Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2008, Article ID 814740, 14 pages
doi:10.1155/2008/814740

Research Article

An Erlang Loss Queue with Time-Phased Batch Arrivals as a Model for Traffic Control in Communication Networks

Moon Ho Lee[1] and Sergey A. Dudin[2]

[1] *Institute of Information and Communication, Chonbuk National University, Chonju 561-765, South Korea*
[2] *Department of Applied Mathematics and Computer Science, Belarusian State University, Minsk 220030, Belarus*

Correspondence should be addressed to Sergey A. Dudin, dudin@madrid.com

Received 16 June 2008; Revised 16 October 2008; Accepted 9 December 2008

Recommended by Oded Gottlieb

A multiserver queueing model that does not have a buffer but has batch arrival of customers is considered. In contrast to the standard batch arrival, in which the entire batch arrives at the system during a single epoch, we assume that the customers of a batch (flow) arrive individually in exponentially distributed times. The service time is exponentially distributed. Flows arrive according to a stationary Poisson arrival process. The flow size distribution is geometric. The number of flows that can be simultaneously admitted to the system is under control. The loss of any customer from an admitted flow, with a fixed probability, implies termination of the flow arrival. Analysis of the sojourn time and loss probability of an arbitrary flow is performed.

1. Introduction

Queueing systems effectively describe the operations of channels and servers of many communication networks, and they have received a great deal of attention in the probabilistic literature, beginning with the pioneering works by the Danish mathematician and engineer A. K. Erlang. A celebration of the 100 year jubilee of the origin of queueing theory will be held in 2009. Erlang's famous loss formula for the customer loss probability for the system $M/M/N/0$ served as the basis for capacity planning and performance evaluation of telecommunication networks for an entire century. Sevastjanov proved in [1] that this loss formula is valid for the more general queueing system $M/G/N/0$ as well. In [2–4], analysis of Erlang's model was extended to the system with (batch Markovian arrival process) *BMAP*.

It was assumed in [2–4] that customers can arrive in batches of random size, with the standard assumption that, during a batch arrival epoch, all batch customers arrive at the

system simultaneously. However, nowadays, it is a typical feature of many communication networks, *IP* networks in particular, that customers arrive in batches, but the arrival of customers of a batch is not instantaneous. To distinguish the standard batches from the batches considered in this paper the latter ones are named flows. The first customer of a flow arrives during the flow arrival epoch, while the rest of the customers arrive individually during random intervals. The flow size is random, and it may not be known a priori during the flow arrival epoch. Such a situation is typical, for example, in modeling transmission of video and multimedia information. This situation is also discussed in [5] with respect to the modeling scheme of alternative packet overflow routing in IP networks. In this scheme, the flow represents a set of packets that must be sequentially routed in the same channel. When a packet arrives, it is determined (e.g., by means of *IP* address) if the packet is a part of a flow previously tracked. If the packet belongs to an existing flow, the packet is marked for transmission. If the flow has not yet been tracked and the buffer and channel capacity is still available, the packet is admitted to the system and the flow count is increased. Otherwise, the flow is routed on an overflow link (or is dropped) and the packet is rejected in the considered channel. Tracked flows are cleared after they are finished. Clearing of an inactive flow is performed if no more packets belonging to this flow are received within a certain time interval. Tracking and clearing of flows is performed by a token mechanism. Physically, the token can be interpreted as a particular timer that is activated during a flow admission epoch and is restarted during epochs of other customers from this flow arrival, and is terminated if a particular fixed timeout expires and a new customer from this flow does not arrive. The number of tokens (timers), which defines the maximum number of flows that can be admitted to the system simultaneously, is a very important control parameter. If this number is too small, the channel may be underutilized. If this number is too large, the channel may become congested. Many packets from admitted flows may be lost, and grade of service is poor. In addition, the delay and jitter of flows may inherently increase. So, the problem of defining the optimal number of tokens is practically important and nontrivial. In [5], performance measures of the *SAPOR* scheme of routing in *IP* networks are evaluated by means of computer simulation. In [6], a description of such a scheme is given in terms of the queueing model.

Also, there is an analogous model for information retrieval in relational databases, where, besides the CPU and disc memory, some additional threads or connections must be provided, to start the user's application processing. In this interpretation, a flow means an application, while customers are queries that will be processed within this application.

In [6], a multiserver Markovian queueing model with a finite buffer that suits the performance evaluation of the scheme of alternative packet overflow routing in *IP* networks, as well as that of other real-world systems with a time distributed arrival of customers in a flow, is constructed and investigated analytically. In particular, the joint distribution of the number of customers and flows in the system is computed. A very important performance measure of this queueing model is the flow sojourn time computed since the instant of a flow arrival until the instant when all customers from this flow finish their arrival and processing in the system.

In [6], the Laplace-Stieltjes transform for a flow sojourn time distribution is calculated only for the case of a single server queue. In this paper, we calculate this distribution for a multiserver queue with losses (without a buffer). The considered model is more general than the model considered in [6], and in the following respect: it was assumed in [6] that the loss of any customer from the admitted flow does not affect the further behavior of the system.

