Journal of Applied Mathematics and Stochastic Analysis, **12**:4 (1999), 371-392.

TRANSIENT ANALYSIS OF A QUEUE WITH QUEUE-LENGTH DEPENDENT MAP AND ITS APPLICATION TO SS7 NETWORK[1]

BONG DAE CHOI[2] and SUNG HO CHOI
KAIST, Department of Mathematics and Center for Applied Math
373-1 Kusong-Dong, Yusong-Gu, Taejon 305-701 Korea
e-mail: bdchoi@math.kaist.ac.kr, shchoi@mathx.kaist.ac.kr

DAN KEUN SUNG
KAIST, Department of Electrical Engineering
373-1 Kusong-Dong, Yussong-Gu, Taejon
305-701 Korea

TAE-HEE LEE and KYU-SEOG SONG
Korea Telecom, Telecommunications Network Laboratory
463-1, Junmin-Dong, Yusong-Gu, Taejon 305-390 Korea
E-mail: thlee@tnrl.kotel.co.kr, kssong@tnrl.kotel.co.kr

(Received March, 1999; Revised August, 1999)

We analyze the transient behavior of a Markovian arrival queue with congestion control based on a double of thresholds, where the arrival process is a queue-length dependent Markovian arrival process. We consider Markov chain embedded at arrival epochs and derive the one-step transition probabilities. From these results, we obtain the mean delay and the loss probability of the nth arrival packet. Before we study this complex model, first we give a transient analysis of an MAP/M/1 queueing system without congestion control at arrival epochs. We apply our result to a signaling system No. 7 network with a congestion control based on thresholds.

Key words: Transient Analysis of Queue, MAP, Congestion Control, SS7 Network.

AMS subject classifications: 60K25, 60K30, 68M20, 90B12.

---

[1]This work was supported in part by KT under Grant 97-5.

[2]Bong Dae Choi is now with the Department of Mathematics, Korea University, Seoul Korea.

1. Introduction

Congestion control based on thresholds [4, 7-10, 15] is aimed to control the traffic causing overload before a significant delay builds up in the network and so to satisfy the quality of service (Qos) requirements of the different classes of traffic. The QoS requirements are often determined by two parameters; the loss probability and the mean delay [5]. S.Q. Li [10] proposed a congestion control with double thresholds consisting of an abatement threshold and an onset threshold to regulate the input rate according to the congestion status. Packets are classified as one of the two priorities: high priority and low priority. When the queue length exceeds the onset threshold, the low priority packets are blocked and lost until the queue length decreases to the abatement threshold. O.C. Ibe and J. Keilson [8] extended this model to the system with N doubles of thresholds ($N \geq 1$) and N different priority packets. For the above systems, they assumed that the arrival processes are a Poisson process [8] and a Markov modulated Poisson process (MMPP) [10], and they obtained the steady state characteristics.

In order to find out performance of a congestion control, first we need to analyze the transient behavior of the system. The Laplace transform and z-transform methods [1, 6, 14, 16] are usually applied within conventional transient analysis. It seems not to be easy to analyze the transient behavior of the system with finite buffer and congestion control based on thresholds by the above transform methods. For transient analysis of such a system, D.S. Lee and S.Q. Li [11, 12] used a discrete time analysis with its time indexed by packet arrivals. They assumed the arrival processes are MMPP [11] and a switched Poisson arrival process [12] and obtained the one-step transition probabilities of an embedded Markov chain. But they considered a congestion control with only one threshold called *partial buffer sharing policy*.

In this paper, we consider the congestion control with double thresholds as in [9] and [10]. We assume that the arrival process is a queue-length dependent Markovian arrival process (MAP). The motivation of this model comes from the study of the congestion control in a signaling system No. 7 (SS7) network [15]. A congestion control called *international control* in a SS7 network is a reactive control with double thresholds, which uses a notification to inform senders about the congestion status of the system. Each sender regulates its traffic load to the system when he receives a notification, and uses timers to resume its traffic load. For such a system, the arrival process can be modeled as a queue-length dependent Markovian arrival process (MAP) [4].

For a transient analysis, we use the discrete time analysis as in [11, 12] but using the advantage of simple notations of MAPs we obtain the one-step transition probabilities by a simpler derivation than that of D.S. Lee and S.Q. Li in [11, 12]. The models dealt with in [11, 12] are special cases of our model. We obtain the mean delay and loss probability of the nth arrival packets. In order to evaluate the performance measures we give an algorithm, which enables us to reduce the complexity of iterated Kolmogorov equation in Section 3. We apply our result to analyze the international control in SS7 networks. In the numerical examples, we show the impact of various parameters such as the value of the thresholds and the length of times and input rates on the transient performances.

This paper is organized as follows: In Section 2, we give a transient analysis of an MAP/M/1 queueing system at arrival epochs in order to provide better understanding of the main result of Section 3. The one-step transition probabilities are derived by using matrix formulation. In Section 3, we consider the congestion control based on double thresholds with queue-length dependent MAP and derive the transient queue length probability at arrival epochs. We give performance measures and an algorithm to compute the performance measures. In Section 4, we describe an analytic modeling of the international control in SS7 networks and present numerical examples and observations.

2. Transient Analysis of MAP/M/1 Queueing System at Arrival Epochs

In this section we study an MAP/M/1 queueing system without any congestion control to provide a better understanding of the system with congestion control in Section 3. We assume that the packets arrive according to a Markovian arrival process (MAP) with representation $(\mathbf{C},\mathbf{D})$, where $\mathbf{C}$ and $\mathbf{D}$ are $m\times m$ matrices [2, 3, 13]. Here $\mathbf{C}$ is the rate matrix of state transitions without an arrival and $\mathbf{D}$ is the rate matrix of state transitions with an arrival. The service time of a packet is assumed to be exponentially distributed with parameter μ. We denote the number of packets in the system and the state of the underlying Markov chain of the MAP at time t by $N(t)$ and $J(t)$ $(1\le J(t)\le m)$, respectively. Then the two-dimensional process $X(t)=(N(t),J(t))$ forms a continuous time Markov chain. Let T_n denote the nth packet arrival epoch. Then we form an embedded Markov chain $\{X_n \mid n\ge 0\}$ defined by

$$X_n=(N(T_n+),J(T_n+)).$$

Let the nth step transition probabilities from X_0 to X_n be denoted by

$$P^n_{j_0,j}(i_0,i)\triangleq P\{X_n=(i,j)\mid X_0=(i_0,j_0)\}$$

and in a matrix form by

$$\mathbf{P}^n(i_0,i)\triangleq [P^n_{j_0,j}(i_0,i)]$$

for $1\le i\le i_0+n$. Let the one-step transition probabilities be denoted by

$$P_{j_0,j}(i_0,i)\triangleq P^1_{j_0,j}(i_0,i)$$

and in a matrix form by

$$\mathbf{P}(i_0,i)\triangleq \mathbf{P}^1(i_0,i).$$

The Chapman-Kolmogorov equations for $\{X_n\}$ are

$$\mathbf{P}^n(i_0,i)=\sum_{k=\max(1,i+1-n)}^{i_0+1}\mathbf{P}(i_0,k)\mathbf{P}^{n-1}(k,i)\text{ for }1<i\le i_0+n. \tag{1}$$

Hence, the queue length probability $\mathbf{P}^n(i_0,i)$ at the nth arrival epoch can be obtained from (1) recursively once the on-step transition probability $\mathbf{P}(i_0,i)$ is known.

2.1 One-Step Transition Probabilities

For convenience, we define the set of all states (i,j), $1 \leq j \leq m$, by level i. Let S be the exit time from level i_0, i.e.,

$$S = \inf(t: N(t) \neq i_0 \mid N(0) = i_0).$$

Then we have the following lemma.

Lemma 1: *For $i_0 \geq 1$,*

$$\int_0^\infty P\{S > s, X(s) = (i_0, k) \mid X(0) = (i_0, j_0)\} ds = ((\mu\mathbf{I} - \mathbf{C})^{-1})_{j_0, k} \tag{2}$$

and

$$\int_0^\infty P\{S > s, X(s) = (0, k) \mid X(0) = (0, j_0)\} ds = (-\mathbf{C}^{-1})_{j_0, k}.$$

Note that the left-hand side of (2) is the expected time that $X(t)$ spends in state (i_0, k) until the Markov chain departs from its level i_0 starting from state (i_0, j_0).

Proof: See Appendix 5.1.

Let the transition probabilities of the underlying Markov chain $\{J(t) \mid t \geq 0\}$ with the first transition of level be denoted by

$$R_{j_0, j} \triangleq P\{X(S) = (i_0 - 1, j) \mid X(0) = (i_0, j_0)\}$$

$$G_{j_0, j} \triangleq P\{X(S) = (i_0 + 1, j) \mid X(0) = (i_0, j_0)\}$$

for $i_0 > 1$, and

$$H_{j_0, j} \triangleq P\{X(S) = (1, j) \mid X(0) = (0, j_0)\}.$$

Define matrices $\mathbf{R}, \mathbf{G}$, and $\mathbf{H}$ as

$$\mathbf{R} = [R_{j_0, j}], \mathbf{G} = [G_{j_0, j}], \mathbf{H} = [H_{j_0, j}],$$

which are all $m \times m$ matrices. Lemma 1 yields the following.

