The Effect of Correlated Service and Inter-arrival Times on System Performance

Gang Uk Hwang

Abstract—In communication networks where communication nodes are connected with finite capacity transmission links, the packet inter-arrival times are strongly correlated with the packet length and the link capacity (or the packet service time). Such correlation affects the system performance significantly, but little attention has been paid to this issue. In this paper, we propose a mathematical framework to study the impact of the correlation between the packet service times and the packet inter-arrival times on system performance. With our mathematical model, we analyze the system performance, e.g., the unfinished work of the system, and show that the correlation affects the system performance significantly. Some numerical examples are also provided.

Keywords—Performance analysis, Correlated queueing system, Unfinished work, PH-type distribution, Communication system.

I. INTRODUCTION

IN communication networks where communication nodes are connected with finite transmission capacity links, the packet inter-arrival times depend on both the packet length and the link capacity. For instance, consider a link of transmission capacity τ, connecting two nodes. The link can not send more than $\tau \times T$ amount of packets during a time period of length T. Consequently, large packets need more service times (i.e., transmission times) than short packets and the inter-arrival times of large packets are longer than those of short packets. Such correlation between the packet service times and the packet inter-arrival times affects the system performance, but most studies have not paid their attention to this issue.

To study the impact of the correlation between the packet service times and the packet inter-arrival times on system performance, we consider two nodes, say, node 1 and node 2, communicating with each other through a finite capacity link. Node 1 generates a packet and transmits the packet to node 2 after its generation. So, the packet arrival process at node 2 can be modelled by an alternating renewal process having two types of periods, say packet generation times and packet service times. Packet generation times capture the times needed for generating new packets at node 1 and packet service times capture the packet transmission times at node 1 through the link. Since node 2 can service an arriving packet after it receives the last bit of the packet, we assume that packet arrivals occur at node 2 at the ends of packet service times. Since an inter-arrival time of two consecutive packets consists of a packet generation time and a

G.U. Hwang is with Division of Applied Mathematics and Telecommunication Program, Korea Advanced Institute of Science and Technology, 373-1 Kuseong-dong, Yuseong-gu, Taejon, 305-701, South Korea (E-mail: guhwang@amath.kaist.ac.kr)

packet transmission time, there is a positive correlation between the packet inter-arrival times and the packet service times. In order to attain the generality of our model, we assume that packet generation times are according to a PH-type distribution and packet service times are according to a general distribution. In analysis, we consider a queue in node 2 which accommodates packets from node 1 and has the same service capacity as the link between node 1 and node 2 (i.e., node 1 and node 2 have the same service capacity). We construct a discrete time Markov chain and analyze the performance of the queue, e.g., the unfinished work of the queueing system. Based on our analysis, we provide some numerical results to see the impact of the correlated service and inter-arrival times on system performance. Our numerical results show that the correlation significantly affects the system performance.

Earlier works related to this issue are found in [1], [2] and the references therein. In [1] they considered a system with Poisson arrivals, where the service time B_n of the nth packet is proportional to the inter-arrival time A_n between the $n - 1$st and nth packets, i.e., $B_n = \zeta A_n$ where ζ is a positive constant. They obtained the LST (Laplace Stieltjes Transform) for the system time, defined by the waiting time in the queue plus the service time, by solving a linear functional equation [3] derived from the evolution equation of the system time. However, they only consider the case of $\zeta \neq 1$ due to the technical difficulties associated with the solution of the functional equation.

There are some other earlier works [4], [5], [6], [7] based on fluid flow models with regard to the issue, where they considered the correlation between inter-arrival and service times by assuming that the work in the queue is continuously removed. However, as illustrated in [1] fluid flow models do not account for the granularity of the arrival and service processes and hence result in inaccuracy in predicting the system performance.

Recently, Hwang and Sohraby [8] considered a correlated queue in an ATM-based MPLS system where the packet generation times are according to a geometric distribution, and analyzed the buffer size distribution of the system. They [9] also analyzed a correlated queue which is an extension of the queue considered in [1]. This paper is an extension of the correlated queue considered in [8], where we use the PH-type distribution for the packet generation time and a general distribution for the packet service time. So, from the analysis of our model we see that the effect of the packet generation time distribution as well as the effect of the correlated service and the inter-arrival times on system performance, which is a contribution of this paper.