In this paper, we assume that, with fixed probability, the entire flow terminates its arrival after the loss of a customer from this flow. This assumption is realistic in various situations, for example, termination of a connection can occur ahead of schedule if the percentage of lost voice or video packets (and quality of speech or movie) becomes unacceptable for the user.

The rest of the paper is organized as follows. In Section 2, the model is described. The steady-state joint distribution of the number of flows and customers in the system is given in Section 3. Sections 4 and 5 contain the main contributions of this paper: an analysis of a flow sojourn time distribution, and an expression for the loss probability of an accepted flow, due to the loss of a particular customer from this flow. Numerical illustrations are presented in Section 6.

2. Mathematical model

We consider a queueing system with N identical servers, where the service time in the server has an exponential distribution with the parameter μ. The system does not have a buffer. Customers arrive at the system in flows. Flows arrive at the system according to a stationary Poisson arrival process with the intensity λ. In accordance with [5], we assume that admission of flows is restricted by means of *tokens*. The total number of available tokens is assumed to be K, $K \geq 1$. The number K can be considered as a control parameter, and various optimization problems can be solved.

If there is no token available at a flow arrival epoch, or all the servers are busy, the flow is rejected. It leaves the system permanently. If the number of available tokens is positive and there is at least one free server, the flow is admitted to the system and the number of available tokens decreases by one. After admission of the flow, the next customer of this flow can arrive at the system in an exponentially distributed time, with the parameter γ.

If there is a free server at the instant of arrival of a customer from the admitted flow, the customer is admitted to the system. Otherwise, the customer is rejected. With probability $1 - q$, $0 \leq q \leq 1$, this rejection does not affect the future behavior of the system. But, with the supplementary probability q, this rejection causes the termination of the flow arrival. However, the customers from this flow, which were previously admitted to the system, can only leave the system after being processed by the server.

The number of customers in a flow has a geometrical distribution, with the parameter θ, $0 < \theta < 1$, that is, the probability that the flow consists of k customers is equal to $\theta^{k-1}(1 - \theta)$, $k \geq 1$, and the mean number of the customers in a flow is equal to $(1 - \theta)^{-1}$.

If an exponentially distributed time (parameter γ) since the arrival of the previous customer of a flow expires, and a new customer does not arrive, this means that the arrival of the flow is finished. The token, which was obtained by this flow upon arrival, is returned to the pool of available tokens. The customers of this flow, which stay in the system during the epoch when the token is returned, must be processed by the system. When the last customer is served, the sojourn time of the flow in the system is considered finished.

It is intuitively obvious that this mechanism of arrivals restriction by means of tokens is reasonable. At the expense of rejecting some flows, it enables a decrease of the probability of the loss of a customer from the admitted flows, flows sojourn time, and jitter. It is important in modeling real-world systems, because the quality of transmission of accepted information units must satisfy the imposed requirements of quality of service (QoS).

3. Stationary distribution of the number of flows and customers in the system

Let i_t, $i_t = \overline{0,N}$, denote the total number of customers in the system during epoch t, $t \geq 0$, and let k_t, $k_t = \overline{0,K}$, denote the number of flows with a token for admission to the system during epoch t, $t \geq 0$. Here, a notation such as $i_t = \overline{0,N}$ means that i_t assumes values from the set $\{0,1,\ldots,N\}$. It is obvious that the two-dimensional process $\zeta_t = \{i_t, k_t\}$, $t \geq 0$, is a finite irreducible regular continuous-time Markov chain.

Let Q be the generator of the Markov chain ζ_t, $t \geq 0$, with blocks $Q_{i,j}$ of dimension $(K+1) \times (K+1)$ consisting of intensities $(Q_{i,j})_{k,k'}$ of the Markov chain ζ_t, $t \geq 0$, transitions from the state (i,k) to the state (j,k'), $k,k' = \overline{0,K}$, $i = \overline{0,N}$, $j = \overline{0,N}$, $i \neq j$ if $k = k'$. The diagonal entries of the matrix $Q_{i,i}$ are negative, and the modulus of the entry $(Q_{i,i})_{k,k}$ defines the total intensity of leaving the state (i,k) of the Markov chain.

