. The analysis of this system is similar to that of the previous one and is given in Section 5.

The interarrival time I_n is again independent of n and with probability a_i is a sum of two independent exponentially distributed r.v.'s with parameters μ/α_i and δ , respectively. Therefore, its LST is $\sum_{i=1}^{N+M} [(a_i \mu/\alpha_i)/(\mu/\alpha_i + s)] [\delta/(\delta + s)]$ and its p.d.f. is $f_{I_n}(t) = \sum_{i=1}^{N+M} a_i [\delta \mu/(\alpha_i \delta - \mu)] [e^{-(\mu/\alpha_i)t} - e^{-\delta t}]$ in case that $\mu \neq \alpha_i \delta$ for all $1 \leq i \leq N + M$. If for some j ($1 \leq j \leq N + M$) $\mu = \alpha_j \delta$, then the j th term in the sum is replaced by $a_j \delta^2 t e^{-\delta t}$.

3. DETERMINISTIC PROPORTIONAL DEPENDENCY WITH ADDITIVE DELAY

The analysis of this system is relatively simple. We begin by expressing the LST $\mathcal{Q}_{W_{n+1}}(s)$ as follows:

$$\mathcal{Q}_{W_{n+1}}(s) = E[e^{-s[W_n + (1-\alpha)B_n - J_n]^+}],$$

where J_n is an exponentially distributed r.v. with parameter δ .

$$\begin{aligned} \mathcal{Q}_{W_{n+1}}(s) &= E[e^{-s[W_n + (1-\alpha)B_n - J_n]^+}] = \int_0^\infty f_{W_n}(w) dw \int_0^\infty \mu e^{-\mu x} dx \\ &\cdot \left[\int_0^{w+(1-\alpha)x} \delta e^{-\delta y} \cdot e^{-s[w+(1-\alpha)x-y]} dy + \int_{w+(1-\alpha)x}^\infty \delta e^{-\delta y} dy \right] \\ &= \frac{\mu \delta}{(\delta - s)(\mu + (1 - \alpha)s)} \mathcal{Q}_{W_n}(s) - \frac{\mu \delta}{(\delta - s)(\mu + (1 - \alpha)\delta)} \mathcal{Q}_{W_n}(\delta) \\ &+ \frac{\mu}{\mu + (1 - \alpha)\delta} \mathcal{Q}_{W_n}(\delta). \end{aligned} \tag{7}$$

Rearranging Eq. (7) and letting $n \rightarrow \infty$, we get

$$\mathcal{Q}(s) = \frac{1 + \gamma s}{(1 + \gamma \delta) [1 + \gamma(s - \delta)]} \mathcal{Q}(\delta), \tag{8}$$

where $\gamma = (1 - \alpha)/\mu$.

Note that the stability condition $\alpha + \mu/\delta > 1$ (or $\gamma \delta < 1$) ensures that the root of $1 + \gamma(s - \delta)$ at $s = \delta - \mu/(1 - \alpha)$ is negative and therefore $\mathcal{Q}(s)$ in Eq. (8) is an analytic function for $\text{Re}(s) > 0$, as required.

Applying the normalization condition $\mathcal{Q}(0) = 1$, we get $\mathcal{Q}(\delta) = (1 - \gamma \delta)/(1 + \gamma \delta)$ and, hence,

$$\mathcal{Q}(s) = \frac{(1 - \gamma \delta)(1 + \gamma s)}{1 + \gamma(s - \delta)} = 1 - \gamma \delta + \gamma \delta \frac{\frac{1 - \gamma \delta}{\gamma}}{s + \frac{1 - \gamma \delta}{\gamma}}. \tag{9}$$

The quantity ${}^{\circ}W(\infty) = 1 - \gamma\delta$ is the probability that the waiting time just prior to an arrival in steady state is zero. It is easy to invert ${}^{\circ}W(s)$ and realize that with probability $1 - \gamma\delta$ the waiting time is zero, and with probability $\gamma\delta$ it is exponentially distributed with parameter $(1 - \gamma\delta)/\gamma = [\mu - (1 - \alpha)\delta]/(1 - \alpha)$.

The waiting time in the system in steady state at the arrival instants thus behaves as the waiting time in an M/M/1 queue with arrival rate δ and service rate $1/\gamma = [\mu/(1 - \alpha)]$. However, the arrival rate to the system is actually $\lambda = [\mu\delta/(\mu + \alpha\delta)]$ and the service rate is clearly μ . Comparing our system with an M/M/1 system with parameters λ, μ , we observe that both systems have the same stability condition. However, in the M/M/1 system, the term that governs the exponent (with a minus sign) of the p.d.f. is $\mu(1 - [\delta/(\mu + \alpha\delta)])$. In our system, the term is $\mu([1/(1 - \alpha)] - \delta/\mu)$, which is larger when the stability condition holds. This implies that the tail probability decreases much faster in our system when compared to the corresponding M/M/1 system.

An extension to this system where J_n can also take negative values (but the interarrival period is, of course, kept positive) appears in the Appendix.

4. RANDOM PROPORTIONAL DEPENDENCY

The analysis of this system is more involved than the previous one, and we use the spectral analysis method described in Kleinrock [17]. We begin by rewriting Eq. (5) as

$$W_{n+1} = (W_n + U_n)^+,$$

where $U_n = (1 - \Omega_n)B_n$ and Ω_n is a r.v. that takes the values α_i with probability a_i for $1 \leq i \leq N + M$, where $1 < \alpha_1 < \alpha_2 < \dots < \alpha_N$ and $\alpha_{N+1} < \alpha_{N+2} < \dots < \alpha_{N+M} \leq 1$. Because B_n is exponentially distributed, the LST of U_n , denoted by $\mathfrak{U}(s)$ is given by

$$\mathfrak{U}(s) = \sum_{i=1}^{N+M} \frac{a_i}{1 - \gamma_i s}, \quad (10)$$

where $\gamma_i = (\alpha_i - 1)/\mu$.