The rest of the paper is organized as follows. In section II we model our system mathematically and analyze it to see the impact of the correlation between the packet service times and the packet inter-arrival times on system performance, e.g., the unfinished work. Some numerical examples are provided. We

give conclusions in section III.

II. SYSTEM MODELLING AND ANALYSIS

For two communication nodes, say node 1 and node 2, communicating through a finite capacity link, we assume that node 1 generates packets and sends them to node 2 through the link. We consider a queue in node 2 which accommodates packets from node 1 and has the same service capacity as the link between node 1 and node 2 (i.e., node 1 and node 2 have the same service capacity). To model the queue mathematically, we assume the following: Time axis is divided into slots of fixed size Δt. The packet arrival process is according to an alternating renewal process having two types of periods, say, packet generation times and packet service times. As mentioned before, packet generation times capture the times needed for generating new packets at node 1 and packet service times capture the packet transmission times at node 1 through the link. So, a packet inter-arrival time at our system consists of a packet generation time and a packet service time and our system has a packet arrival at the end of each packet service time.

Packet service times are assumed to be according to a general distribution with Probability Generating Function (PGF) $P(z)$ as follows:

$$P(z) = \sum_{i=1}^{M} a_i z^i,$$

where a_i denotes the probability that a packet service time is of length i (slots), $1 \le i \le M < \infty$. Here, M is a fixed number. Even though we assume a discrete probability distribution for the packet service times, a discrete probability distribution can be used as an approximation of the continuous packet service times by taking Δt as small as possible. We further assume that the packet service times are independent and identically distributed (i.i.d.).

Packet generation times are assumed to be according to a discrete time Phase type distribution (PH type distribution) [10]. A discrete time PH type distribution is defined by considering an $m + 1$ state Markov chain with transition probability matrix of the form

$$\begin{vmatrix} \mathbf{T} & \mathbf{T}^0 \\ \mathbf{0} & 1 \end{vmatrix}, \tag{1}$$

where $\mathbf{T}$ is a substochastic matrix, such that $\mathbf{I} - \mathbf{T}$ is nonsingular, and $\mathbf{T}^0$ is a column vector satisfying $\mathbf{T}e + \mathbf{T}^0 = e$. Here, e is a column vector all of whose elements are equal to 1.

The state space of the Markov chain is given by $\{1, \cdots, m, m + 1\}$ and the first m states are called transient states and the last state $m + 1$ is called an absorbing state. The initial probability vector is $(\boldsymbol{\alpha}, \alpha_{m+1})$ where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_m)$. Then, the discrete time PH type distribution is defined by the distribution of time until absorption in the finite Markov chain defined in (1). For analysis, we assume that $\alpha_{m+1} = 0$ and $\mathbf{T} + \mathbf{T}^0 \boldsymbol{\alpha}$ is an irreducible matrix for convenience. We use $J(t)$ to denote the state of the finite Markov chain defined in (1) at time t. In addition, T_{ij} and T_i^0 denote the (i, j)-th element and the i-th element of the matrix $\mathbf{T}$ and the column vector $\mathbf{T}^0$, respectively. It is worthwhile to mention that any general distribution can be approximated by a PH type distribution arbitrarily closely. For more details, refer to [10].

To model our system mathematically, we further assume the following:

i) state transitions of the Markov chain in (1) occur at slot boundaries.

ii) Whenever the Markov chain in (1) visits the absorbing state $m + 1$ at time, say, t_1, it makes a transition to state $i \in \{1, \cdots, m\}$ immediately with probability α_i at time t_1. That is, we have $J(t_1) = m + 1$ and $J(t_1^+) = i$. Moreover, the current packet generation time ends at time t_1 and the service time of the generated packet begins at time t_1^+. In this case, we say that a packet service time with state i begins at time t_1^+ for convenience.

iii) When a packet service time begins, the Markov chain in (1) stops its evolution until the packet service time ends. That is, there is no state change of the Markov chain in (1) during packet service time periods.