Here, we introduce the following notation:

(i) $\gamma^- = \gamma(1-\theta)$, $\gamma^+ = \gamma\theta$;

(ii) $C_K = \operatorname{diag}\{0,1,\ldots,K\}$, that is, the diagonal matrix with the diagonal entries $\{0,1,\ldots,K\}$;

(iii) I is an identity matrix, O is a zero matrix, $\widehat{I} = \operatorname{diag}\{1,\ldots,1,0\}$, $\mathbf{e}$ is a column vector of dimension $K+1$ consisting of 1 s, $\mathbf{0}$ is a row vector of dimension $K+1$ consisting of 0 s;

(iv)

$$A = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \gamma^- & -\gamma & 0 & \cdots & 0 & 0 \\ 0 & 2\gamma^- & -2\gamma & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & K\gamma^- & -K\gamma \end{pmatrix}; \tag{3.1}$$

(v)

$$A_1 = \begin{pmatrix} -\gamma & O & \cdots & O & O \\ \gamma^- & -2\gamma & \cdots & O & O \\ O & 2\gamma^- & \cdots & O & O \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O\cdots & (K-1)\gamma^- & -K\gamma \end{pmatrix}; \tag{3.2}$$

(vi) square matrices E^+, E^-, of the corresponding size have all zero entries, except the entries $(E^+)_{k,k+1}$, $(E^-)_{k,k-1}$, which are equal to 1. If the size of a matrix is not indicated as a suffix explicitly, it is assumed to be equal to $K+1$;

(vii) $\otimes$ is the sign of the Kronecker product of matrices, $\delta_{i,j}$ is the Kronecker delta.

Lemma 3.1. *The infinitesimal generator Q of the Markov chain $\zeta_t,\ t \geq 0$, has the following three-block-diagonal structure:*

$$Q = \begin{pmatrix} Q_{0,0} & Q_{0,1} & O & \cdots & O & O \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots & O & O \\ O & Q_{2,1} & Q_{2,2} & \cdots & O & O \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ O & O & O & \cdots & Q_{N-1,N-1} & Q_{N-1,N} \\ O & O & O & \cdots & Q_{N,N-1} & Q_{N,N} \end{pmatrix}, \tag{3.3}$$

where the nonzero blocks $Q_{i,j}$ are computed by

$$\begin{aligned} Q_{i,i} &= A - \lambda\widehat{I} - i\mu I, \quad Q_{i+1,i} = (i+1)\mu I, \quad i = \overline{0, N-1}, \\ Q_{N,N} &= A + (1-q)\gamma^{+}C_K + q\gamma^{+}\left(C_K E^{-}\right) - N\mu I, \\ Q_{i-1,i} &= \gamma^{+}C_K + \lambda E^{+}, \quad i = \overline{1, N}. \end{aligned} \tag{3.4}$$

The proof of the lemma consists of analysis of the Markov chain $\zeta_t,\ t \geq 0$, transitions during the infinitesimal interval of time, and combining the corresponding transition intensities with the matrix blocks.

Because the Markov chain $\zeta_t = \{i_t, k_t\},\ t \geq 0$, is irreducible and regular, and has a finite state space, the stationary probabilities exist:

$$\pi(i,k) = \lim_{t\to\infty} P\{i_t = i,\ k_t = k\}, \quad i = \overline{0, N},\ k = \overline{0, K}. \tag{3.5}$$

We can apply an effective and numerically stable algorithm for calculation of such probabilities for Markov chains with a generator of form (3.3), (elaborated in [4]).

Once the stationary probabilities $\pi(i,k)$ have been computed, we can calculate various performance measures of the system. The most essential measures are as follows.

Corollary 3.2. *The mean number T of customers processed by the system during unit time (throughput) is computed by*

$$T = \mu \sum_{i=1}^{N} \sum_{k=0}^{K} i\pi(i,k). \tag{3.6}$$

The probability $P_b^{(loss)}$ of an arbitrary flow rejection upon arrival is computed by

$$P_b^{(loss)} = \sum_{i=0}^{N-1} \pi(i,K) + \sum_{k=0}^{K} \pi(N,k). \tag{3.7}$$

4. Distribution of the sojourn time

Let $V_b(x)$ denote the distribution function of the sojourn time of an arbitrary flow in the system under study, and let $v_b(s)$ denote its Laplace-Stieltjes transform (LST):

$$v_b(s) = \int_0^\infty e^{-sx}\, dV_b(x), \quad \operatorname{Re} s > 0. \tag{4.1}$$

We derive an expression for the LST $v_b(s)$ by means of the very powerful method of collective marks (method of additional event, method of catastrophes); see [7, 8]. To this end, we interpret the variable s as the intensity of a particular virtual stationary Poisson process of catastrophes. So, the function $v_b(s)$ denotes the probability that no catastrophe arrives during the sojourn time of an arbitrary (tagged) flow.