Following the "spectrum factorization" approach (see Eqs. (8.35) and (8.36) in Kleinrock [17]), we need to factor $\mathfrak{U}(s) - 1$ as follows:

$$\mathfrak{U}(s) - 1 = \frac{\Psi_+(s)}{\Psi_-(s)}, \quad (11)$$

where $\Psi_+(s)$ is an analytic function of s for $\text{Re}(s) > 0$ with no zeros in this half plane and $\Psi_-(s)$ is an analytic function of s for $\text{Re}(s) < D$ (D is a positive constant to be determined later) with no zeros in this half plane.

Once we have the preceding factorization, we immediately obtain the LST of the waiting time, up to a constant (see Eq. (8.41) in Kleinrock [17]), that is,

$$\mathcal{W}(s) = \frac{L \cdot s}{\Psi_+(s)}, \quad (12)$$

and the constant L is determined from the normalization condition $\lim_{s \rightarrow 0} \mathcal{W}(s) = 1$.

To form the factorization in Eq. (11), we write

$$\mathcal{U}(s) - 1 = \frac{\mathcal{N}(s)}{\mathcal{V}(s)},$$

where from Eq. (10)

$$\begin{aligned} \mathcal{N}(s) &= \sum_{j=1}^{N+M} a_j \prod_{\substack{i=1 \\ i \neq j}}^{N+M} (1 - \gamma_i s) - \prod_{i=1}^{N+M} (1 - \gamma_i s) \\ \mathcal{V}(s) &= \prod_{i=1}^{N+M} (1 - \gamma_i s) \end{aligned}$$

and we have to study the location of the roots of the polynomials $\mathcal{N}(s)$ and $\mathcal{V}(s)$. Let K be the degree of these polynomials. Then, if $\alpha_{N+M} \neq 1$ we have $K = N + M$, and if $\alpha_{N+M} = 1$ we have $K = N + M - 1$.

From the definition of $\mathcal{V}(s)$, it is easy to see that the roots of this polynomial are γ_i^{-1} , $1 \leq i \leq K$, and they are all distinct. Furthermore, for $1 \leq i \leq N$ these roots are positive, whereas for $N + 1 \leq i \leq K$ they are negative. In the following theorem, we determine the location of the roots of $\mathcal{N}(s)$.

THEOREM 1: *If the stability condition holds, that is, $\sum_{j=1}^{N+M} a_j \alpha_j > 1$, then $\mathcal{N}(s)$ has $N - 1$ distinct positive real roots, a single root at $s = 0$ and $K - N$ distinct negative real roots.*

PROOF: For $1 \leq k \leq K$, define

$$\Gamma_k = \gamma_k^{-1} = \frac{\mu}{\alpha_k - 1}. \quad (13)$$

Therefore, $\Gamma_1 > \Gamma_2 > \dots > \Gamma_N > 0 > \Gamma_{N+1} > \Gamma_{N+2} > \dots > \Gamma_K$.

Computing the polynomial $\mathcal{N}(s)$ at $s = \Gamma_k$, we obtain for $1 \leq k \leq K$

$$\mathcal{N}(\Gamma_k) = a_k \prod_{\substack{i=1 \\ i \neq k}}^K (1 - \gamma_i \Gamma_k) = a_k \prod_{\substack{i=1 \\ i \neq k}}^K \left(1 - \frac{\Gamma_k}{\Gamma_i} \right). \quad (14)$$

The quantity $\mathcal{N}(\Gamma_N)$ is positive because $\Gamma_i > \Gamma_N > 0$ for any $1 \leq i \leq N - 1$. The quantity $\mathcal{N}(\Gamma_{N-1})$ is negative as $1 - \Gamma_N/\Gamma_{N-1}$ is the only negative term in the preceding product. In fact, the sign of $\mathcal{N}(\Gamma_k)$ for $1 \leq k \leq N$ is $(-1)^{N-k}$ and therefore $\mathcal{N}(\Gamma_N)$, $\mathcal{N}(\Gamma_{N-1})$, \dots , $\mathcal{N}(\Gamma_1)$ have alternating signs. Similarly, $\mathcal{N}(\Gamma_{N+1})$ is positive because $\Gamma_i < \Gamma_{N+1} < 0$ for any $N + 2 \leq i \leq K$. Again, we see that the sign of $\mathcal{N}(\Gamma_k)$ for $N + 1 \leq k \leq K$ is $(-1)^{k-N+1}$ and therefore

$\mathcal{N}(\Gamma_{N+1}), \mathcal{N}(\Gamma_{N+2}), \dots, \mathcal{N}(\Gamma_K)$ have alternating signs. Consequently, there are at least $N - 1$ distinct positive real roots, one root at $s = 0$, and at least $K - N - 1$ distinct negative real roots.

Because both $\mathcal{N}(\Gamma_{N+1})$ and $\mathcal{N}(\Gamma_N)$ are positive and $\mathcal{N}(0) = 0$, it is clear that the additional real root of $\mathcal{N}(s)$ is either between Γ_{N+1} and 0 or between 0 and Γ_N . This is true because $\mathcal{N}(s)$ is a polynomial of degree K . The location of that additional root is determined by computing the derivative of $\mathcal{N}(s)$ with respect to s at $s = 0$:

$$\left. \frac{d\mathcal{N}(s)}{ds} \right|_{s=0} = \sum_{i=1}^{N+M} \gamma_i - \sum_{j=1}^{N+M} a_j \sum_{\substack{i=1 \\ i \neq j}}^{N+M} \gamma_i = \frac{1}{\mu} \sum_{j=1}^{N+M} a_j (\alpha_j - 1) > 0,$$

where the last inequality follows from the stability condition $\sum_{j=1}^{N+M} a_j \alpha_j > 1$.