iv) When a packet service time with state i ends at time, say t_2, the Markov chain in (1) evolves again, so that it takes either a new transient state j (i.e., $J(t_2) = j$) with probability $T_{ij}, 1 \le j \le m$ or the absorbing state $m + 1$ immediately (i.e., $J(t_2) = m+1$) with probability T_i^0. If the new state is one of the transient states $\{1, \cdots, m\}$, then a new packet generation time (of length greater than 0) begins at time t_2. If the new state is the absorbing state $m + 1$, then we assume that we have a packet generation time of length 0. So, as described in ii) it makes a transition to a new state $j \in \{1, \cdots, m\}$ immediately (i.e., $J(t_2^+) = j$) with probability α_j at t_2, and a new packet service time begins at t_2^+.

Now we are ready to analyze our system to get the unfinished work of the system in steady state. Here, the unfinished work at time t is defined by the total amount of time needed for the system to service all the packets in the system at time t. In order to analyze our system, we first construct an embedded Markov chain by taking the ends of the slots in packet generation times and the ends of packet arrival slots as our embedded points. Here, the packet arrival slot denotes the last slot of the packet service time. Refer to Fig. 1. We define the following two random variables:

$X_n :=$ the unfinished work of the system at the n-th embedded point

$A_n :=$ the service time of the packet arriving just before the n-th embedded point.

We first obtain the evolution equation for the random variable X_n. If there is no packet arrival just before the n-th embedded point, i.e., the n-th embedded point is the end of a slot in a packet generation time, we have $A_n = 0$. If there is a packet arrival just before the n-th embedded point, i.e., the n-th embedded point is the end of a packet arrival slot, the (conditional) PGF of A_n is given by $P(z)$. Note that, since we assume that the service capacities of node 1 and node 2 are equal, the service time A_n of a packet at node 2 also has $P(z)$ as its PGF. Accordingly, it is easy to show that $\{X_n\}$ has the following evolution equations:

$$X_{n+1} = \begin{cases} A_{n+1}, & \text{if } X_n < A_{n+1} \text{ and } A_{n+1} > 0, \\ X_n, & \text{if } X_n \ge A_{n+1} \text{ and } A_{n+1} > 0, \\ (X_n - 1)^+, & \text{if } A_n = 0. \end{cases} \tag{3}$$

Next, for analysis we introduce a discrete time embedded Markov chain $\{J_n\}$ with state space $\{1, \cdots, m\}$, defined by $J_n \stackrel{\triangle}{=} J(s_n^+)$ where s_n is the n-th embedded point. Then, J_n
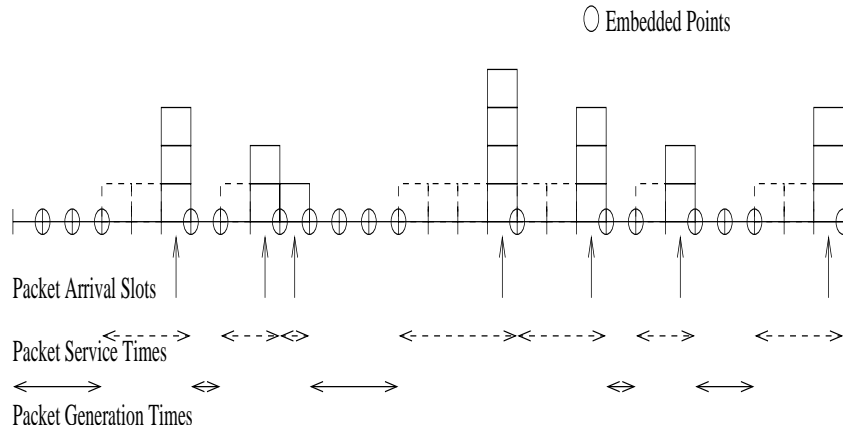
Fig. 1 Packet arrival process