Let $v(s,i,n,k)$ denote the probability that a catastrophe from the stationary Poisson process of catastrophes with intensity s will not arrive during the rest of the tagged flow sojourn time in the system conditional that, at the given instant, the number of flows processed in the system is equal to k, $k = \overline{1,K}$, the number of customers is equal to i, $i = \overline{0,N}$, and the number of customers from the tagged flow equals n, $n = \overline{0,i}$. The system of linear algebraic equations for functions $v(s,i,n,k)$ is derived by means of the law of the total probability, expressed as follows:

$$\begin{aligned}
&v(s,i,n,k)\big(s + \lambda\big(1-\delta_{k,K}\big)\big(1-\delta_{i,N}\big) + i\mu + k\gamma\big) \\
&\quad = \lambda\big(1-\delta_{k,K}\big)\big(1-\delta_{i,N}\big)v(s,i+1,n,k+1) \\
&\qquad + (i-n)\mu v(s,i-1,n,k) + n\mu v(s,i-1,n-1,k) \\
&\qquad + \gamma^+\big(1-\delta_{i,N}\big)\big(v(s,i+1,n+1,k) + (k-1)v(s,i+1,n,k)\big) \\
&\qquad + \gamma^+\delta_{i,N}\big((1-q)kv(s,i,n,k) + q + q(k-1)v(s,i,n,k-1)\big) \\
&\qquad + \gamma^-\Gamma_n(s) + \gamma^-(k-1)v(s,i,n,k-1), \\
&\qquad\quad n = \overline{0,i}, \quad i = \overline{0,N}, \quad k = \overline{1,K},
\end{aligned} \tag{4.2}$$

where

$$\begin{aligned}
\Gamma_n(s) &= \sum_{l=1}^{n} \binom{n}{l} (-1)^{l+1} \frac{\mu l}{\mu l + s} \\
&= \prod_{l=1}^{n} \frac{\mu l}{\mu l + s}, \quad n \geq 1,\ \Gamma_0(s) = 1.
\end{aligned} \tag{4.3}$$

Proof of (4.2) can be clarified by considering that

(i) $\lambda(1-\delta_{k,K})(1-\delta_{i,N})$ is the intensity of new flows gaining admission to the system;

(ii) $n\mu$ is the intensity of service completion of customers from the tagged flow;

(iii) $(i-n)\mu$ is the intensity of service completion of customers that are not from the tagged flow;

(iv) $\gamma^{+}(1-\delta_{i,N})$ is the intensity of arrival of new customers from a tagged flow, when the system is not full;

(v) $\gamma^{+}(1-\delta_{i,N})(k-1)$ is the intensity of arrival of new customers from other flows, when the system is not full;

(vi) $\gamma^{+}\delta_{i,N}(1-q)k$ is the intensity of arrival of new customers from all admitted flows to the full system, which will not cause the termination of the corresponding flow;

(vii) $\gamma^{+}\delta_{i,N}q(k-1)$ is the intensity of the termination (due to an arriving customer loss) of nontagged flows;

(viii) $\gamma^{+}\delta_{i,N}q$ is the intensity of the termination (due to an arriving customer loss) of the tagged flow;

(ix) γ^{-} is the intensity of finishing the arrival of customers from the tagged flow;

(x) $\gamma^{-}(k-1)$ is the intensity of finishing the arrival of customers from the nontagged flows;

(xi) $\Gamma_n(s)$ is probability of no catastrophe arrival after finishing the arrival of customers from the tagged flow, when n customers from this flow are served in the system at the arrival finishing epoch.

To solve system (4.2), the following vector notation is useful:

$$\begin{aligned}
\mathbf{v}(s,i,n) &= \big(v(s,i,n,1),\dots,v(s,i,n,K)\big)^{T}, \quad n=\overline{0,i},\\
\mathbf{v}(s,i) &= \big(\mathbf{v}(s,i,0),\dots,\mathbf{v}(s,i,i)\big)^{T}, \quad i=\overline{0,N},\\
\mathbf{v}(s) &= \big(\mathbf{v}(s,0),\dots,\mathbf{v}(s,N)\big)^{T},\\
\mathcal{B}_i(s) &= \gamma^{-}\big(\mathbf{e}_K,\Gamma_1(s)\mathbf{e}_K,\dots,\Gamma_i(s)\mathbf{e}_K\big)^{T}, \quad i=\overline{0,N},\\
\mathcal{B}(s) &= \big(\mathcal{B}_0(s),\dots,\mathcal{B}_N(s)\big)^{T}+q\gamma^{+}\big(0,\dots,0,\mathbf{e}_{K(N+1)}\big)^{T}.
\end{aligned} \tag{4.4}$$

Let the matrix function $\boldsymbol{\Omega}(s)$ be defined by its entries:

$$\begin{aligned}
\Omega_{i,i}(s) &= -I_{i+1}\otimes\big(sI-\widehat{Q}_{i,i}\big),\\
\Omega_{i,i+1} &= D_1^{(i)}\otimes\widehat{Q}_{i,i+1}+D_2^{(i)}\otimes\gamma^{+}I_K,\\
\Omega_{i,i-1} &= \begin{pmatrix}
i\mu & 0 & 0 & \cdots & 0\\
\mu & (i-1)\mu & 0 & \cdots & 0\\
\dots & \ddots & \ddots & \dots & \dots\\
0 & 0 & \cdots & (i-1)\mu & \mu\\
0 & 0 & \cdots & 0 & i\mu
\end{pmatrix}\otimes I_K,
\end{aligned}$$