Because the derivative is positive and $\Gamma_{N+1} > 0$, the additional root must be between Γ_{N+1} and 0 and therefore $\mathcal{N}(s)$ has $K - N$ negative roots and $N - 1$ positive roots that were already proven to be real and distinct. This completes the proof of Theorem 1. ■

Let the distinct roots of $\mathcal{N}(s)$ be denoted by $\sigma_1, \sigma_2, \dots, \sigma_K$, with $\sigma_1 > \sigma_2 > \dots > \sigma_K$ and $\sigma_N = 0$. Note that from the proof of Theorem 1 we know that $\Gamma_i > \sigma_i > \Gamma_{i+1}$ for $1 \leq i \leq K - 1$. Therefore, $\sigma_j \gamma_i \neq 1$ for $1 \leq i, j \leq K$.

From Theorem 1, we conclude that $\mathcal{U}(s) - 1$ can be factorized as stated in Eq. (11) with

$$\Psi_+(s) = \frac{\prod_{i=N}^K (s - \sigma_i)}{\prod_{i=N+1}^K (1 - \gamma_i s)},$$

$$\Psi_-(s) = - \frac{\prod_{i=1}^N (1 - \gamma_i s)}{\left(\prod_{i=1}^K (-\gamma_i) \right) \left(\prod_{i=1}^{N-1} (s - \sigma_i) \right)}.$$

Because $\gamma_i < 0$ and $\sigma_i < 0$ for $N + 1 \leq i \leq K$, it follows that $\Psi_+(s)$ is analytic for $\text{Re}(s) > 0$ and has no zeros in this half plane, as required. Similarly, because $\gamma_i > 0$ for $1 \leq i \leq N$ and $\sigma_i > 0$ for $1 \leq i \leq N - 1$, it follows that $\Psi_-(s)$ is analytic for $\text{Re}(s) < \gamma_N^{-1}$ and has no zeros in this half plane (so $D = \gamma_N^{-1}$ in this case).

Using Eq. (12), we obtain

$$\mathcal{Q}(s) = \frac{L \prod_{i=N+1}^K (1 - \gamma_i s)}{\prod_{i=N+1}^K (s - \sigma_i)},$$

and from the normalization condition we have that $L = \prod_{i=N+1}^K (-\sigma_i)$. Note that L is also the probability that an arriving packet need not queue. In summary, the LST $\mathcal{Q}_W(s)$ of the waiting time is given by

$$\mathcal{Q}_W(s) = \frac{\left(\prod_{i=N+1}^K (-\sigma_i) \right) \left(\prod_{i=N+1}^K (1 - \gamma_i s) \right)}{\prod_{i=N+1}^K (s - \sigma_i)}. \tag{15}$$

Note that it is very easy to invert the preceding LST to obtain the distribution of the waiting time. Taking the n th derivative of $\mathcal{Q}_W(s)$ with respect to s at $s = 0$, we obtain the n th moment of the waiting time. In particular, the expected waiting time is given by

$$E[W] = \sum_{i=N+1}^K (\gamma_i - \sigma_i^{-1}),$$

and the second moment is given by

$$E[W^2] = \sum_{i=N+1}^K \sum_{\substack{j=N+1 \\ j \neq i}}^K (\gamma_i \gamma_j - (\sigma_i \sigma_j)^{-1}) - 2 \sum_{i=N+1}^K \sigma_i^{-1} \sum_{i=N+1}^K (\gamma_i - \sigma_i^{-1}).$$

Then, the variance is obtained from the standard formula,

$$\text{Var}(W) = E[W^2] - (E[W])^2 \tag{16}$$

Remark 1: The computational effort in determining the LST of $\mathcal{Q}_W(s)$ lies in the determination of the roots σ_i , $N + 1 \leq i \leq K$. However, because each root is known to be real and because we know that $\Gamma_{N+1} < \sigma_{N+1} < 0$ and $\Gamma_i < \sigma_i < \Gamma_{i-1}$ for $N + 2 \leq i \leq K$, it is very easy to determine the roots with any simple search procedure.

Remark 2: The preceding analysis does not depend on the values of α_i and μ separately but on the values and the sign of the parameters $\gamma_i = (\alpha_i - 1)/\mu$. Consequently, it is easy to generalize our system to the case where μ itself is a r.v. and with probability a_{ij} it takes the value μ_j whereas Ω_n takes the value α_i . This is equivalent to the case where with probability a_{ij} we define $\gamma_{ij} = (\alpha_i - 1)/\mu_j$ and adjust Eq. (10) such that the sum is taken over all values γ_{ij} .

Remark 3: Let $W \triangleq \lim_{n \rightarrow \infty} W_n$, $U \triangleq \lim_{n \rightarrow \infty} U_n$. For a G/G/1 system with $\rho \cong 1$ (but remaining strictly less than 1, preserving stability), one can show that (see Kingman [15]),

$$\Pr[W \leq w] \cong 1 - \exp \left\{ \frac{2E[U]}{E[U^2]} w \right\}. \tag{17}$$

Recall that $U \triangleq B - I$, where B, I are the generic r.v.'s of the service time and the interarrival time, respectively. Denote by W^{ind} the generic r.v. of the waiting

time in an equivalent GI/G/1 system (in which the r.v.'s B and I are independent). From Eq. (17), $W \leq_{st} W^{ind}$ (where $A \leq_{st} B$ means that A is stochastically smaller than B) if and only if $E[B \cdot I] - E[B]E[I] \geq 0$. That is, in heavy traffic, the waiting time in a G/G/1 system with service and interarrival times *positively* correlated is stochastically smaller than in the equivalent GI/G/1 system. We conjecture that this fact holds not only in heavy traffic but for any load. For our system with random proportional dependency, we have $E[B \cdot I] - E[B]E[I] = (1/\mu^2) \sum_{i=1}^K a_i \alpha_i$, and in the following numerical examples we show that the average and the variance of the waiting time in the equivalent GI/M/1 system are smaller than in our system for all considered loads.