Lemma 2:

$$\mathbf{R} = (\mu\mathbf{I} - \mathbf{C})^{-1}\mu\mathbf{I} \tag{3}$$

$$\mathbf{G} = (\mu\mathbf{I} - \mathbf{C})^{-1}\mathbf{D} \tag{4}$$

$$\mathbf{H} = (-\mathbf{C})^{-1}\mathbf{D}. \tag{5}$$

Proof: See Appendix 5.2.

Now we are ready to arrive at the one-step transition probabilities.

Proposition 3: *For $i \leq i_0 + 1$,*

$$\mathbf{P}(i_0, i) = \begin{cases} \mathbf{R}^{i_0 - i + 1}\mathbf{G} & \textit{if } i > 1 \\ \mathbf{R}^{i_0}\mathbf{H} & \textit{if } i = 1. \end{cases}$$

Proof: We will show that

$$\mathbf{P}(i_0, i) = \mathbf{RP}(i_0 - 1, i) \quad \text{for } i_0 \geq i$$

and

$$\mathbf{P}(i-1, i) = \begin{cases} \mathbf{G} & \text{if } i > 1 \\ \mathbf{H} & \text{if } i = 1. \end{cases}$$

(*i*) First, consider the case of $i_0 \geq i$: In this case, the next arrival occurs after at least one service completion and, therefore, $T_1 > S$. By the strong Markov property, we have

$$P\{X(T_1) = (i, j) \mid X_0 = (i_0, j_0), X(S) = (i_0 - 1, k)\} = P_{k,j}(i_0 - 1, i). \tag{6}$$

Using (6), we have

$$P_{j_0, j}(i_0, i) = P\{X(T_1) = (i, j) \mid X_0(i_0, j_0)\}$$

$$= \sum_{k=1}^{m} P\{X(S) = (i_0 - 1, k) \mid X_0 = (i_0, j_0)\}$$

$$\cdot P\{X(T_1) = (i, j) \mid X_0 = (i_0, j_0), X(S) = (i_0 - 1, k)\}$$

$$= \sum_{k=1}^{m} R_{j_0, k} P_{k, j}(i_0 - 1, i).$$

The above equation can be rewritten in the matrix form as

$$\mathbf{P}(i_0, i) = \mathbf{RP}(i_0 - 1, i). \tag{7}$$

(*ii*) Secondly, consider the case of $i_0 = i - 1$: In this case, the next transition occurs with an arrival, i.e., $S = T_1$. From the definitions of $\mathbf{G}$ and $\mathbf{H}$, we have

$$P_{j_0, j}(i-1, i) = P\{X(T_1) = (i, j) \mid X_0 = (i-1, j_0)\}$$

$$= P\{X(S) = (i, j) \mid X_0 = (i-1, j_0)\}$$

$$= \begin{cases} G_{j_0, j} & i > 1 \\ H_{j_0, j} & i = 1. \end{cases}$$

The above equation can be rewritten in the matrix form as

$$\mathbf{P}(i-1,i)=\begin{cases}\mathbf{G} & \text{if } i>1\\ \mathbf{H} & \text{if } i=1.\end{cases} \tag{8}$$

From (7) and (8), we get

$$\mathbf{P}(i_0,i)=\mathbf{RP}(i_0-1,i)=\ldots$$

$$=\begin{cases}\mathbf{R}^{i_0-i+1}\mathbf{P}(i-1,i)=\mathbf{R}^{i_0-i+1}\mathbf{G} & \text{for } i>1\\ \mathbf{R}^{i_0}\mathbf{P}(0,1)=\mathbf{R}^{i_0}\mathbf{H} & \text{for } i=1.\end{cases}$$

□

2.2 Special Cases

From Proposition 3, we can obtain the following corollaries, which agree with the results in [11, 12].

Corollary 4: (Corollary 3 in [11]) *If the arrival process is an MMPP with the representation* $(\mathbf{Q},\Lambda)$, *where* $\mathbf{Q}$ *is the infinitesimal generator of the underlying Markov chain, and* Λ *is the arrival rate matrix, then for* $i\le i_0+1$,

$$\mathbf{P}(i_0,i)=\begin{cases}((\mu\mathbf{I}-\mathbf{Q}+\Lambda)^{-1}\mu\mathbf{I})^{i_0-i+1}(\mu\mathbf{I}-\mathbf{Q}+\Lambda)^{-1}\Lambda & i>1\\ ((\mu\mathbf{I}-\mathbf{Q}+\Lambda)^{-1}\mu\mathbf{I})^{i_0}(\Lambda-\mathbf{Q})^{-1}\Lambda & i=1.\end{cases}$$

Proof: An MMPP with the representation $(\mathbf{Q},\Lambda)$ is an MAP with the representation $(\mathbf{C},\mathbf{D})$, where

$$\mathbf{C}=\mathbf{Q}-\Lambda,\quad \mathbf{D}=\Lambda.$$

Therefore, the result follows from (3) and Proposition 3 directly. □

Corollary 5: (Proposition 2.1 in [12]) *If the arrival process is a switched Poisson process with the representation* $(\mathbf{Q},\Lambda)$, *where*

$$\mathbf{Q}=\begin{pmatrix}-r_0 & r_0\\ r_1 & -r_1\end{pmatrix}\qquad \Lambda=\begin{pmatrix}\lambda_0 & 0\\ 0 & \lambda_1\end{pmatrix},$$

then for $i\le i_0+1$ *and* $i>1$,

$$\mathbf{P}(i_0,i)=\begin{pmatrix}\alpha_0 z_1^{i-i_0-1}+\beta_0 z_2^{i-i_0-1} & \alpha_1' z_1^{i-i_0-1}+\beta_1' z_2^{i-i_0-1}\\ \alpha_0' z_1^{i-i_0-1}+\beta_0' z_2^{i-i_0-1} & \alpha_1 z_1^{i-i_0-1}+\beta_1 z_2^{i-i_0-1}\end{pmatrix},$$

where for $j=0,1$,

$$\alpha_j=\frac{\lambda_j(\mu+\lambda_{1-j}+r_{1-j}-\mu z_1)}{\mu^2 z_1(z_2-z_1)}\qquad \beta=\frac{\lambda_j(\mu+\lambda_{1-j}+r_{1-j}-\mu z_2)}{\mu^2 z_2(z_1-z_2)}$$

$$\alpha'_j = \frac{\lambda_j r_{1-j}}{\mu^2 z_1(z_2 - z_1)} \qquad \beta' = \frac{\lambda_j r_{1-j}}{\mu^2 z_2(z_1 - z_2)}$$

and z_1, z_2 *are the roots of the quadratic equation*

$$(\mu + \lambda_0 + r_0 - \mu z)(\mu + \lambda_1 + r_1 - \mu z) - r_0 r_1 = 0.$$

Proof: Since

$$\mathbf{R} = (\mu \mathbf{I} - \mathbf{Q} + \Lambda)^{-1}\mu \mathbf{I} = (\mathbf{I} - \tfrac{1}{\mu}\mathbf{Q} + \tfrac{1}{\mu}\Lambda)^{-1}$$

$$\mathbf{G} = (\mu \mathbf{I} - \mathbf{Q} + \Lambda)^{-1}\Lambda = (\mathbf{I} - \tfrac{1}{\mu}\mathbf{Q} + \tfrac{1}{\mu}\Lambda)^{-1}\tfrac{1}{\mu}\Lambda,$$

we have

$$\mathbf{P}(i_0, i) = \mathbf{R}^{i_0 - i + 1}\mathbf{G} = [(I - \tfrac{1}{\mu}\mathbf{Q} + \tfrac{1}{\mu}\Lambda)^{-1}]^{i_0 - i + 2}\tfrac{1}{\mu}\Lambda.$$