$$\begin{aligned}
\widehat{Q}_{i,i} &= A_1 - \lambda \widehat{I}_K - i\mu I_K, \quad i = \overline{0, N-1}, \\
\widehat{Q}_{N,N} &= A_1 + (1-q)\gamma^+ \left(C_{K-1} + I\right) + q\gamma^+ C_{K-1} E^- - N\mu I_K, \\
\widehat{Q}_{i,i+1} &= \gamma^+ C_{K-1} + \lambda E_K^+, \quad i = \overline{0, N-1},
\end{aligned} \tag{4.5}$$

where the matrix $D_1^{(i)}$ ($D_2^{(i)}$) is obtained from the identity matrix I_{i+1} by means of supplementing it from the right (left) side with the column $\mathbf{0}_{i+1}^T$. The matrix $D_3^{(i)}$ is obtained from the identity matrix I_i by means of supplementing it from the top with the row $\{1, 0, \ldots, 0\}$.

Using this notation, system (6.1) can be easily expressed as

$$\mathbf{\Omega}(s)\mathbf{v}(s) = -\mathcal{B}(s). \tag{4.6}$$

Considering this relation and using the law of the total probability, we can easily prove the following statement.

Theorem 4.1. *The LST $v_b(s)$ is calculated by*

$$v_b(s) = P_b^{(loss)} + \sum_{i=0}^{N-1} \sum_{k=0}^{K-1} \pi(i,k) v(s, i+1, 1, k+1), \tag{4.7}$$

where the vector $\mathbf{v}(s)$ consisting of the conditional Laplace-Stieltjes transforms $v(s,i,l,k)$, $l = \overline{0,i}$, $i = \overline{0,N}$, $k = \overline{1,K}$, is computed by

$$\mathbf{v}(s) = -\mathbf{\Omega}^{-1}(s)\mathcal{B}(s). \tag{4.8}$$

Remark 4.2. It can be shown that the diagonal entries of the matrix $\Omega(s)$ dominate in each row. Thus (see [9]) the matrix $\Omega(s)$ is nonsingular for any s, $\operatorname{Re} s > 0$.

Corollary 4.3. *The mean sojourn time V_b of an arbitrary flow is computed by*

$$V_b = -\sum_{i=0}^{N-1} \sum_{k=0}^{K-1} \pi(i,k) \left. \frac{\partial v(s, i+1, 1, k+1)}{\partial s} \right|_{s=0}, \tag{4.9}$$

where the values $(\partial v(s, i+1, 1, k+1)/\partial s)|_{s=0}$ are computed as the corresponding entries of the vector $(d\mathbf{v}(s)/ds)|_{s=0}$, which is calculated by

$$\left. \frac{d\mathbf{v}(s)}{ds} \right|_{s=0} = \mathbf{\Omega}^{-1}(0) \left(- \left. \frac{d\mathcal{B}(s)}{ds} \right|_{s=0} + \mathbf{e} \right). \tag{4.10}$$

The mean sojourn time $V_b^{(accept)}$ *of an arbitrary accepted flow is computed by*

$$V_b^{(accept)} = \frac{V_b}{1 - P_b^{(loss)}}. \tag{4.11}$$

5. Loss probability of an arbitrary admitted flow

In this section, we derive the expression for the loss probability of an arbitrary admitted flow. Recall that the admitted flow can be lost, with probability q, due to the loss of any customer from this flow.

Let $u(s,i,n,k)$ denote the probability that a catastrophe will not arrive during the rest of the tagged flow sojourn time in the system conditional that, at the given instant, the number of flows processed in the system is equal to k, $k = \overline{1,K}$, the number of customers in the system is equal to i, $i = \overline{0,N}$, the number of customers from the tagged flow is equal to n, $l = \overline{0,i}$, and the flow will not be rejected (terminated) due to the loss of a particular customer from this flow.

Analogous to the previous section, we can prove the following statement.

Theorem 5.1. *The loss probability* $P_a^{(loss)}$ *of an arbitrary admitted flow is calculated by*

$$P_a^{(loss)} = 1 - \frac{\sum_{i=0}^{N-1}\sum_{k=0}^{K-1}\boldsymbol{\pi}(i,k)u(0,i+1,1,k+1)}{1 - P_b^{(loss)}}, \tag{5.1}$$

where the functions $u(s,i,n,k)$, $n = \overline{0,i}$, $i = \overline{0,N}$, $k = \overline{1,K}$, *are the entries of the vector* $\mathbf{u}(s)$ *which is calculated by formula*

$$\mathbf{u}(s) = -\boldsymbol{\Omega}^{-1}(s)\widetilde{\mathcal{B}}(s), \tag{5.2}$$

where $\widetilde{\mathcal{B}}(s)$ *is the vector defined by*

$$\widetilde{\mathcal{B}}(s) = \left(\mathcal{B}_0(s), \ldots, \mathcal{B}_N(s)\right)^T. \tag{5.3}$$

Remark 5.2. It is worthy of mention that, in contrast to the classical Erlang loss model, where the loss probability can be calculated by a simple analytical expression, calculation of the loss probability of an arbitrary flow, and an arbitrary admitted flow in the model with flow arrivals, requires the development of an algorithmic tool. Such a tool is the main contribution of this paper.