$$
\begin{pmatrix}
\mathbf{T} & \mathbf{T}^0\boldsymbol{\alpha}a_1 & \mathbf{T}^0\boldsymbol{\alpha}a_2 & \mathbf{T}^0\boldsymbol{\alpha}a_3 & \cdots & \mathbf{T}^0\boldsymbol{\alpha}a_{M-1} & \mathbf{T}^0\boldsymbol{\alpha}a_M \\
\mathbf{T} & \mathbf{T}^0\boldsymbol{\alpha}a_1 & \mathbf{T}^0\boldsymbol{\alpha}a_2 & \mathbf{T}^0\boldsymbol{\alpha}a_3 & \cdots & \mathbf{T}^0\boldsymbol{\alpha}a_{M-1} & \mathbf{T}^0\boldsymbol{\alpha}a_M \\
0 & \mathbf{T} & \mathbf{T}^0\boldsymbol{\alpha}\sum_{n=1}^{2}a_n & \mathbf{T}^0\boldsymbol{\alpha}a_3 & \cdots & \mathbf{T}^0\boldsymbol{\alpha}a_{M-1} & \mathbf{T}^0\boldsymbol{\alpha}a_M \\
0 & 0 & \mathbf{T} & \mathbf{T}^0\boldsymbol{\alpha}\sum_{n=1}^{3}a_n & \cdots & \mathbf{T}^0\boldsymbol{\alpha}a_{M-1} & \mathbf{T}^0\boldsymbol{\alpha}a_M \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & \mathbf{T}^0\boldsymbol{\alpha}\sum_{n=1}^{M-1}a_n & \mathbf{T}^0\boldsymbol{\alpha}a_M \\
0 & 0 & 0 & 0 & \cdots & \mathbf{T} & \mathbf{T}^0\boldsymbol{\alpha}
\end{pmatrix}. \tag{2}
$$

denotes the state of the Markov chain in (1) just after the n-th embedded point s_n. Note that the transition probability matrix of $\{J_n\}$ is given by the irreducible matrix $\mathbf{T} + \mathbf{T}^0\boldsymbol{\alpha}$. Then, it is easy to show that a sequence $\{(X_n, J_n)\}$ is a Markov chain with state space $[0, M] \times [1, m]$ and it has the transition probability matrix given in (2).

Since $\{(X_n, J_n)\}$ is irreducible with a finite state space, the steady state probability vector for the Markov chain $\{(X_n, J_n)\}$ exists [11]. Introduce vectors $\boldsymbol{\pi}_k, 0 \le k \le M$, defined by

$$
\begin{aligned}
\boldsymbol{\pi}_k &= (\pi_{k1}, \cdots, \pi_{km}), \\
\pi_{kl} &= \lim_{n\to\infty} P\{X_n = k, J_n = l\}, \ 1 \le l \le m.
\end{aligned}
$$

From (2) we have the following balance equations:

$$
\begin{aligned}
\boldsymbol{\pi}_0 &= \boldsymbol{\pi}_0\mathbf{T} + \boldsymbol{\pi}_1\mathbf{T} \tag{4} \\
\boldsymbol{\pi}_n &= \boldsymbol{\pi}_0\mathbf{T}^0\boldsymbol{\alpha}a_n + \boldsymbol{\pi}_1\mathbf{T}^0\boldsymbol{\alpha}a_n + \cdots + \boldsymbol{\pi}_{n-1}\mathbf{T}^0\boldsymbol{\alpha}a_n \\
&\quad + \boldsymbol{\pi}_n\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n}a_i + \boldsymbol{\pi}_{n+1}\mathbf{T}, 1 \le n \le M-1, \tag{5} \\
\boldsymbol{\pi}_M &= \boldsymbol{\pi}_0\mathbf{T}^0\boldsymbol{\alpha}a_M + \boldsymbol{\pi}_1\mathbf{T}^0\boldsymbol{\alpha}a_M + \cdots \\
&\quad + \boldsymbol{\pi}_{M-1}\mathbf{T}^0\boldsymbol{\alpha}a_M + \pi_M\mathbf{T}^0\boldsymbol{\alpha}.
\end{aligned}
$$