Here,

$$\mathbf{I} - \tfrac{1}{\mu}\mathbf{Q} + \tfrac{1}{\mu}\Lambda = \begin{pmatrix} \frac{\mu + r_0 + \lambda_0}{\mu} & \frac{-r_0}{\mu} \\ \frac{-r_1}{\mu} & \frac{\mu + r_1 + \lambda_1}{\mu} \end{pmatrix}$$

is diagonalizable with eigenvalues z_1 and z_2 which are roots of the quadratic equation

$$(\mu + \lambda_0 + r_0 - \mu z)(\mu + \lambda_1 - r_1 - \mu z) - r_0 r_1 = 0$$

i.e.,

$$\mathbf{I} - \tfrac{1}{\mu}\mathbf{Q} + \tfrac{1}{\mu}\Lambda = \mathbf{U}\begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}\mathbf{U}^{-1},$$

where

$$\mathbf{U} = \begin{pmatrix} \mu + r_1 + \lambda_1 - \mu z_1 & r_0 \\ r_1 & \mu + r_0 + \lambda_0 - \mu z_2 \end{pmatrix},$$

whose columns are the eigenvectors of $\mathbf{I} - \frac{1}{\mu}\mathbf{Q} + \frac{1}{\mu}\Lambda$ with respect to z_1 and z_2, respectively. Then by simple calculation, we can obtain

$$\mathbf{P}(i_0, i) = \mathbf{U}\begin{pmatrix} z_0^{i - i_0 - 2} & 0 \\ 0 & z_2^{i - i_0 - 2} \end{pmatrix}\mathbf{U}^{-1}\tfrac{1}{\mu}\Lambda$$

$$= \begin{pmatrix} \alpha_0 z_1^{i - i_0 - 1} + \beta_0 z_2^{i - i_0 - 1} & \alpha'_1 z_1^{i - i_0 - 1} + \beta'_1 z_2^{i - i_0 - 1} \\ \alpha'_0 z_1^{i - i_0 - 1} + \beta'_0 z_2^{i - i_0 - 1} & \alpha_1 z_1^{i - i_0 - 1} + \beta_1 z_2^{i - i_0 - 1} \end{pmatrix}. \qquad \square$$

For $i = 1$, we can obtain similarly

$$\mathbf{P}(i_0, 1) = \mathbf{U}\begin{pmatrix} z_1^{i_0} & 0 \\ 0 & z_2^{i_0} \end{pmatrix}\mathbf{U}^{-1}\Lambda.$$

This agrees with the result of Proposition 2.2 in [12].

3. Transient Analysis of the Congestion Control with Double Threshold

In this section we consider a congestion control with double thresholds consisting of an onset threshold M and an abatement threshold L (see Figure 1).

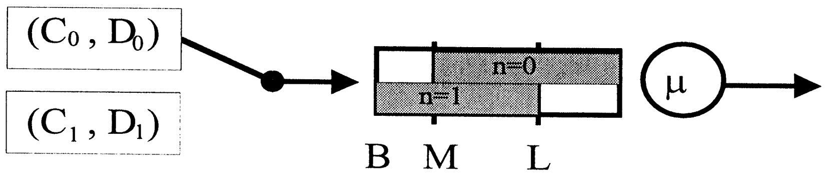


Figure 1: The model of the congestion control with two thresholds

Let B denote the buffer size. Until the queue length reaches the threshold M, the congestion status of the buffer is assigned to 0. Once the queue length exceeds M from below, we assign the congestion status of the buffer to 1 during the period until the queue length crosses L from above. When the queue length crosses L from above, the congestion status of the buffer is assigned to 0 again and the procedure is repeated. We assume that the packets arrive according to an MAP with the representation $(\mathbf{C}_0, \mathbf{D}_0)$, when the congestion status of the buffer is 0 and the packets arrive according to an MAP with the representation $(\mathbf{C}_1, \mathbf{D}_1)$, when the congestion status of the buffer is 1 and $\mathbf{C}_0, \mathbf{D}_0, \mathbf{C}_1$ and $\mathbf{D}_1$ are $m \times m$ matrices. We will describe the matrices $\mathbf{C}_0, \mathbf{D}_0$, $\mathbf{C}_1$ and $\mathbf{D}_1$ in detail for modeling of congestion control in SS7 networks in Section 4.

Let $I(t)$ denote the congestion status of the buffer at time t. Then $X(t) = (I(t), N(t), J(t))$ forms a continuous time Markov chain and $X_n = (I(T_n+), N(T_n+), J(T_n+))$ forms an embedded Markov chain of the Markov chain $\{X(t) \mid t \geq 0\}$, where T_n is the nth packet arrival epoch. Let the nth step transition probabilities be denoted by $\mathbf{P}^n(\xi_0, i_0; \xi, i)$ which is an $m \times m$ matrix, where

$$[\mathbf{P}^n(\xi_0, i_0; \xi, i)]_{j_0, j} = P\{X_n = (\xi, i, j) \mid X_0 = (\xi_0, i_0, j_0)\}.$$

Let the one-step transition probabilities of X_n be denoted by

$$\mathbf{P}(\xi_0, i_0; \xi, i) = \mathbf{P}^1(\xi_0, i_0; \xi, i).$$

Then we have the following Chapman-Kolmogorov's equations: For $1 \leq i_0 \leq M$

$$\mathbf{P}^n(0, i_0; \xi, i) = \sum_{k = \max(1, i+1-n)}^{\min(i_0+1, M)} \mathbf{P}(0, i_0; 0, k)\mathbf{P}^{n-1}(0, k; \xi, i)$$

$$+1_{\{i_0 = M\}}\mathbf{P}(0, M; 1, M+1)\mathbf{P}^{n-1}(1, M+1; \xi, i), \tag{9}$$

and, for $L+1 \leq i_0 \leq B$,

$$\mathbf{P}^n(1, i_0; \xi, i) = \sum_{k = \max(L+1, i+ \ -n)}^{\min(i_0+1, B)} \mathbf{P}(1, i_0; 1, k)\mathbf{P}^{n-1}(1, k; \xi, i)$$

$$+\sum_{k = \max(1, i+1-n)}^{L} \mathbf{P}(1, i_0; 0, k)\mathbf{P}^{n-1}(0, k; \xi, i), \tag{10}$$

where 1_A is the indication function of set A. Hence, the queue length probability $\mathbf{P}^n$ $(\xi_0, i_0; \xi, i)$ at the nth arrival epoch can be obtained iteratively once the one-step transition probability matrices $\mathbf{P}(\xi_0, i_0; \xi, i)$ are known.

3.1 One-Step Transition Probabilities

Let the exit times from levels be denoted by

$$S_0 = \inf(t: N(t) \neq i_0 \mid N(0) = i_0, \xi(0) = 0) \quad \text{for } 0 \leq i_0 \leq M$$

$$S_1 = \inf\,(t: N(t) \neq i_0 \mid N(0) = i_0, \xi(0) = 1) \quad \text{for } L+1 \leq i_0 \leq B.$$

Let the transition probabilities of $J(t)$ with the first transition of $N(t)$ be denoted by

$$R_{j_0, j} \triangleq P\{X(S_0) = (0, i_0 - 1, j) \mid X(0) = (0, i_0, j_0)\} \quad (i_0 \neq 0)$$

$$G_{j_0, j} \triangleq P\{X(S_0) = (0, i_0 + 1, j) \mid X(0) = (0, i_0, j_0)\} \quad (i_0 \neq M)$$

$$H_{j_0, j} \triangleq P\{X(S_0) = (0, 1, j) \mid X(0) = (0, 0, j_0)\}$$

$$\widetilde{R}_{j_0, j} \triangleq P\{X(S_1) = (1, i_0 - 1, j) \mid X(0) = (1, i_0, j_0)\} \quad (i_0 \neq L+1)$$

$$\widetilde{G}_{j_0, j} \triangleq P\{X(S_1) = (1, i_0 + 1, j) \mid X(0) = (1, i_0, j_0)\} \quad (i_0 \neq B)$$

$$\widetilde{H}_{j_0, j} \triangleq P\{X(S_1) = (0, B-1, j) \mid X(0) = (0, B, j_0)\}.$$

Define the matrices

$$\mathbf{R}_0 = [R_{j_0, j}], \mathbf{G}_0 = [G_{j_0, j}], \mathbf{H}_0 = [H_{j_0, j}]$$

$$\mathbf{R}_1 = [\widetilde{R}_{j_0, j}], \mathbf{G}_1 = [\widetilde{G}_{j_0, j}], \mathbf{H}_1 = [\widetilde{H}_{j_0, j}]$$

which are all $m \times m$ matrices. Then we can establish the following lemma similar to Lemma 2.

Lemma 6:

$$\mathbf{R}_0 = (\mu\mathbf{I} - \mathbf{C}_0)^{-1}\mu\mathbf{I}, \mathbf{G}_0 = (\mu\mathbf{I} - \mathbf{C}_0)^{-1}\mathbf{D}_0, \mathbf{H}_0 = (-\mathbf{C}_0)^{-1}\mathbf{D}_0$$

$$\mathbf{R}_1 = (\mu\mathbf{I} - \mathbf{C}_1)^{-1}\mu\mathbf{I}, \mathbf{G}_1 = (\mu\mathbf{I} - \mathbf{C}_1)^{-1}\mathbf{D}_1, \mathbf{H}_1 = \mathbf{G}_1.$$

Note that

$$P\{X(S_0) = (1, M+1, j) \mid X(0) = (0, M, j_0)\} = G_{j_0, j}$$

and

$$P\{X(S_1) = (0, L, j) \mid X(0) = (1, L+1, j_0)\} = \widetilde{R}_{j_0, j}.$$

Now we are ready to obtain the one-step transition matrix.