6. Numerical examples

To demonstrate the feasibility of the proposed means of the characteristics calculation and to provide insight into the dependence of the main system performance measures on its parameters and the number of token K, we present the results of various numerical experiments.

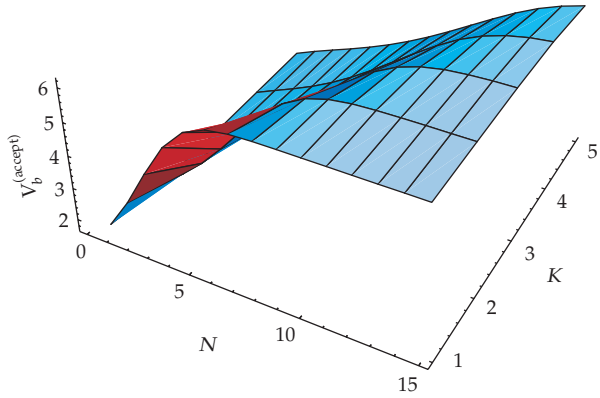

Figure 1: Dependence of the mean sojourn time $V_b^{(accept)}$ of an arbitrary admitted flow on the number of servers N and the number of tokens K.

Experiment 6.1. In this experiment, we fix the following parameters of the system: $q = 0.3,\ \lambda = 2,\ \mu = 1,\ \gamma = 2,\ \theta = 0.9$. Figure 1 shows the dependence of the mean sojourn time $V_b^{(accept)}$ of an arbitrary admitted flow on the number of servers N and the number of tokens K.

Note that the mean sojourn time $V_b^{(accept)}$ increases as the number of servers N increases. This surprising effect is easily explained as follows. The sojourn time of an arbitrary admitted flow increases with increasing N, because the probability of the flow termination due to the loss of a customer from this flow decreases with increasing N.

Experiment 6.2. In this experiment, we fix the following parameters of the system $\lambda = 1,\ \mu = 2,\ \gamma = 2,\ \theta = 0.9$ and assume that the number of tokens $K = 5$. Figure 2 illustrates the dependence of the loss probability $P_a^{(loss)}$ of an arbitrary admitted flow on the number of servers N, and the probability q of flow termination when a customer from this flow is rejected.

As is expected, the loss probability of an arbitrary admitted flow increases when the probability q of flow termination due to its customer rejection increases. This effect declines as the number of servers increases.

Experiment 6.3. In this experiment, we show the effect of changing the parameter θ, which defines the number of customers in a flow. We fix the same parameters as in the previous experiment, and set $q = 0.6$. Figure 3 illustrates the dependence of the loss probability $P_a^{(loss)}$ of an arbitrary admitted flow on the number of servers N and the parameter θ.

The loss probability of an arbitrary admitted flow increases as the parameter θ (and average number of customers in a flow) increases. This effect declines as the number of servers increases.

Experiment 6.4. It is evident that the decision maker controlling the operation of the model with the flow arrival aims to increase the number of tokens K, to provide the highest throughput of the system. However, an increase of K evidently leads to greater competition between flows, and an increase of the loss probability $P_a^{(loss)}$ of an arbitrary admitted flow. To protect the interests of users, some QoS requirements should be imposed. The natural

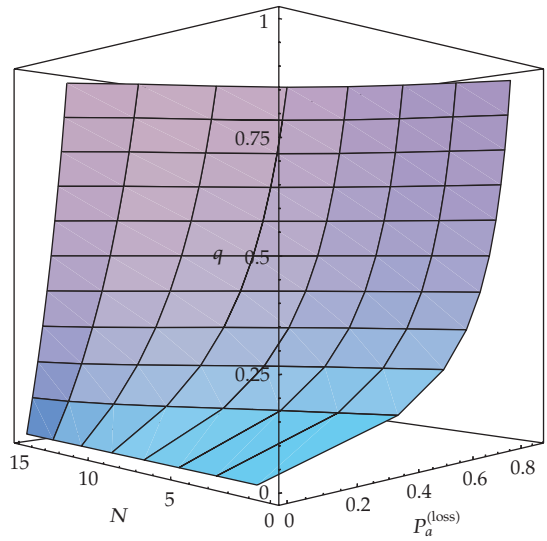

Figure 2: Dependence of the loss probability $P_a^{(loss)}$ of an arbitrary admitted flow on the number of servers N and the probability q.