For analysis, we introduce $\boldsymbol{q}_k, 0 \le k \le M$, defined by

$$
\boldsymbol{q}_k = \sum_{n=0}^{k} \boldsymbol{\pi}_n.
$$

Then, from (4) and (5) we obtain, for $1 \le k \le M-1$

$$
\begin{aligned}
\boldsymbol{q}_k &= \boldsymbol{\pi}_0\mathbf{T} + \boldsymbol{\pi}_1\mathbf{T} + \sum_{n=1}^{k}\sum_{j=0}^{n-1}\boldsymbol{\pi}_j\mathbf{T}^0\boldsymbol{\alpha}a_n \\
&\quad + \sum_{n=1}^{k}\boldsymbol{\pi}_n\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n}a_i + \sum_{n=1}^{k}\boldsymbol{\pi}_{n+1}\mathbf{T} \\
&= \sum_{j=0}^{k-1}\sum_{n=j+1}^{k}\boldsymbol{\pi}_j\mathbf{T}^0\boldsymbol{\alpha}a_n + \sum_{n=1}^{k}\boldsymbol{\pi}_n\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n}a_i \\
&\quad + \sum_{n=0}^{k+1}\boldsymbol{\pi}_n\mathbf{T} \\
&= \sum_{n=0}^{k}\boldsymbol{\pi}_n\left(\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k}a_i + \mathbf{T}\right) + \boldsymbol{\pi}_{k+1}\mathbf{T} \\
&= \boldsymbol{q}_k\left(\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k}a_i + \mathbf{T}\right) + \boldsymbol{\pi}_{k+1}\mathbf{T}. \tag{6}
\end{aligned}
$$

Without loss of generality we may assume $\sum_{i=1}^{k} a_i < 1$ for $1 \le k \le M-1$. Consequently, we have

$$
0 \le \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k}a_i + \mathbf{T} < \mathbf{T}^0\boldsymbol{\alpha} + \mathbf{T}.
$$

By the irreducibility of the matrix $\mathbf{T} + \mathbf{T}^0\boldsymbol{\alpha}$, the spectral radius of $\mathbf{T} + \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k}a_i$ is less than 1 (see Chapter II, Corollary 2.2 in [12]), which guarantees the invertibility of the matrix $\mathbf{I} -$

$\mathbf{T} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k} a_i$. Therefore, from (6) we obtain

$$\boldsymbol{q}_k = \boldsymbol{\pi}_{k+1}\mathbf{T}\left[\mathbf{I} - \mathbf{T} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k} a_i\right]^{-1}, 1 \leq k \leq M-1. \quad (7)$$

Observe that $\boldsymbol{q}_{M-1} + \boldsymbol{\pi}_M = \boldsymbol{z}$ where the row vector $\boldsymbol{z}$ satisfies

$$\boldsymbol{z}(\mathbf{T} + \mathbf{T}^0\boldsymbol{\alpha}) = \boldsymbol{z}, \boldsymbol{z}\boldsymbol{e} = 1. \quad (8)$$

Then, we have

$$\begin{aligned}
\boldsymbol{z} &= \boldsymbol{q}_{M-1} + \boldsymbol{\pi}_M \\
&= \boldsymbol{\pi}_M\mathbf{T}\left[\mathbf{I} - \mathbf{T} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{M-1} a_i\right]^{-1} + \boldsymbol{\pi}_M \\
&= \boldsymbol{\pi}_M\left[\mathbf{I} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{M-1} a_i\right]\left[\mathbf{I} - \mathbf{T} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k} a_i\right]^{-1} \quad (9)
\end{aligned}$$

By a similar argument given above, we can also show the invertibility of the matrix $\mathbf{I} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{k} a_i$ for $1 \leq k \leq M-1$. Therefore, from (9) we obtain

$$\boldsymbol{\pi}_M = \boldsymbol{z}\left[\mathbf{I} - \mathbf{T} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{M_1} a_i\right]\left[\mathbf{I} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{M-1} a_i\right]^{-1}. \quad (10)$$

Since the probability vector $\boldsymbol{z}$ can be obtained from (8), we can also obtain $\boldsymbol{\pi}_M$ from the equation (10). In addition, From (5) and the following equation

$$\boldsymbol{q}_{M-1} = \boldsymbol{z} - \boldsymbol{\pi}_M, \quad (11)$$

we get, for $1 \leq n \leq M-1$,

$$\begin{aligned}
\boldsymbol{\pi}_n &= \sum_{k=0}^{n-1}\boldsymbol{\pi}_k\mathbf{T}^0\boldsymbol{\alpha}a_n + \boldsymbol{\pi}_n\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n} a_i + \boldsymbol{\pi}_{n+1}\mathbf{T} \\
&= \boldsymbol{q}_n\mathbf{T}^0\boldsymbol{\alpha}a_n + \boldsymbol{\pi}_n\mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n-1} a_i + \boldsymbol{\pi}_{n+1}\mathbf{T}.
\end{aligned}$$