Proposition 7: *For* $1 \leq i_0 \leq M$,

$$\mathbf{P}(0, i_0; 0, i) = \begin{cases} \mathbf{R}_0^{i_0 - i + 1}\mathbf{G}_0 & 2 \leq i \leq M \\ \mathbf{R}_0^{i_0}\mathbf{H}_0 & i = 1, \end{cases}$$

$\mathbf{P}(0, M; 1, M+1) = \mathbf{G}_0$ *and for* $L + 1 \leq i_0 \leq B$,

$$\mathbf{P}(1, i_0; 0, i) = \begin{cases} \mathbf{R}_1^{i_0 - L}\mathbf{R}_0^{L - i + 1}\mathbf{G}_0 & 2 \leq i \leq L \\ \mathbf{R}_1^{i_0 - L}\mathbf{R}_0^{L}\mathbf{H}_0 & i = 1, \end{cases}$$

$$\mathbf{P}(1, i_0; 1, i) = \mathbf{R}_1^{i_0 - i + 1}\mathbf{G}_1 \quad L + 1 \leq i \leq B,$$

and $\mathbf{P}(1, B; 1, B) = \mathbf{G}_1 + \mathbf{R}_1\mathbf{G}_1$.

The proof is identical to that of Proposition 3 and is therefore omitted.

3.2 Performance Measures and Algorithm

Once the transient queue length probability $\mathbf{P}^n(\xi_0, i_0; \xi, i)$ at arrival epochs is obtained, the performance measures can be easily evaluated. Let d_n be the delay of the nth arrival packet and $\mathbf{e}_{j_0} = (0, \ldots, 0, \overbrace{1}^{j_0\text{th}}, 0, \ldots, 0)$. Under the initial condition $X_0 = (\xi_0, i_0, j_0)$, both the mean and the variance of d_n are obtained as

$$E\{d_n \mid X_0 = (\xi_0, i_0, j_0\} = \frac{1}{\mu} \sum_{i=1}^{\min(i_0 + n, M)} (i-1) \cdot \mathbf{e}_{j_0}\mathbf{P}^n(\xi_0, i_0; 0, i)\mathbf{e}$$

$$+ \frac{1}{\mu} \sum_{i = L+1}^{\min(i_0 + n, B)} (i-1) \cdot \mathbf{e}_{j_0}\mathbf{P}^n(\xi_0, i_0; 1, i)\mathbf{e} \tag{11}$$

$$Var\{d_n \mid X_0 = (\xi_0, i_0, j_0)\} = \frac{1}{\mu^2} \sum_{i=1}^{\min(i_0 + n, M)} (i-1)^2 \cdot \mathbf{e}_{j_0}\mathbf{P}^n(\xi_0, i_0; 0, i)\mathbf{e}$$

$$+\frac{1}{\mu^2}\sum_{i=L+1}^{\min(i_0+n,B)}(i-1)^2\cdot \mathbf{e}_{j_0}\mathbf{P}^n(\xi_0,i_0;1,i)\mathbf{e}$$

$$-[E\{d_n \mid X_0=(\xi_0,i_0,j_0)\}]^2, \tag{12}$$

where $\mathbf{e}$ is the column vector whose elements are all 1. Let $p^n_{\text{loss}}(\xi_0,i_0,j_0)$ denote the loss probability of the nth arrival packet under the initial condition $X_0=(\xi_0,i_0,j_0)$. Then

$$p^n_{\text{loss}}(\xi_0,i_0,j_0)=\mathbf{e}_{j_0}\cdot\mathbf{P}^{n-1}(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e}. \tag{13}$$

When the matrices $\mathbf{R}_0$ and $\mathbf{R}_1$ are diagonalizable, we can greatly reduce the complexity of the iterated Kolmogorov equations (9) and (10) using their eigenvalues and eigenvectors as in [11, 12]. But there is no evidence that there exist distinct and real eigenvalues of the matrices $\mathbf{R}_0$ and $\mathbf{R}_1$, and it is not easy to obtain the eigenvectors numerically. Therefore, we introduce another algorithm to reduce the complexity of the iterative Kolmogorov equations (9) and (10). To obtain the performance measures, we only need to calculate the column vectors $\mathbf{P}^n(\xi_0,i_0;\xi,i)\mathbf{e}$ in (11) and (12), and $\mathbf{P}^{n-1}(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e}$ in (13). Using the fact that $\mathbf{P}(0,i_0;0,i)=\mathbf{R}_0\mathbf{P}(0,i_0-1;0,i)$ and $\mathbf{P}(1,i_0;\xi,i)=\mathbf{R}_1\mathbf{P}(1,i_0-1;\xi,i)$, we can reduce the complexity to obtain $\mathbf{P}^n(\xi_0,i_0;\xi,i)\mathbf{e}$. First consider the case of $2\le i_0\le M-1$ and $\xi_0=0$. Since $\mathbf{P}(0,i_0;0,k)=\mathbf{R}_0,\mathbf{P}(0,i_0-1;0,k)$, from (9) we obtain

$$\mathbf{P}^n(0,i_0;\xi,i)\mathbf{e}=\sum_{k=\max(1,i+1-n)}^{i_0+1}\mathbf{P}(0,i_0;0,k)\mathbf{P}^{n-1}(0,k;\xi,i)\mathbf{e}$$

$$=\mathbf{P}(0,i_0;0,i_0+1)\mathbf{P}^{n-1}(0,i_0+1;\xi,i)\mathbf{e}$$

$$+\mathbf{R}_0\sum_{k=\max(1,i+1-n)}^{i_0}\mathbf{P}(0,i_0-1;0,k)\mathbf{P}^{n-1}(0,k;\xi,i)\mathbf{e}$$

$$=\mathbf{G}_0\mathbf{P}^{n-1}(0,i_0+1;\xi,i)\mathbf{e}+\mathbf{R}_0\mathbf{P}^n(0,i_0-1;\xi,i)\mathbf{e}.$$

Therefore, we can calculate $\mathbf{P}^n(0,i_0;\xi,i)\mathbf{e}$ iteratively starting with

$$\mathbf{P}^n(0,1;\xi,i)\mathbf{e}=\mathbf{R}_0\mathbf{H}_0\mathbf{P}^{n-1}(0,1;\xi,i)\mathbf{e}+\mathbf{G}_0\mathbf{P}^{n-1}(0,2;\xi,i)\mathbf{e},$$

once we have $\mathbf{P}^{n-1}(0,i_0;\xi,i)\mathbf{e}$ for all i_0 and i. Similarly, from (9) and (10) we can obtain the following:

$$\mathbf{P}^n(0,1;\xi,i)\mathbf{e}=\mathbf{G}_0\mathbf{P}^{n-1}(0,2;\xi,i)\mathbf{e}+\mathbf{R}_0\mathbf{H}_0\mathbf{P}^{n-1}(0,1;\xi,i)\mathbf{e}$$

$$\mathbf{P}^n(0,M;\xi,i)\mathbf{e}=\mathbf{G}_0\mathbf{P}^{n-1}(1,M+1;\xi,i)\mathbf{e}+\mathbf{R}_0\mathbf{P}^n(0,M-1;\xi,i)\mathbf{e}$$

$$\mathbf{P}^n(1,L+1;\xi,i)\mathbf{e}=\mathbf{G}_1\mathbf{P}^{n-1}(1,L+2;\xi,i)\mathbf{e}+\mathbf{R}_1\mathbf{G}_1\mathbf{P}^{n-1}(1,L+1;\xi,i)\mathbf{e}$$

$$+\mathbf{R}_1^2\mathbf{P}^n(0,L-1;\xi,i)\mathbf{e}$$

$$\mathbf{P}^n(1,i_0;\xi,i)\mathbf{e} = \mathbf{G}_1\mathbf{P}^{n-1}(1,i_0+1;\xi,i)\mathbf{e} + \mathbf{R}_1\mathbf{P}^n(1,i_0-1;\xi,i)\mathbf{e}$$

$$\text{for } L+2 \le i_0 \le B-1$$

$$\mathbf{P}^n(1,B;\xi,i)\mathbf{e} = \mathbf{G}_1\mathbf{P}^{n-1}(1,B;\xi,i)\mathbf{e} + \mathbf{R}_1\mathbf{P}^n(1,B-1;\xi,i)\mathbf{e}. \tag{14}$$

Once we obtain $\mathbf{P}^{n-1}(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e}$, we can calculate the loss probability $p^n_{\text{loss}}(\xi_0,i_0,j_0)$ from (13). By substituting $\mathbf{P}^{n-1}(\xi_0,i_0\xi,i)\mathbf{e}$ for $\mathbf{P}^{n-1}(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e}$ and $\mathbf{P}^n(\xi_0,i_0;\xi,i)\mathbf{e}$ for $\mathbf{P}^n(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e}$ in iteration (14) we can obtain $\mathbf{P}^n(\xi_0,i_01,B)\mathbf{G}_1\mathbf{e}$ iteratively starting with

$$\mathbf{P}(\xi_0,i_0;1,B)\mathbf{G}_1\mathbf{e} = \begin{cases} \mathbf{G}_1^2\mathbf{e} & i_0 = B-1, \xi_0 = 1 \\ \mathbf{R}_1\mathbf{G}_1^2\mathbf{e} & i_0 = B, \xi_0 = 1 \\ 0 & \text{otherwise.} \end{cases}$$

4. Application to SS7 Network

There are three types of congestion controls in SS7 networks such as international control, national option with congestion priorities, and national option without congestion priorities [15]. Here we will describe the international control. when a message signal unit (MSU) is received at a Signaling Transfer Point (STP) for the congested link whose congestion status is 1, it is passed to Level 2 for transmission and a Transfer Controlled (TFC) packet is sent back to the originating Signaling Point (SP) which sent the MSU, for the initial packet and for every m_0 packet (default value of m_0 is 1 in this paper, but by a simple modification we can deal with the model with m_0 larger than 1).