An upper bound on the average wait for a GI/G/1 system was developed in Kingman [16]. Using the same techniques, we obtain an upper bound for a G/G/1 system (and, hence, for our system with random proportional dependency). We have

$$E[W] \leq \frac{\text{Var}(W)}{-2E[U]} = \frac{\sum_{i=1}^K a_i (1 - \alpha_i)^2}{2\mu \sum_{i=1}^K a_i (\alpha_i - 1)}. \quad (18)$$

The following lower bound on the average wait was developed in Kingman [16]:

$$E[W] \geq \frac{E[(U)^+]^2}{-2E[U]} = \frac{\sum_{i=N+1}^K a_i (1 - \alpha_i)^2}{\mu \sum_{i=1}^K a_i (\alpha_i - 1)}. \quad (19)$$

Another lower bound on the average wait for our system can be obtained using similar techniques as exercise 2.8 in Kleinrock [18]

$$\begin{aligned} E[W] &\geq \frac{\text{Var}(W)}{-2E[U]} - \frac{E[U]}{2} - \alpha_N + 1 \\ &= \frac{\sum_{i=1}^K a_i (1 - \alpha_i)^2}{2\mu \sum_{i=1}^K a_i (\alpha_i - 1)} + \frac{\sum_{i=1}^K a_i (\alpha_i - 1)}{2\mu} - \alpha_N + 1 \end{aligned} \quad (20)$$

and for small α_N this is a tight bound (see Eq. (18)).

In Figures 2-5, we depict the average and the variance of the waiting time of a system with random proportional dependency. For comparison purposes, we also depict the same quantities in an equivalent GI/M/1 system in which the service time is exponentially distributed with parameter μ , and the interarrival times are independent and sampled from a probability distribution whose LST

is $\mathcal{A}(s) = \sum_{i=1}^{N+M} a_i \frac{\mu/\alpha_i}{\mu/\alpha_i + s}$. Namely, with probability a_i , the interarrival time is exponentially distributed with parameter μ/α_i . Recall that the analysis of this GI/M/1 system requires the determination of a single root between 0 and 1 of the equation $s = \mathcal{A}(\mu - \mu s)$. In all figures, we consider a system with $N = 3$, $M = 2$, and $a_i = 0.2$, $1 \leq i \leq 5$.

In Figures 2 and 3, we use $\alpha_1 = 1.4$, $\alpha_2 = 1.6$, $\alpha_4 = 0.1$, and $\alpha_5 = 0.2$, and we depict the average and the variance of the waiting time versus the largest proportional parameter α_3 . As expected, both quantities are decreasing with increasing α_3 . In Figures 4 and 5, we keep $\sum_{i=1}^{N+M} a_i \alpha_i$ constant. In particular, we use $\alpha_1 = 1.2$, $\alpha_2 = 1.3$, and $\alpha_4 = 0.1$, and $\alpha_3 + \alpha_5$ is kept constant. The average and the variance of the waiting time are depicted as a function of the largest proportional parameter α_3 . It is interesting to note that, although $\alpha_3 + \alpha_5$ is kept constant, both the average and the variance increase with increasing α_3 (and decreasing α_5). This implies that decreasing α_5 has a more pronounced effect on the performance of the queuing system. The reason is that increasing α_3 and decreasing α_5 while keeping their sum constant increases the variability of the arrival process and, hence, increases the average and the variance of the waiting time, as shown in Figures 4 and 5.

In all cases, we observe that the equivalent GI/M/1 system exhibits much larger averages and variances of the waiting time. This implies that correlations

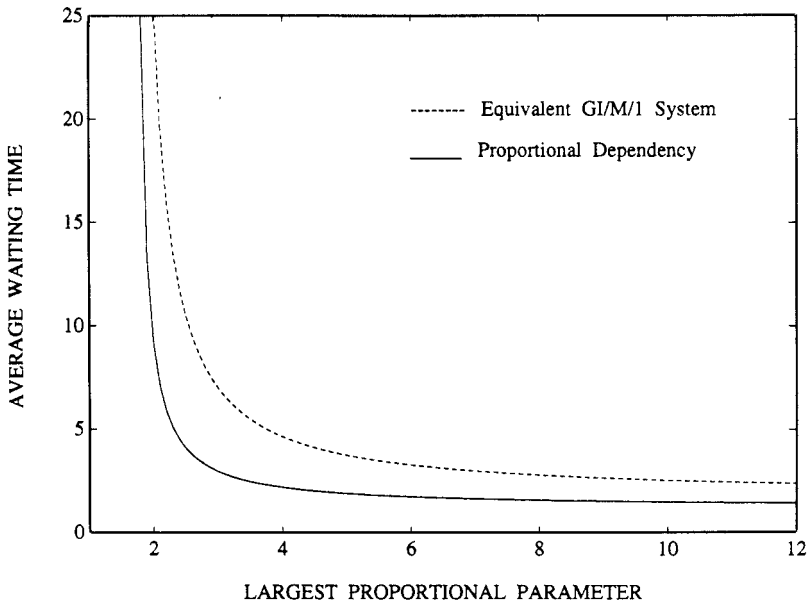


FIGURE 2. Average waiting time versus largest proportional parameter: $a_i = 0.2$, $1 \leq i \leq 5$; $\alpha_1 = 1.4$, $\alpha_2 = 1.6$, $\alpha_4 = 0.2$, $\alpha_5 = 0.1$.