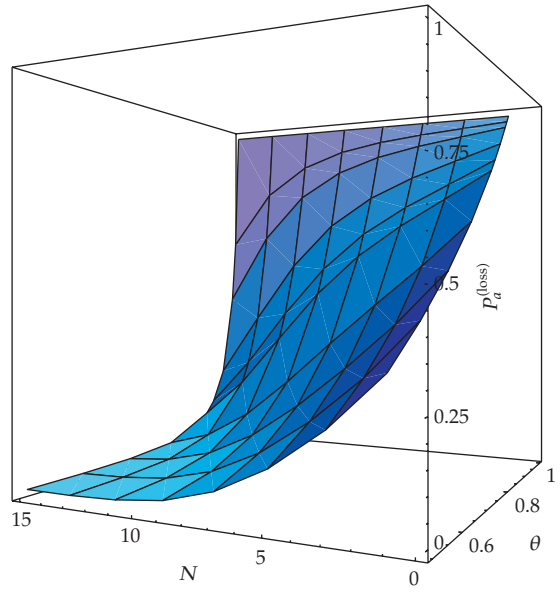

Figure 3: Dependence of the loss probability $P_a^{(loss)}$ of an arbitrary admitted flow on the number of servers N and the parameter θ.

requirement is to ensure that the loss probability of an arbitrary admitted flow is less than a particular preassigned level ε.

So the following optimization problem arises:

$$K \longrightarrow \max \tag{6.1}$$

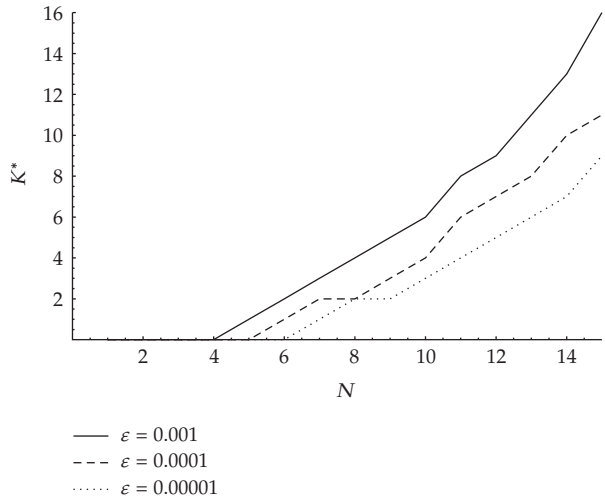

Figure 4: Dependence of the optimal number K^* of tokens on the number of servers N.

subject to restriction

$$P_a^{(\text{loss})} < \varepsilon. \tag{6.2}$$

In this experiment, we consider this optimization problem, and illustrate the effect of the number of servers N, probability q of the flow termination when a customer from this flow is rejected, and parameter θ that characterizes the distribution of the number of customers in a flow.

Firstly, we fix the following parameters of the system: $q = 0.3,\ \lambda = 1,\ \mu = 5,\ \gamma = 2,\ \theta = 0.95$. Figure 4 shows the dependence of the value K^*, which is the solution to problem (6.1), (6.2), on the number of servers N for various admissible values ε of the loss probability.

It is clear from this figure that if the number N of servers is less than or equal to four, even when the number of tokens $K = 1$, the loss probability $P_a^{(\text{loss})}$ is greater than $\varepsilon,\ \varepsilon \leq 0.001$. For $N = 5$, the loss probability $P_a^{(\text{loss})} < 0.001$ can be guaranteed. For $N = 6$ and $N = 7$ correspondingly, the loss probability $P_a^{(\text{loss})} < 0.0001$ and $P_a^{(\text{loss})} < 0.00001$ can be guaranteed. By increasing the number of servers N, the number K^* can be rapidly increased.

Secondly, we fix the followingparameters of the system: $N = 15,\ \lambda = 1,\ \mu = 5,\ \gamma = 2,\ \theta = 0.95$. Figure 5 shows the dependence of K^*, which is the solution to problem (6.1), (6.2), on the probability q of flow termination, when a customer from this flow is rejected, for various values ε of admissible loss probability.

As is expected, an increase of the probability q implies a decrease of K^*.

Finally, we fix the following parameters of the system: $N = 10,\ \lambda = 1,\ \mu = 2,\ \gamma = 2,\ q = 0.4$. Figure 6 shows the dependence of K^* on the parameter θ.

Because the mean number of customers in a flow is equal to $(1-\theta)^{-1}$, this number rapidly increases when θ approaches 1. Correspondingly, K^*, which is the solution to problem (6.1), (6.2), rapidly decreases. For values of θ equal to 0.997, 0.9997, 0.99997, that correspond to the mean number of customers in a batch equal to 330, 3330, and 33330, respectively,

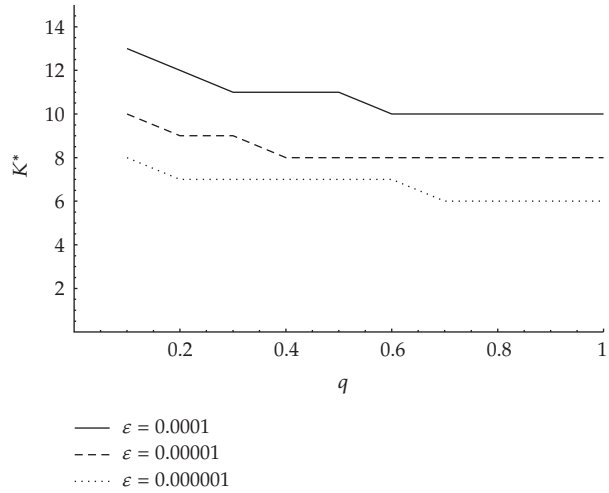

Figure 5: Dependence of the optimal number K^* of tokens on the probability q.