We assume there are S identical SPs sending packets to a STP and we consider an output buffer of the STP and the packets sent to the output buffer (see Figure 2).

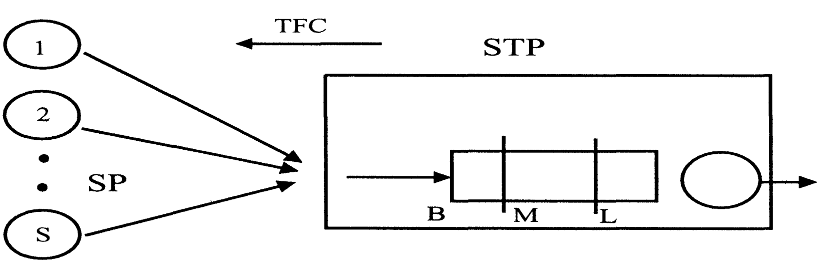


Figure 2. The model of the congestion control in SS7 networks

If an SP receives a TFC packet from the STP, the traffic load toward the STP is reduced by one step, and two timers T_{29} and T_{30} are started where the length of T_{30} is greater than that of T_{29}. Until T_{29} times out further TFC packets are ignored in order not to reduce traffic too rapidly. If a TFC is received after the expiry of T_{29} but before T_{30} expires, the traffic load is reduced by one more step and both T_{29} and T_{30} are restarted. This reduction continues until the last step when maximum reduction is obtained. If T_{30} expires, then the traffic load is increased by one step and T_{30} is restarted. This is repeated until the full load has been resumed. For simplicity, we assume the length of T_{29} to be equal to zero. The extension to the model with nonzero T_{29} is similar to the modeling in [4]. Even though the lengths of T_{30} is deterministic, we assume that the length of T_{30} has an exponential distribution with a mean, which is a deterministic value for the analytical modeling as in [9]. Let $\sigma = \frac{1}{E[T_{30}]}$.

Let K be the maximum reduction step of the traffic load in an SP. Define the state of an SP as k $(0 \leq k \leq K)$ if the SP has reduced its traffic load k times since the beginning with full load. Assume that each SP whose state is k sends packets according to a Poisson process with rate λ_k $(\lambda_0 \geq \lambda_1 \geq \ldots \leq \lambda_K)$. Let $Y_k(t)$ be the number of SPs in state k at time t. When $(Y_1(t), Y_2(t), \ldots, Y_K(t)) = (y_1, y_2, \ldots, y_K)$, $S - \sum_{k=1}^{K} y_k$ is the number of SPs in state 0 and the total arrival rate to the STP is $\lambda_0(S - \sum_{k=1}^{K} y_k) + \sum_{k=1}^{K} \lambda_k y_k$. Hence, $J(t) = (Y_1(t), Y_2(t), \ldots, Y_K(t))$ governs the arrival rate and so it can be defined as the underlying process of the arrival to the STP with the state space consisting of $(y_1, \ldots, y_K)$ listed in the lexicographic order, where $y_i \geq 0$ and $\sum_{k=1}^{K} y_k \leq S$. The total number m of the states equals $\frac{(K+S)!}{K!S!}$.

Let $\mathbf{A}((y_1, \ldots, y_K), (y'_1, \ldots, y'_K))$ denote the transition rates from the state $(y_1, y_2, \ldots, y_K)$ to the state $(y'_1, y'_2, \ldots, y'_K)$ which are the elements of an $m \times m$ matrix $\mathbf{A}$. Let $\mathbf{y} = (y_1, \ldots, y_K)$. For example, $\mathbf{A}(\mathbf{y}, \mathbf{y})$ denotes the diagonal elements of matrix $\mathbf{A}$. Let $\mathbf{e}_i$ be a vector whose elements are all zero except for the ith element which is 1, i.e., $\mathbf{e}_i = (0, \ldots, \overset{ith}{\widehat{1}}, \ldots, 0)$. Let $\mathbf{e}_0 = (0, \ldots, 0)$ and $\mathbf{e}_{K+1} = \mathbf{e}_K$, for the sake of convenience. We are ready to find the rate matrices $\mathbf{C}_0, \mathbf{D}_0, \mathbf{C}_1$ and $\mathbf{D}_1$ of the underlying process $\{J(t): t \geq 0\}$ for our model in this section. Independently of the congestion status of the buffer of the STP, if a T_{30} of an SP whose state is k expires, the state of the SP will be changed to $k-1$. Hence,

$$\mathbf{C}_n(\mathbf{y}, \mathbf{y} - \mathbf{e}_k + \mathbf{e}_{k-1}) = \sigma \cdot y_k \text{ for } n = 0, 1, \quad 1 \leq k \leq K.$$

When the congestion status of the buffer of the STP is 0, there is no transition of the state with an arrival of the underlying process $J(t)$, since there is no TFC generating from STP.

$$\mathbf{D}_0(\mathbf{y}, \mathbf{y}) = \lambda_0(S - \sum_{k=1}^{K} y_k) + \sum_{k=1}^{K} \lambda_k y_k.$$

When the congestion status of the buffer of the STP is 1, each SP who sends a packet to the STP will receive a TFC and reduce the traffic load by one step and restart T_{30}. Therefore, if one of y_k's SPs whose states are k sends a packet to the STP, its state will be changed into $k+1$.

$$\mathbf{D}_1(\mathbf{y}, \mathbf{y} - \mathbf{e}_k + \mathbf{e}_{k+1}) = \lambda_k y_k \text{ for } 0 \leq k \leq K.$$

The diagonal elements of the matrices $\mathbf{C}_0$ and $\mathbf{C}_1$ are negative values to make $\mathbf{C}_0 e + \mathbf{D}_0 e = 0$ and $\mathbf{C}_1 e + \mathbf{D}_1 e = 0$, respectively. The elements of the matrices $\mathbf{C}_0$, $\mathbf{D}_0$, $\mathbf{C}_1$ and $\mathbf{D}_1$ not mentioned above are all zeros.

For all numerical examples, we assume that $S = 10$, $K = 1$ and that the time scale is normalized by the mean service time of a packet, i.e., $\mu = 1.0$. Let the buffer capacity B be equal to 50. For Figures 3 through Figure 6, we assume $\lambda_0 = 0.08$ and $\lambda_1 = 0.04$.

Figure 3, Figure 4, and Figure 5 display the mean delay and the loss probability of packets in terms of functions of time, when $T_{30} = 100$. For Figure 3, we let L be fixed at 25 and the initial state by (0, 25, 0). As M decreases, the congestion control is triggered earlier and therefore the mean delay and the loss probability of packets decrease as shown in Figure 3.