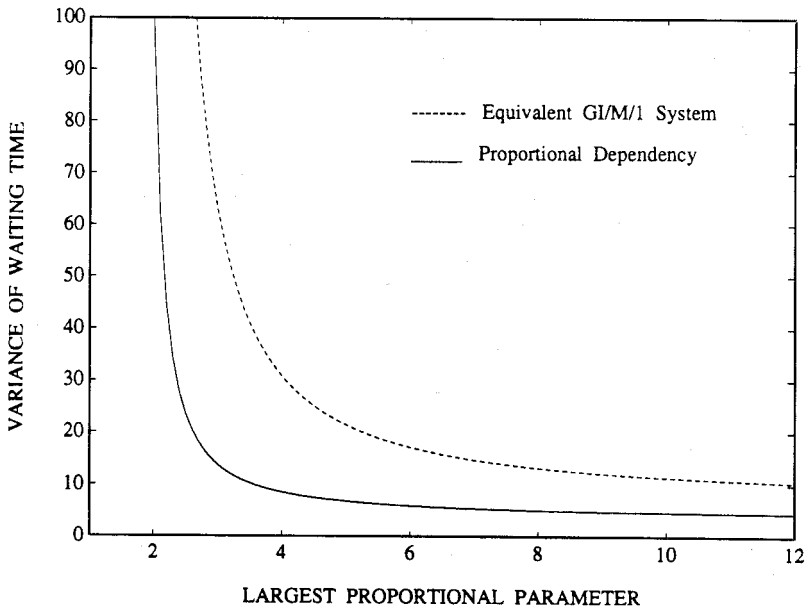


FIGURE 3. Variance of waiting time versus largest proportional parameter: $a_i = 0.2$, $1 \leq i \leq 5$; $\alpha_1 = 1.4$, $\alpha_2 = 1.6$, $\alpha_4 = 0.2$, $\alpha_5 = 0.1$.

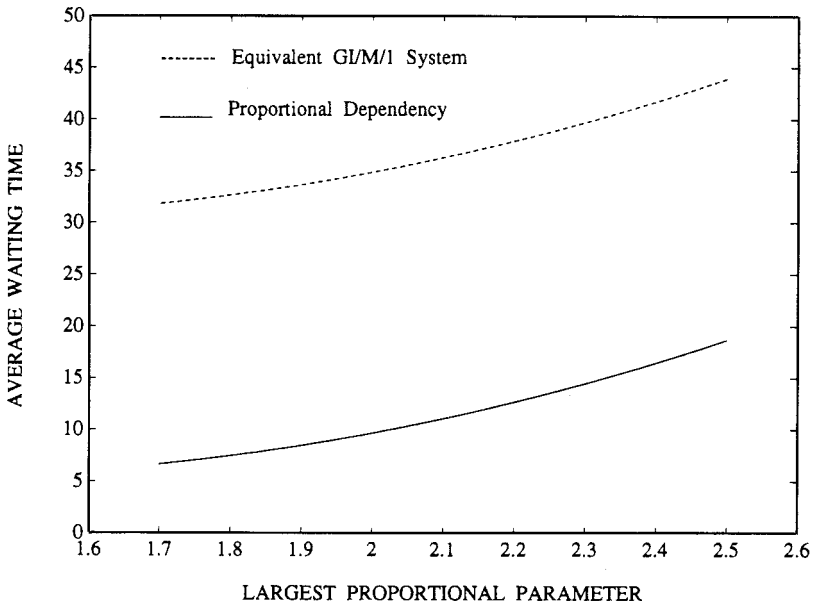


FIGURE 4. Average waiting time versus largest proportional parameter: $\alpha_i = 0.2$, $1 \leq i \leq 5$; $\alpha_1 = 1.2$, $\alpha_2 = 1.3$, $\alpha_4 = 0.1$, $\alpha_3 + \alpha_5 = 2.6$.

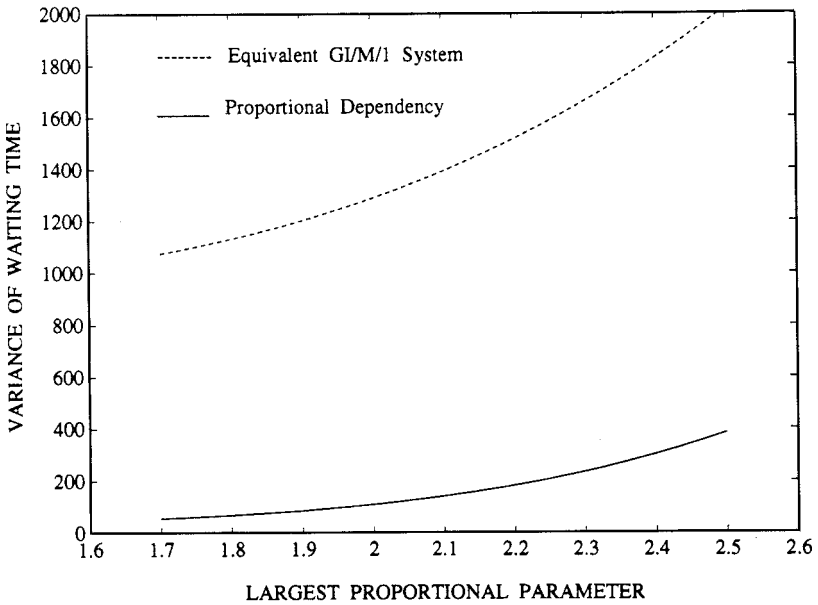


FIGURE 5. Variance of waiting time versus largest proportional parameter: $a_i = 0.2, 1 \leq i \leq 5; \alpha_1 = 1.2, \alpha_2 = 1.3, \alpha_4 = 0.1, \alpha_3 + \alpha_5 = 2.6.$

between service times and interarrival times have a smoothing effect on the system. This has also been observed in Cidon et al. [4], Conolly [5], Conolly and Choo [6], and Conolly and Hadidi [7,8] for different types of correlations.