Hence, we finally get

$$\begin{aligned}
\boldsymbol{\pi}_n &= \left[\boldsymbol{q}_n\mathbf{T}^0\boldsymbol{\alpha}a_n + \boldsymbol{\pi}_{n+1}\mathbf{T}\right] \\
&\quad \times \left[\mathbf{I} - \mathbf{T}^0\boldsymbol{\alpha}\sum_{i=1}^{n-1} a_i\right]^{-1}, \quad (12) \\
\boldsymbol{q}_{n-1} &= \boldsymbol{q}_n - \boldsymbol{\pi}_n. \quad (13)
\end{aligned}$$

Starting from $\boldsymbol{\pi}_M$ and $\boldsymbol{q}_{M-1}$ given in (10) and (11), respectively, we can compute $\boldsymbol{\pi}_n$ and $\boldsymbol{q}_{n-1}$ from (12) and (13) iteratively for $1 \leq n \leq M-1$.

The next step is to find the distribution of the unfinished work at a packet arrival slot in steady state. For doing this, we introduce a random variable X to denote the unfinished work in the system at an embedded point in steady state, i.e., $P\{X = n\} = \boldsymbol{\pi}_n\boldsymbol{e}$. Similarly, we introduce a random variable U to denote the unfinished work in the system at a packet arrival

slot. Then, it is easy to show that the probability that $U = n$ is given by

$$P\{U = n\} = \frac{\boldsymbol{\pi}_n\mathbf{T}^0}{\sum_{k=0}^{M}\boldsymbol{\pi}_k\mathbf{T}^0}.$$

Now we compute the distribution of V, the unfinished work in the system at a packet arrival slot, including the service time of the newly arriving packet. Since U is the unfinished work in the system seen by a newly arriving packet, the distribution of V can be computed from

$$\begin{aligned}
P\{V = k\} &= \sum_{n=0}^{k-1} P\{U = n\}a_k \\
&\quad + P\{U = k\}\sum_{i=1}^{k} a_i, k \geq 1. \quad (14)
\end{aligned}$$

Note that we can determine the buffer size of the system with certain level of packet loss probability based on the distribution of V.

We end this section with some numerical results based on our analysis. In the numerical study, we use actual measured data for the packet service time, denoted by S. The data for the packet size distribution was obtained from [13]. In [13] the packet size data is given in bytes, so we should convert the packet size in bytes into its corresponding service time. Here, we assume that the link capacity is 48×8 Mbps and $\Delta t = 10^{-6}$, and consequently, a segment of 48 bytes are serviced in each slot of length Δt. The resulting distribution of the service time is given in Fig. 2.

To see the impact of the correlated service and inter-arrival times on system performance, e.g., the unfinished work, we consider a geometric distribution for the packet generation times, denoted by G, as follows:

$$P\{G = k\} = (1-p)^k p, \quad k \geq 0.$$

In addition, for the comparison purpose we consider the corresponding GI/GI/1 queue having inter-arrival time $G + S$ and the packet service time S. We compute the unfinished works for our system and the corresponding GI/GI/1 queue in cases of $p = 0.1$ and $p = 0.3$, and plot the resulting complementary distributions $P\{U > n\}$ in Fig. 3 on a log scale. As seen in the figure, the unfinished work distributions with and without the correlations are far away from each other. Moreover, the means and variances are different for both systems: When $p = 0.3$ the mean and variance of our system are 18.112 and 93.332, respectively, while the mean and the variance of the corresponding GI/GI/1 queue are 37.050 and 2066.001, respectively. Therefore, from our observation we see that the correlation plays an important role in system performance.