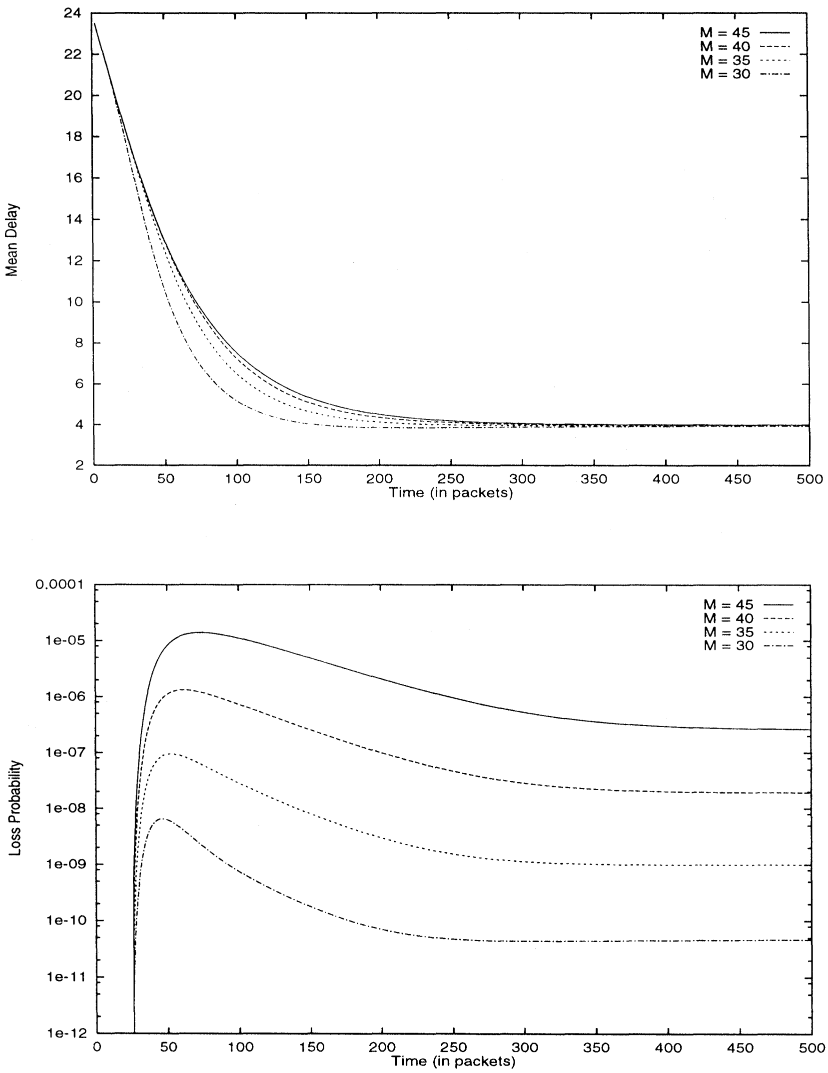


Figure 3. The mean delay and the loss probability of packets for the case of $T_{30} = 100, L = 25$ and that the initial state equals to (0, 25, 0)

In Figure 4, we consider an epoch when the queue length exceeds the onset threshold M as the initial epoch, i.e., $X_0 = (1, M+1, 1)$. Since the congestion control is triggered from the initial epoch, each SP receives a TFC packet when it sends a packet until the queue length crosses the abatement threshold L. SPs receiving a TFC packet reduce their traffic load and therefore the total offered load until the buffer decreases and the mean delay of packets decreases as shown in Figure 4. The loss probability of packets increases initially but it begins to decrease soon since the congestion control is triggered. After a time interval, the mean delay and the loss probability increase slightly as in Figure 4. This is because the total offered load increases again after the queue length crosses L.

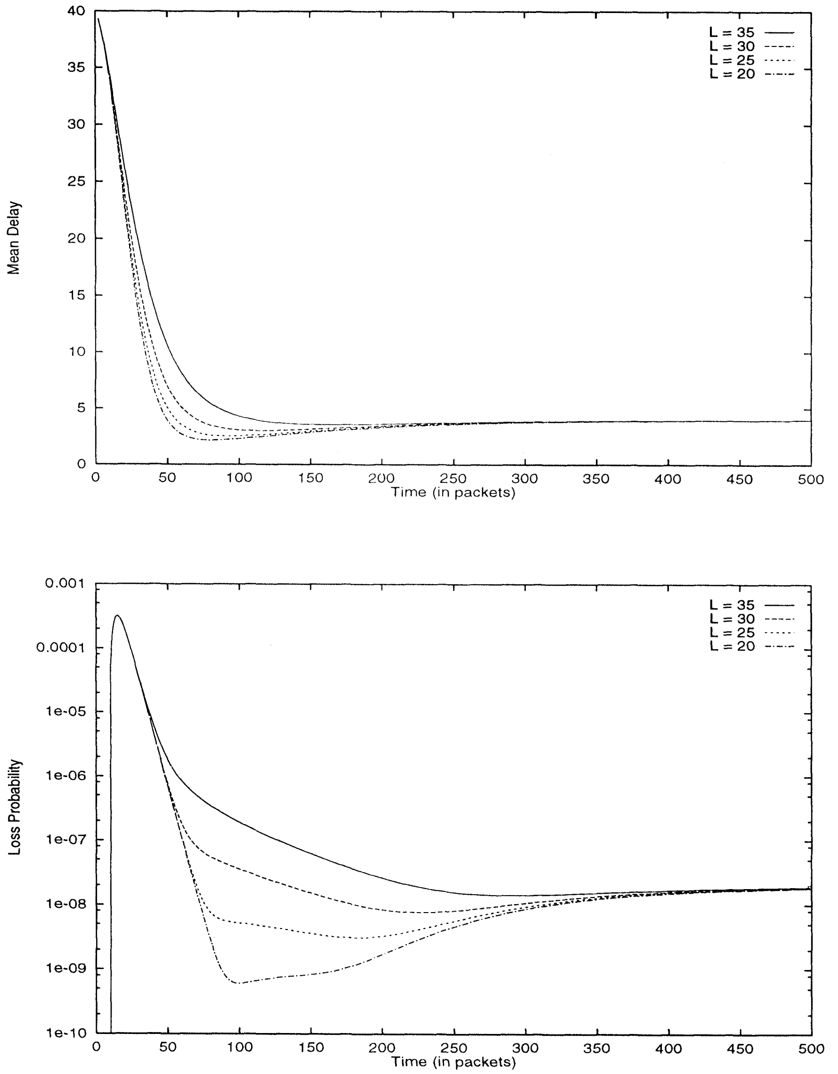


Figure 4. The mean delay and the loss probability of packets for the case of $T_{30} = 100$, $M = 40$ and that the initial state equals to $(1, M+1, 1)$

In Figure 5, we compare the loss probabilities with different M and different initial state X_0 for a fixed L ($L = 25$). Figure 5 shows that the loss probabilities converge to the same value for the same M independently of the initial value as the time increases.

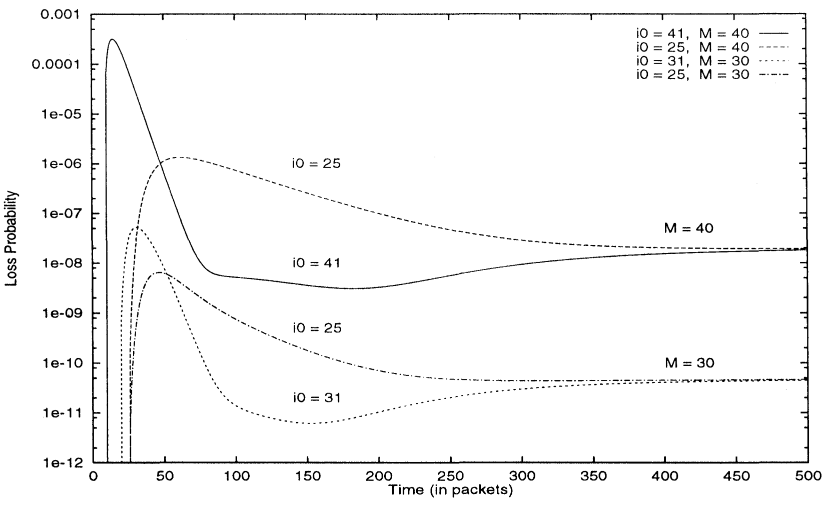
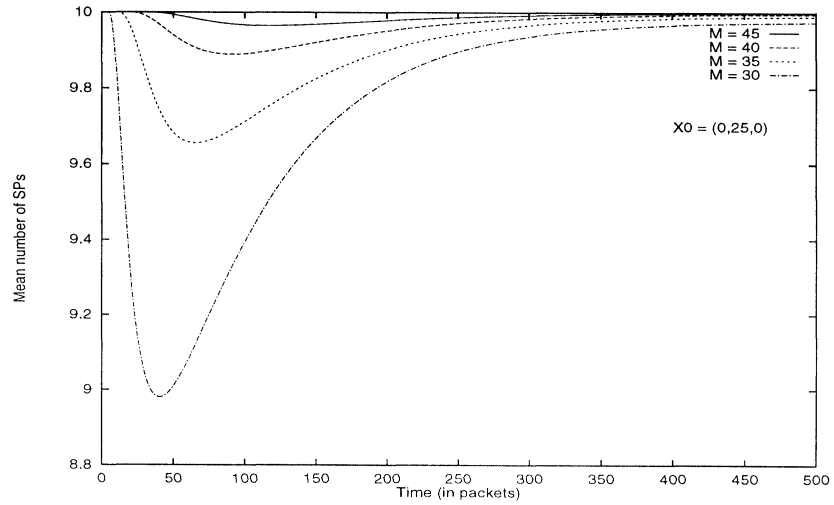


Figure 5. The mean delay and the loss probability of packets for the case of $T_{30} = 100$, $L = 25$

Define $F^n(i_0, j_0, \xi_0)$ as the mean number of SPs, which send packets in their full traffic load at time n (in packets). Then $F^n(i_0, j_0, \xi_0)$ is calculated by $\mathbf{P}^n(\xi_0, i_0; \xi, i)$ as

$$F^n(i_0, j_0, \xi_0) = S - \sum_{\xi} \sum_{i} e_{j_0} \mathbf{P}^n(\xi_0, i_0; \xi, i) e^*,$$

where $e^* = (0, 1, 2, \ldots, S)$. As in the case of $p^n_{\text{loss}}(i_0, j_0, \xi_0)$, $\mathbf{P}^n(\xi_0, i_0; \xi, i)e^*$ can be evaluated iteratively by substituting $\mathbf{P}^{n-1}(\xi_0, i_0; \xi, i)e$ and $\mathbf{P}^n(\xi_0, i_0; \xi, i)e$ for $\mathbf{P}^{n-1}(\xi_0, i_0; \xi, i)e^*$ and $\mathbf{P}^{n-1}(\xi_0, i_0; \xi, i)e^*$ in (14), respectively. Figure 6 displays $F^n(0, L, 0)$ and $F^n(1, M+1, 1)$ in terms of functions of time. For a fixed L, as M decreases and for a fixed M, as L decreases, the mean number of SPs with full traffic load decreases as in Figure 6. Hence there is trade-off between the loss probability (and the mean delay) and the throughput in terms of $F^n(i_0, j_0, \xi_0)$.