5. RANDOM PROPORTIONAL DEPENDENCY WITH ADDITIVE DELAY

We now combine the two models of Sections 3 and 4. As before, we start by rewriting Eq. (6) as

$$W_{n+1} = (W_n + U_n)^+,$$

where $U_n = (1 - \Omega_n)B_n - J_n$, Ω_n is defined in Section 4, and J_n is an independent exponentially distributed r.v. with parameter δ . The LST of U_n , denoted by $\mathcal{U}(s)$, is given in this case by (recall that $\gamma_i = (\alpha_i - 1)/\mu$)

$$\mathcal{U}(s) = \frac{\delta}{\delta - s} \sum_{i=1}^{N+M} \frac{a_i}{1 - \gamma_i s}. \tag{21}$$

We now need to factorize $\mathcal{U}(s) - 1$ as in Eq. (11) and then to obtain the LST of the waiting time as in Eq. (12). To that end, we write

$$\mathcal{U}(s) - 1 = \frac{\mathcal{N}(s)}{\mathcal{Q}(s)},$$

where, from Eq. (21),

$$\mathcal{N}(s) = \delta \sum_{j=1}^{N+M} a_j \prod_{\substack{i=1 \\ i \neq j}}^{N+M} (1 - \gamma_i s) - (\delta - s) \prod_{i=1}^{N+M} (1 - \gamma_i s),$$

$$\mathcal{V}(s) = (\delta - s) \prod_{i=1}^{N+M} (1 - \gamma_i s),$$

and we have to study the location of the roots of the polynomials $\mathcal{N}(s)$ and $\mathcal{V}(s)$. Let K be the degree of these polynomials. Then, if $\alpha_{N+M} \neq 1$, we have $K = N + M + 1$, and if $\alpha_{N+M} = 1$ we have $K = N + M$.

From the definition of $\mathcal{V}(s)$, it is easy to see that the roots of this polynomial are γ_i^{-1} , $1 \leq i \leq K - 1$, and δ . Furthermore, for $1 \leq i \leq N$, the roots γ_i^{-1} are positive, whereas for $N + 1 \leq i \leq K - 1$, the roots γ_i^{-1} are negative. The root at δ is also positive. In the following theorem, we determine the location of the roots of $\mathcal{N}(s)$.

THEOREM 2: *If the stability condition holds, that is, $\sum_{j=1}^{N+M} a_j \alpha_j + \mu/\delta > 1$, then $\mathcal{N}(s)$ has N distinct positive real roots, a single root at $s = 0$, and $K - N$ distinct negative real roots.*

PROOF: We can apply exactly the same technique as in the proof of Theorem 1 to show the existence of at least $N - 1$ distinct positive real roots, a single root at $s = 0$, and at least $K - N - 2$ distinct negative real roots of the polynomial $\mathcal{N}(s)$ because the last term of $\mathcal{N}(s)$ becomes zero when $s = \Gamma_k$ (recall that $\Gamma_k = \gamma_k^{-1}$). Similarly, the location of another negative real root between Γ_{N+1} and 0 can be proved because the derivative of $\mathcal{N}(s)$ is positive at $s = 0$ when the stability condition holds. Therefore, we only need to show that there exists another positive root of $\mathcal{N}(s)$, which is different from the ones found before. Recall that the sign of $\mathcal{N}(\Gamma_k)$ is $(-1)^{N-k}$ for $1 \leq k \leq N$. This is true here because the sign of $\mathcal{N}(\Gamma_k)$ is only determined by the sign of the first term of $\mathcal{N}(s)$ ($\sum_{j=1}^{N+M} a_j \delta \prod_{i=1, i \neq j}^{N+M} (1 - \gamma_i s)$) because the last term ($(\delta - s) \prod_{i=1}^{N+M} (1 - \gamma_i s)$) vanishes at all these points. Therefore, the sign of $\mathcal{N}(\Gamma_1)$ (where Γ_1 is the largest) is $(-1)^{N-1}$. Now, we will explore the sign of $\mathcal{N}(s)$ as $s \rightarrow \infty$. By rewriting $\mathcal{N}(s)$ as

$$\mathcal{N}(s) = \prod_{i=1}^{N+M} \left(1 - \frac{s}{\Gamma_i} \right) \left[s - \delta + \sum_{j=1}^{N+M} \frac{a_j \delta}{1 - \gamma_j s} \right],$$

it becomes clear that the sign of $\mathcal{N}(s)$ as $s \rightarrow \infty$ is $(-1)^N$ because $\Gamma_i \leq 0$ for $N + 1 \leq i \leq K - 1$ and $\Gamma_i > 0$ for $1 \leq i \leq N$. Therefore, there must be another positive real root of $\mathcal{N}(s)$ between Γ_1 and ∞ . There is only one root in this region because the total number of roots of $\mathcal{N}(s)$ is K . ■

Let the distinct roots of $\mathcal{N}(s)$ be denoted by $\sigma_1, \sigma_2, \dots, \sigma_K$, with $\sigma_1 > \sigma_2 > \dots > \sigma_K$ and $\sigma_{N+1} = 0$. Note that from the proof of Theorem 2 we know that $\sigma_j \gamma_j \neq 1$ for $1 \leq j \leq K$.