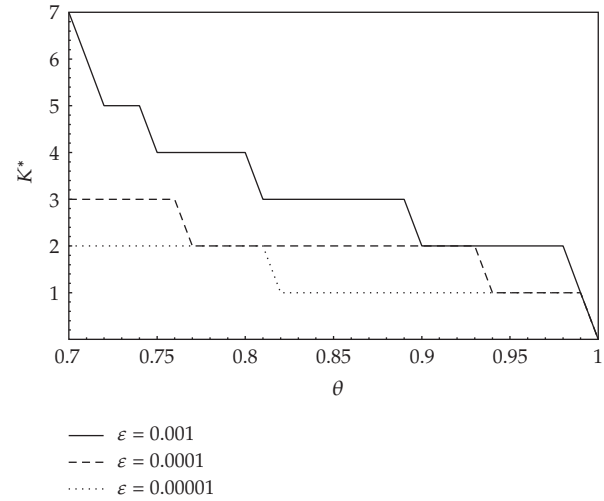

Figure 6: Dependence of K^* of tokens on the parameter θ.

the system cannot provide the required QoS, even when $K = 1$, where the values of ε are 10^{-5}, 10^{-4}, 10^{-3}, respectively.

7. Conclusion

We analyzed the sojourn time distribution, loss probability of an arbitrary flow, and arbitrary admitted flow in a multiserver loss queueing system with a flow arrival of customers. The results can be exploited for performance evaluation and capacity planning of systems of *IP* telephony and other systems of information transmission that do not have buffers but have time distributed (flow) arrival of customers. Results can be easily extended to the cases of the *MAP* arrival process of flows and phase-type service time distribution. However, this particular account of the phase-type service time distribution results in an inherent increase

of the state space of the Markov chain under study. Results can also be generalized to the case of an arbitrary distribution of the number of customers in a flow. However, in this case, the state space of the Markov chain also inherently increases, because it is necessary to record the number of customers that previously arrived in each flow. The alternative means of weakening the assumptions about the exponential distribution of interarrival times in a flow and the geometric batch size distribution is to assume a phase-type mechanism of customers arrival in a flow. Customer arrivals occur at the instants of transitions of the underlying Markov process of the phase-type distribution. Analysis of such a model is in progress.

Acknowledgment

This research was supported by Korea Research Foundation, Grant KRF-2007-521-D00330, and Small and Medium Business Administration in Korea.

References

[1] B. A. Sevastjanov, "Erlang formula in telephone systems under the arbitrary distribution of conversations duration," in *Proceedings of the 3rd All-Union Mathematical Congress*, vol. 4, pp. 68–70, Academy of Science, Moscow, Russia, 1959.

[2] C. S. Kim, A. Dudin, V. Klimenok, and V. Khramova, "Erlang loss queueing system with batch arrivals operating in a random environment," *Computers & Operations Research*, vol. 36, no. 3, pp. 674–697, 2009.

[3] V. Klimenok, "Characteristics calculation for multi-server queue with losses and bursty traffic," *Automatic Control and Computer Sciences*, vol. 32, pp. 393–415, 1999.

[4] V. Klimenok, C. S. Kim, D. Orlovsky, and A. Dudin, "Lack of invariant property of the Erlang loss model in case of *MAP* input," *Queueing Systems*, vol. 49, no. 2, pp. 187–213, 2005.

[5] A. A. Kist, B. Lloyd-Smith, and R. J. Harris, "A simple IP flow blocking model," in *Proceedings of the 19th International Teletraffic Congress on Performance Challenges for Efficient Next Generation Networks (ITC '05)*, pp. 355–364, Beijing, China, August-September 2005.

[6] M. H. Lee, S. Dudin, and V. Klimenok, "Queueing model with time-phased batch arrivals," in *Proceedings of the 20th International Teletraffic Congress on Managing Traffic Performance in Converged Networks (ITC '07)*, vol. 4516 of *Lecture Notes in Computer Science*, pp. 719–730, Ottawa, Canada, June 2007.

[7] H. Kasten and J. Th. Runnenburg, *Priority in Waiting Line Problems*, Mathematisch Centrum, Amsterdam, The Netherlands, 1956.

[8] D. van Dantzig, "Chaînes de Markof dans les ensembles abstraits et applications aux processus avec régions absorbantes et au problème des boucles," *Annales de l'Institut Henri Poincaré*, vol. 14, pp. 145–199, 1955.

[9] F. R. Gantmakher, *The Matrix Theory*, Science, Moscow, Russia, 1967.