To see the effect of the packet generation time distribution on the performance of our correlated system, we consider the following mixture of geometric distributions for the packet generation time:

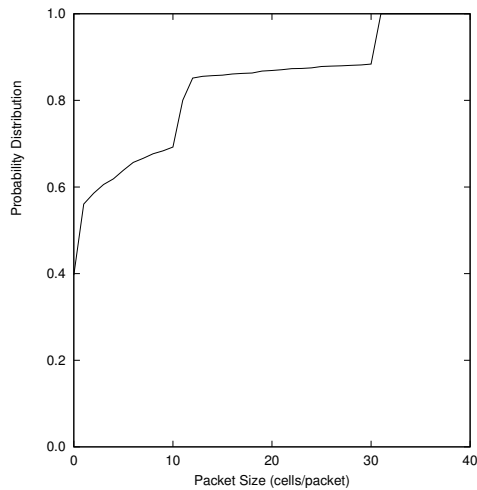$$P\{G = k\} = \sum_{i=1}^{m} c_i(1-q_i)^k q_i, k \geq 0,$$
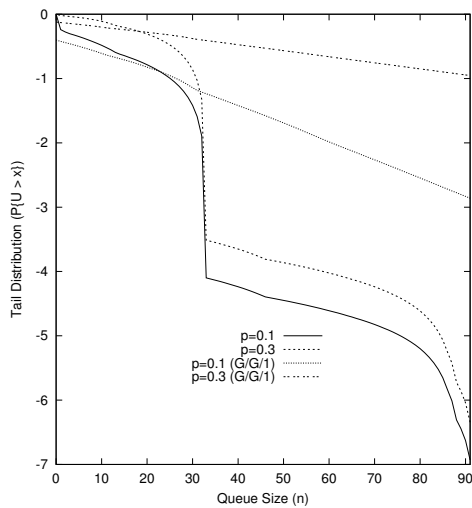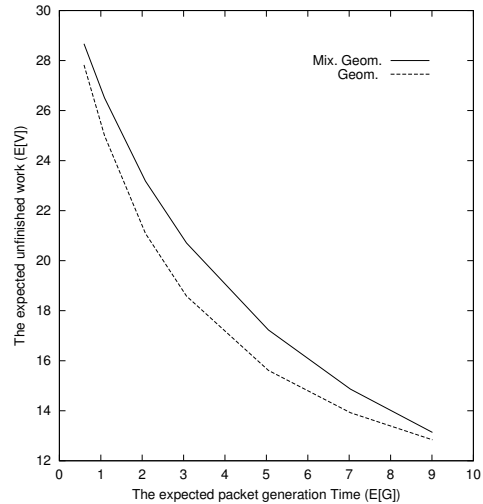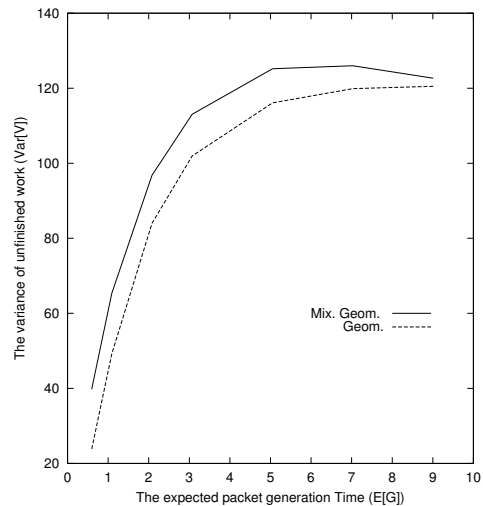
Fig. 2 Packet size distribution

Fig. 4 The effect of $E[G]$ on the expectation of the unfinished work

Fig. 3 The effect of correlations on performance

Fig. 5. The effect of $E[G]$ on the variance of the unfinished work

where $0 < q_i < 1, 0 \leq c_i \leq 1$ and $\sum_{i=1}^{m} c_i = 1$. Then PH type representation of the distribution is as follows:

$$\mathbf{T} = \begin{pmatrix} 1-q_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1-q_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-q_{m-1} & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1-q_m \end{pmatrix},$$

$$\mathbf{T}^0 = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_{m-1} \\ q_m \end{pmatrix}, \boldsymbol{\alpha} = (c_1, \cdots, c_m).$$