From Theorem 2, we conclude that $\mathcal{U}(s) - 1$ can be factorized as stated in Eq. (11) with

$$\Psi_+(s) = \frac{\prod_{i=N+1}^K (s - \sigma_i)}{\prod_{i=N+1}^K (1 - \gamma_i s)},$$

$$\Psi_-(s) = -\frac{(\delta - s) \prod_{i=1}^N (1 - \gamma_i s)}{\left(\prod_{i=1}^K (-\gamma_i)\right) \left(\prod_{i=1}^N (s - \sigma_i)\right)}.$$

Using Eq. (12), we obtain

$$\mathcal{Q}(s) = \frac{L \prod_{i=N+1}^K (1 - \gamma_i s)}{\prod_{i=N+2}^K (s - \sigma_i)}, \tag{22}$$

and from the normalization condition we have that $L = \prod_{i=N+2}^K (-\sigma_i)$. Note that L is also the probability that an arriving packet need not queue. In summary, the LST $\mathcal{Q}(s)$ of the waiting time is given by

$$\mathcal{Q}(s) = \frac{\left(\prod_{i=N+2}^K (-\sigma_i)\right) \left(\prod_{i=N+1}^K (1 - \gamma_i s)\right)}{\prod_{i=N+2}^K (s - \sigma_i)}. \tag{23}$$

Note that this expression is very similar to Eq. (15) except that the locations of the roots σ_i will be different in this case.

Taking the corresponding derivatives, we obtain the expected waiting time

$$E[W] = \gamma_{N+1} \sum_{i=N+2}^K (\gamma_i - \sigma_i^{-1})$$

and the second moment

$$E[W^2] = \sum_{i=N+2}^K \sum_{\substack{j=N+2 \\ j \neq i}}^K (\gamma_i \gamma_j - (\sigma_i \sigma_j)^{-1}) - 2 \sum_{i=N+2}^K \sigma_i^{-1} \sum_{i=N+2}^K (\gamma_i - \sigma_i^{-1})$$

$$- 2\gamma_{N+1} \sum_{i=N+2}^K (\gamma_i - \sigma_i^{-1}).$$

6. SUMMARY

This paper presents the analysis of the customer waiting time in systems where the interarrival time I_{n+1} between customers n and $n + 1$ depends on the service time B_n of customer n through a *proportionality* relation and B_n is an exponentially distributed r.v. We considered different scenarios of increasing complexity and provided efficient computational methods for their analysis. The general conclusion, which confirms the initial intuition, is that such correlations have a smoothing effect on the behavior of the queueing systems and emphasizes the general consensus that independence assumptions are usually pessimistic in practical environments.

Our motivation for investigating such systems originated from rate control policies that are popular in new high-speed network architectures. However, the models developed in the paper and the associated solutions are of general interest and potentially applicable to other environments. Extensions of these models to service times with general distributions and for proportional parameters that are continuous random variables are currently under study.

References

1. Bala, K., Cidon, I., & Sohraby, K. (1990). Congestion control for high-speed packet switch. *Proceedings of INFOCOM '90*, San Francisco, June: 520–526.
2. Boxma, O.J. (1979). On a tandem queueing model with identical service times at both counters, Pts. I and II. *Advances in Applied Probability* 11: 616–659.
3. Cidon, I., Gopal, I., & Guérin, R. (1991). Bandwidth management and congestion control in plaNET. *IEEE Communications Magazine* 29(10): 54–63.
4. Cidon, I., Guérin, R., Khamisy, A., & Sidi, M. (1993). Analysis of a correlated queue in a communication system. *IEEE Transactions on Information Theory* 39(2): 456–465.
5. Conolly, B.W. (1968). The waiting time process for a certain correlated queue. *Operations Research* 16: 1006–1015.
6. Conolly, B.W. & Choo, Q.H. (1979). The waiting time process for a generalized correlated queue with exponential demand and service. *SIAM Journal on Applied Mathematics* 37(2): 263–275.
7. Conolly, B.W. & Hadidi, N. (1969). A comparison of the operational features of conventional queues with a self-regulating system. *Applied Statistics* 18: 41–53.
8. Conolly, B.W. & Hadidi, N. (1969). A correlated queue. *Journal of Applied Probability* 6: 122–136.
9. Elwalid, A.I. & Mitra, D. (1991). Analysis and design of rate-based congestion control of high-speed networks, I: Stochastic fluid models, access regulation. *Queueing Systems Theory and Applications (QUESTA)* 9: 29–64.
10. Fendick, K.W., Saksena, V.R., & Whitt, W. (1989). Dependence in packet queues. *IEEE Transactions on Communications* COM-37(11): 1173–1183.
11. Ghahramani, S. & Wolff, R.W. (1989). A new proof of finite moment conditions for GI/G/1 busy periods. *Queueing Systems Theory and Applications (QUESTA)* 4(2): 171–178.
12. Guérin, R. & Gün, L. (1992). A unified approach to bandwidth allocation and access control in fast packet-switched networks. *Proceedings of INFOCOM '92*, Florence, May, pp. 1–12.
13. Hadidi, N. (1981). Queues with partial correlation. *SIAM Journal on Applied Mathematics* 40(3): 467–475.
14. Hadidi, N. (1985). Further results on queues with partial correlations. *Operations Research* 33: 203–209.