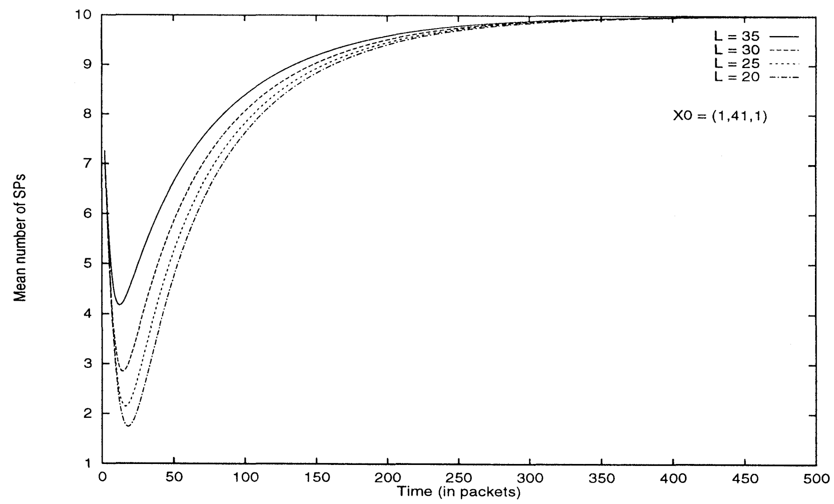


Figure 6. The mean number of SPs with their full traffic load for the case $T_{30} = 100$, $L = 25$, $M = 40$ and that the initial states equal to $(0, L, 0)$ and $(1, M+1, 1)$, respectively.

In Figures 7 and 8, we consider the control of SP with traffic reduction and the timer. We consider the case of $M = 40$ and $L = 25$. Figure 7 shows the impact of the length of the timer T_{30} on the transient performance. As T_{30} increases, the time of resuming full traffic load of each SP is delayed and therefore the loss probability and the mean delay of packets decrease as in Figure 7. We assume that each SP receiving TFC only send $d\%$ of its full packets. Then $\lambda_1 = \frac{d}{100}\lambda_0$. We consider two cases: $\lambda_0 = 0.08$ and $\lambda_0 = 0.12$. As d decreases, the total offered load decreases and therefore the mean delay decreases as shown in Figure 8.

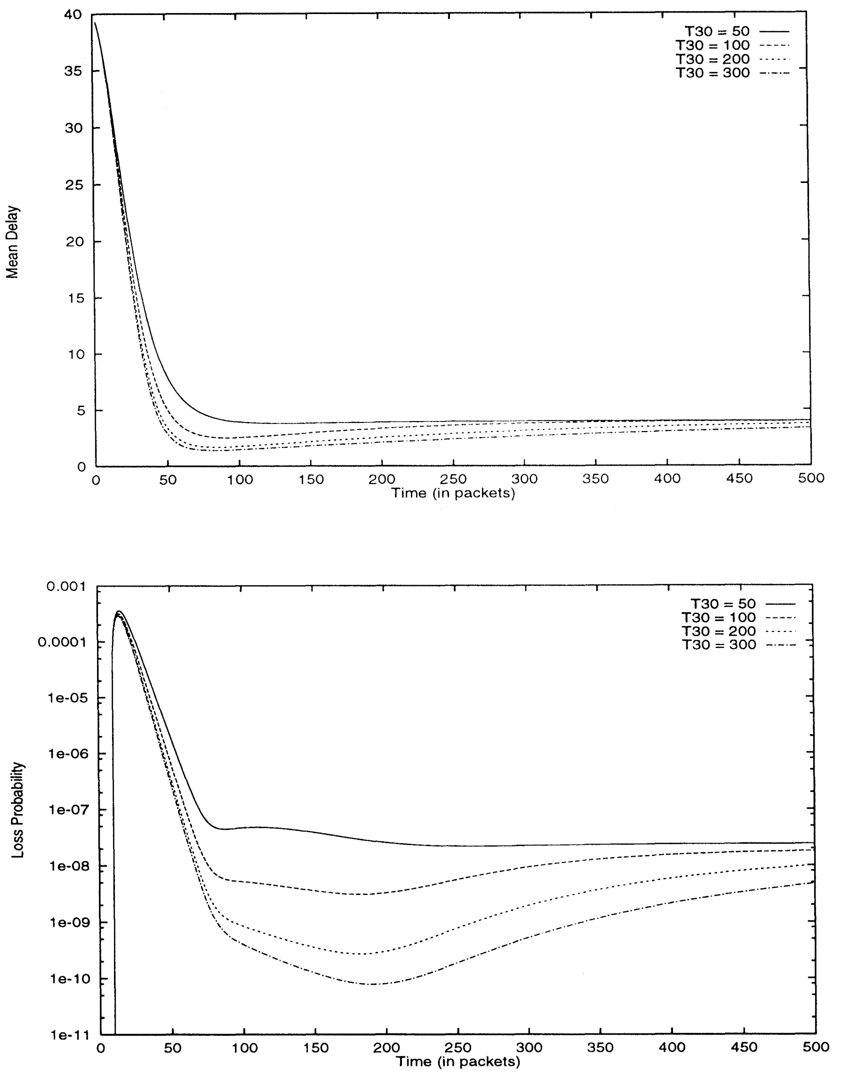
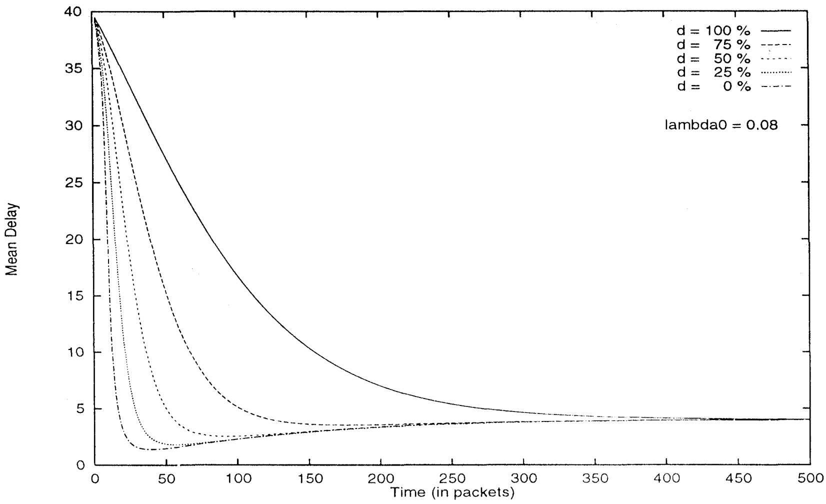


Figure 7. The mean delay and the loss probability of packets for the case of $L = 25$, $M = 40$ and that the initial state equal to $(1, M+1, 1)$



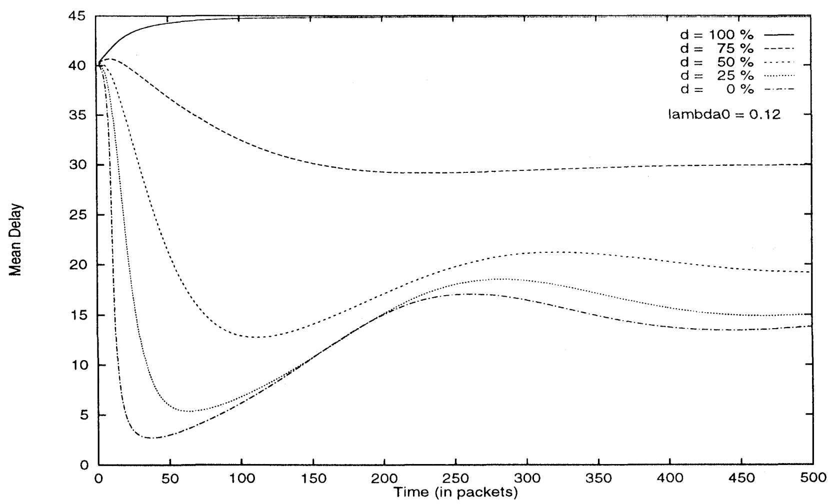


Figure 8. The mean delay of packets for the case of $L = 25$, $M = 40$ and that the initial state equal to $(1, M+1, 1)$ with $\lambda_0 = 0.08$ and $\lambda_0 = 0.12$, respectively

References

[1] Abate, J. and Whitt, W., Transient behavior of the M/M/1 queue via Laplace transforms, *Adv. Appl. Prob.* **20** (1988), 145-178.