In the numerical computation, we change the mean $E[G]$ of the packet generation time distribution and examine the behaviors of the mean and the variance of the unfinished work at a

packet arrival slot. We use $m = 2, q_1 = 10/11, q_2 = 1/11$ and $\tilde{\boldsymbol{\alpha}} = (\beta, 1 - \beta)$. We change the value of β from 0.1 to 0.95, so that the value of $E[G]$ changes from 9.01 to 0.595. The results are given in Fig. 4 and Fig. 5. As seen in the figure, the mean of the unfinished work is decreasing as the expected packet generation time is increasing. On the other hand, the variance of the unfinished work is increasing as the expected packet generation time is increasing. Next, to see the effect of the packet generation time distribution on the unfinished work, we also plot the mean and the variance of the unfinished work when we have a geometric distribution having the same mean $E[G]$ for the packet generation time distribution. The results are also plotted in Fig. 4 and Fig. 5. As seen in the figures, the distribution of the packet generation time also affects system performance.

388

III. CONCLUSION

The main contribution of this paper is to propose a mathematical framework to study the impact of the correlation between the packet service times and the packet inter-arrival times on system performance. In order to attain the generality of our mathematical model, we use the PH-type distribution for the packet generation times and a general distribution for the packet service times.

With our framework we have analyzed the system performance, e.g., the unfinished work of the system and have shown that the correlation considered in this paper affects the system performance significantly.

REFERENCES

[1] I. Cidon, R. Guerin, A. Khamisy, and M. Sidi, "Analysis of a Correlated Queue in a Communication System", *IEEE Transactions on Information Theory*, Vol. 39, No. 2, pp. 456–465, 1993.

[2] I. Cidon, R. Guerin, A. Khamisy, and M. Sidi, "On queues with interarrival times proportional to service times", *Proceeding of INFOCOM '93*.

[3] M. Kuczma, B. Choczewski and R. Ger., *Iterative Functional Equations*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1990.

[4] D. Anick, D. Mitra, and M.M. Sondhi, "Stochastic theory of a data-handling system with multiple sources," *Bell System Tech. J.*, Vol. 61, pp. 1871-1894, 1982.

[5] A.I. Elwalid, D. Mitra, and T.E. Stern, "Statistical multiplexing of Markov modulated sources: Theory and computational algorithms," *Proc. 13th Int. Teletraffic Cong.*, Copenhagen, pp. 495–500, 1991.

[6] D. Mitra, "Stochastic theory of a fluid model of producers and consumers coupled by a buffer," *Advances in Applied Probability*, Vol. 20, pp. 646–676, 1988.

[7] T.E. Stern and A.I. Elwalid, "Analysis of separable Markov-modulated models for information-handling systems," *Advances in Applied Probability*, Vol. 23, pp. 105–139, 1991.

[8] G.U. Hwang and K. Sohraby, "Modelling and Analysis of a Buffer in an ATM-based MPLS LER System," *Proceeding of Globecom 2002*, CQRS-06-3, Taipei, Taiwan, 2002.

[9] G.U. Hwang and K. Sohraby, "Performance of Correlated Queues: The Impact of Correlated Service and Inter-arrival Times," *Performance Evaluation*, Vol. 55, No. 1-2, 129-145, 2004.

[10] M.F. Neuts, *Matrix Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, 1981.

[11] S. Asmussen, *Applied Probability and Queues*, John Wiley and Sons, 1987.

[12] H. Minc, *Nonnegative Matrices*, John Wiley & Sons, Inc., 1988.

[13] WAN packet Size Distribution, [Online], Available WWW: http://www.nlanr.net/NA/Learn/packetsizes.html.

Gang U. Hwang received the B.S., M.S., and Ph. D. degrees in Mathematics (Applied Probability) from KAIST, Taejeon, Republic of Korea, in 1991, 1993 and 1997, respectively.

From February 1997 to March 2000, he joined Electronics and Telecommunications Research Institute (ETRI), Taejeon, Republic of Korea, where he was a senior member of research staff. From March 2000 to February 2002, he was at the School of Interdisciplinary Computing and Engineering in University of Missouri - Kansas City as a visiting scholar. Since March 2002, he has been with Division of Applied Mathematics at KAIST, where he is an Associate Professor. His research interests include teletraffic theory and traffic engineering issues for next generation communication networks.