15. Kingman, J.F.C. (1962). On queues in heavy traffic. *Journal of the Royal Statistical Society* 24: 383-392.
16. Kingman, J.F.C. (1962). Some inequalities for the queue GI/G/1. *Biometrika* 49: 315-324.
17. Kleinrock, L. (1975). *Queueing systems, Vol. 1: Theory*. New York: John Wiley & Sons.
18. Kleinrock, L. (1976). *Queueing systems, Vol. 2: Computer applications*. New York: John Wiley & Sons.
19. Langaris, C. (1986). A correlated queue with infinitely many servers. *Journal of Applied Probability* 23: 155-165.
20. Langaris, C. (1987). Busy-period analysis of a correlated queue with exponential demand and service. *Journal of Applied Probability* 24: 476-485.
21. Loynes, R.M. (1962). The stability of a queue with non-independent inter-arrival and service times. *Proceedings of the Cambridge Philosophical Society* 58(3): 497-520.

APPENDIX: PROPORTIONAL DEPENDENCY WITH ADDITIVE AND SUBTRACTIVE DELAY

In this Appendix, we extend the results of Section 3 to the case where the packet delay jitter caused by previous traveling through the network can take both positive and negative values. This means that packets can arrive sooner or later than expected based only on the service time. However, because most networks preserve a first-come first-served ordering, we do not allow the interarrival time to be negative. Because in our case $I_{n+1} = \alpha B_n + J_n$, the LST $\mathcal{Q}_{n+1}(s)$ can be expressed as

$$\mathcal{Q}_{n+1}(s) = E[e^{-s[W_n + B_n - (\alpha B_n + J_n)^+]^+}],$$

where $\alpha < 1$ and J_n is a random variable to be defined. As long as J_n takes only positive values, the term $(\alpha B_n + J_n)$ is always positive and, hence, is identical to $(\alpha B_n + J_n)^+$. Here we allow J_n to take also negative values (but I_{n+1} is kept nonnegative). To that end, we assume that with probability a J_n is an additive exponentially distributed r.v. (denoted by \hat{J}_n) with parameter δ (as in Section 3) and with probability $(1 - a)$ a subtractive exponentially distributed r.v. with parameter ν (denoted by \tilde{J}_n). Note that because $\alpha < 1$ then $[W_n + B_n - (\alpha B_n - \tilde{J}_n)^+]^+ = [W_n + B_n - (\alpha B_n - \tilde{J}_n)^+]$. Hence,

$$\mathcal{Q}_{n+1}(s) = aE[e^{-s[W_n + (1-\alpha)B_n - \hat{J}_n]^+}] + (1 - a)E[e^{-s[W_n + B_n - (\alpha B_n - \tilde{J}_n)^+]^+}], \quad (24)$$

$$\begin{aligned} E[e^{-s[W_n + B_n - (\alpha B_n - \tilde{J}_n)^+]^+}] &= \int_0^\infty f_{W_n}(w) dw \int_0^\infty \mu e^{-\mu x} dx \\ &\quad \cdot \left[\int_0^{\alpha x} \nu e^{-\nu y} e^{-s[w + (1-\alpha)x + y]} dy + \int_{\alpha x}^\infty \nu e^{-\nu y - s(w+x)} dy \right], \\ &= \frac{\mu\nu}{(\nu + s)(\mu + (1-\alpha)s)} \mathcal{Q}_n(s) - \frac{\mu\nu}{(\nu + s)(\mu + \nu\alpha + s)} \mathcal{Q}_n(s) \\ &\quad + \frac{\mu}{\mu + \nu\alpha + s} \mathcal{Q}_n(s) \\ &= \frac{\mu s(1 + \gamma s) + \nu(\mu + \nu\alpha + s)}{(\nu + s)(\mu + \nu\alpha + s)(1 + \gamma s)} \mathcal{Q}_n(s), \end{aligned} \quad (25)$$

where $\gamma = (1 - \alpha)/\mu$.

Applying Eqs. (25) and (7) in Eq. (24) and letting $n \rightarrow \infty$, we get

$$\begin{aligned} \mathcal{Q}(s) &= \frac{as}{(s - \delta)(1 + \gamma\delta)} \mathcal{Q}(\delta) - \frac{a\delta}{(s - \delta)(1 + \gamma s)} \mathcal{Q}(s) \\ &\quad + (1 - a) \frac{\mu s(1 + \gamma s) + \nu(\mu + \nu\alpha + s)}{(\nu + s)(\mu + \nu\alpha + s)(1 + \gamma s)} \mathcal{Q}(s). \end{aligned} \quad (26)$$

Rearranging Eq. (26) and applying the normalization condition (${}^{\circ}W(0) = 1$), we get

$${}^{\circ}W(s) = \frac{{}^{\circ}W(\delta)}{1 + \gamma\delta}$$

$$\frac{a(\mu + \nu - \gamma\mu\nu + s)(1 + \gamma s)}{a[\mu(1 - \delta\gamma - \gamma\nu) + \nu] - \delta[1 - a + \nu\gamma(1 - \gamma\mu)] + [1 + \gamma(a\mu - \delta + \nu + \gamma\mu\nu)]s + \gamma^2 s^2}$$

with

$${}^{\circ}W(\delta) = \frac{(1 + \gamma\delta)[- \delta(1 - a) + a\mu - a\delta\gamma\mu + a\nu - \delta\gamma\nu - a\gamma\mu\nu + \delta\gamma^2\mu\nu]}{a(\mu + \nu - \gamma\mu\nu)}. \quad (27)$$

The stability condition is derived from $E[B_n - (\alpha B_n - J_n)^+] < 0$ and is $\gamma - a\delta + (1 - a)(1/\nu - \mu/[\nu(\mu + \nu\alpha)]) < 0$. It can be verified that this condition is equivalent to ${}^{\circ}W(\delta) > 0$.