[2] Choi, B.D., Hwang, G.U. and Han, D.H., Supplementary variable method applied to the MAP/G/1 queueing system, *J. of the Aust. Math. Soc. Series B-Applied Math* **39** (1998), 86-96.

[3] Choi, B.D. and Hwang, G.U., The MAP, M/G_1,G_2/1 queue with preemptive

priority, *J. of Appl. Math. and Stoch. Anal.* **10**:4 (1997), 407-421.

[4] Choi, B.D., Choi, S.H., Kim, B. and Sung, D.K., Analysis of priority queueing system based on thresholds and its application to signaling system No. 7 with congestion control, *Computer Networks and ISDN*, to appear.

[5] Choi, B.D. and Lee, Y., Queueing system with multiple delay and loss priorities for ATM networks, submitted.

[6] Choi, B.D. and Hwang, G.U., A transient analysis for the unfinished work and waiting time in the MAP/G/1 queue with vacations and state-dependent service times, submitted.

[7] Collier, B.R. and Kim, H.S., Traffic rate and shape control with queue threshold congestion recognition, *IEEE ICC'96* **2** (1996), 746-750.

[8] Oliver, C.I. and Keilson, J., Overload control in finite-buffer multiclass message systems, *Telecomm. Sys.* **2** (1994), 121-140.

[9] Lee, K.J. and Lim, Y., Performance analysis of the congestion control scheme in the signaling system No. 7, *Proc. INFOCOM '89*, Ottawa, **Apr.** (1989), 691-700.

[10] Li, S.Q., Overload control in a finite message storage buffer, *IEEE Trans. on Commun.* **37**:12, (1989), 1330-1338.

[11] Lee, D.S. and Li, S.Q., Transient analysis of multi-server queues with Markov-modulated Poisson arrivals and overload control, *Perf. Eval.* **16** (1992), 49-66.

[12] Lee, D.S. and Li, S.Q., Transient analysis of a switched Poisson arrival queue under overload control, *Perf. Eval.* **17** (1993), 13-29.

[13] Lucantoni, D.M., Meier-Hellstern, K.S. and Neuts, M.F., A single-server queue with server vacations and a class of non-renewal arrival processes, *Adv. Appl. Prob.* **22** (1990), 676-705.

[14] Lucantoni, D.M., Choudhury, G.L. and Whitt, W., The transient BMAP/G/1 queue, *Stoch. Models* **10**:1 (1994), 145-182.

[15] Manfield, D.R., Millsteed, G. and Zukerman, M., Congestion controls in SS7 signaling networks, *IEEE Commun. Mag.* **June** (1993), 50-57.

[16] Sharma, O.P. and Shobha, B., Transient behavior of a double-channel Markovian queue with limiting waiting space, *Queueing Systems* **3** (1988), 89-96.

5. Appendix

5.1 The Proof of Lemma 1

For $i_0 \geq 1$, define

$$P_{j_0,j}(s) \triangleq Pr\{S > s, X(s) = (i_0, j) \mid X(0) = (i_0, j_0)\}.$$

Then we have the following Chapman-Kolmogorov equations:

$$P_{j_0,j}(s + \Delta s) = Pr\{S > s + \Delta s, X(s + \Delta s) = (i_0, j) \mid X(0) = (i_0, j_0)\}$$

$$= \sum_{k=1}^{m} Pr\{S > s, X(s) = (i_0, k) \mid X(0) = (i_0, j_0)\}$$

$$\cdot Pr\{S > s+\Delta s, X(s+\Delta s) = (i_0, j) \mid X(0) = (i_0, j_0), S > s, X(s) = (i_0, k)\}$$

$$= \sum_{k=1}^{m} P_{j_0,k}(s) Pr\{S > s+\Delta s, X(s+\Delta s) = (i_0, j) \mid S > s, X(s) = (i_0, k)\},$$

where the last equality follows from the Markov property. By subtracting $P_{j_0,j}(s)$ from both sides of the above equation and dividing both sides by the equation Δs, we get

$$\frac{P_{j_0,j}(s+\Delta s) - P_{j_0,j}(s)}{\Delta s}$$

$$= \sum_{k \neq j} P_{j_0,k}(s) \frac{Pr\{S > s+\Delta s, X(s+\Delta s) = (i_0, j) \mid S > s, X(s) = (i_0, k)\}}{\Delta s}$$

$$+ P_{j_0,j}(s) \frac{Pr\{S > s+\Delta s, X(s+\Delta s) = (i_0, j) \mid S > s, X(s) = (i_0, j)\} - 1}{\Delta s}.$$

By passing to the limit $\Delta s \to 0$ in both sides of the above equation and using the definition of the matrix $\mathbf{C} - \mu\mathbf{I}$ we can obtain

$$\frac{d}{ds} P_{j_0,j}(s) = \sum_{k=1}^{m} P_{j_0,k}(s)(\mathbf{C} - \mu\mathbf{I})_{k,j}.$$

The above equation can be rewritten in matrix form as

$$\frac{d}{ds}\mathbf{P}(s) = \mathbf{P}(s)(\mathbf{C} - \mu\mathbf{I}),$$

where $\mathbf{P}(s) = [P_{j_0,j}(s)]$. By integrating both sides of the above equation and using $\mathbf{P}(0) = \mathbf{I}$ and $\mathbf{P}(\infty) = \mathbf{0}$, we can obtain $\mathbf{P}(\infty) - \mathbf{P}(0) = -\mathbf{I} = \int_0^\infty \mathbf{P}(s)ds(\mathbf{C} - \mu\mathbf{I})$, i.e.,

$$\int_0^\infty \mathbf{P}(s)ds = (\mu\mathbf{I} - \mathbf{C})^{-1}. \tag{15}$$

For $i_0 = 0$, we can obtain the result similarly to the above.

5.2 The Proof of Lemma 2

For $i_0 \geq 1$, define

$$R_{j_0,j}(t) \triangleq Pr\{S \leq t, X(S) = (i_0 - 1, j) \mid X(0) = (i_0, j_0)\}.$$

Note that $R_{j_0,j} = \lim_{t\to\infty} R_{j_0,j}(t)$. From the definition of $R_{j_0,j}(t)$, we have the following Chapman-Kolmogorov equation:

$$R_{j_0, j}(t+\Delta t) - R_{j_0, j}(t)$$

$$= \sum_{k=1}^{m} Pr\{S > t, X(t) = (i_0, k) \mid X(0) = (i_0, j_0)\}$$

$$\cdot Pr\{t < S \leq t + \Delta t, X(S) = (i_0 - 1, j) \mid S > t, X(t) = (i_0, k), X(0) = (i_0, j_0)\}$$

$$= \sum_{k=1}^{m} Pr\{S > t, X(t) = (i_0, k) \mid X(0) = (i_0, j_0)\}$$

$$\cdot Pr\{t < S \leq t + \Delta t, X(S) = (i_0 - 1, j) \mid S > t, X(t) = (i_0, k)\},$$

where the last equality follows from the Markov property of $\{X(t): t \geq 0\}$. Dividing both sides of the above equation by Δt and taking the limit $\Delta t \rightarrow 0$, we obtain

$$\frac{d}{dt} R_{j_0, j}(t) = \sum_{k=1}^{m} Pr\{S > t, X(t) = (i_0, k) \mid X(0) = (i_0, j_0)\}(\mu \mathbf{I})_{k, j}.$$

The above equation can be rewritten in matrix form as

$$\frac{d}{dt}\mathbf{R}(t) = \mathbf{P}(t)\mu\mathbf{I},$$

where $\mathbf{R}(t) = [R_{j_0, j}(t)]$. Since $\mathbf{R}(0) = 0$, by integrating both sides of the above equation, we can obtain

$$\mathbf{R}(t) = \int_0^t \mathbf{P}(s)ds\mu\mathbf{I}. \tag{16}$$

Hence from (16) and (15), we have

$$\mathbf{R} = \lim_{t \rightarrow \infty} \mathbf{R}(t) = \int_0^\infty \mathbf{P}(s)ds\mu\mathbf{I} = (\mu\mathbf{I} - \mathbf{C})^{-1}\mu\mathbf{I}.$$

Similarly, we can get $\mathbf{G} = (\mu\mathbf{I} - \mathbf{C})^{-1}\mathbf{D}$, $\mathbf{H} - (-\mathbf{C})^{-1}\mathbf{D}$, whose proof is omitted.