Journal of the Korean Institute of Industrial Engineers
Vol. 25, No. 1, pp. 130-141, March 1999.

A Queueing Model with Loss and Time Priority for Optimal Buffer Control in ATM

H. W. Lee · E. Y. Kim
Sungkyunkuan University School of Systems Managemt Engineering

손실 우선과 시간 우선이 공존하는 ATM에서의 최적 버퍼 제어를 위한 대기 행렬 모형

이호우 · 김은영

This paper deals with a priority queueing model in an ATM system. Two types of customers are considered. Type-1 customers have push-put priority over type-2 customers. Type-1 customers can enter the service only when the number of type-2 customers is less than a threshold T.

We derive the joint probability of the number of customers in the buffer, the mean waiting time, and the loss probabilities of each type. We also propose an optimal control policy that satisfies a given quality of service.

1. Introduction

The asyncronous transfer mode(ATM) is considered as the basis for the future B-ISDN which integrates different types of information services such as voice, data and video communication. Each type of information services has its own quality of service(QoS) requirements, taking cell loss probability and cell transfer delay into consideration. Numerous studies on buffer control strategies have been presented for the purpose of effective management of the QoSs.

For finite buffer systems, a space control strategy is composed of service discipline and the buffer access control discipline. The former is concerned with the rule of customer selection for next service, while the latter deals with the rule of customer acceptance.

In general, "loss priority" and "time priority" rules are applied to the disciplines. Time priority scheme is used to reduce the waiting time of high priority customers. The usual HOL(Head of Line) mechanisms have suggested differnt delay characteristics for different types of customers [3], [10]. Loss priority scheme is used to reduce the loss probability of high priority customers usually at the expense of low priority customers. Two loss priority mechanisms have been proposed and studied by Kröner [6], Kröner et al. [7] and Rothermel [9]: "push-out" and "partial buffer sharing". In the push-out mechanism a high priority customer arriving to the full buffer pushes out one of the low priority customers. In the partial buffer sharing mechanism, an arriving low-priority customer is denied of admission if the buffer occupancy reaches a given threshold, whereas high-priority customers are accepted as long as there is a vacancy. Ahn and Lee[1][2] analyzed the partial buffer sharing scheme with threshold and proposed an algorithm to obtain the optimal buffer size for each class of customers.

Application of only one of those priority schemes may cost efficiency and effectiveness that might have been avoided if the other scheme is applied. Recently, some systems in which both time and loss priority mechanisms were implemented were studied by Gravey & Hebuterne [5] and Neuts [8].

In this paper, we build an $M_1, M_2/M/1/B+1$ queueing model to analyze the system with both loss and time priorities. Type-1 customers have loss-priority over type-2 customers while type-2 customers have time-priority over type-1 customers. We derive the joint distribution of the number of customers in the buffer, the mean queue waiting times and the loss probabilities. We also derive the optimal buffer control policy that satisfies the given QoSs.

2. The system and the model

The queueing system studied in this paper is characterized as follows :

1) There are two types of customers: type-1 and type-2. Type-i $(i=1,2)$ customers arrive according to a Poisson process with rate λ_i.
2) The system has the buffer of size B (thus there can be $B+1$ customers in the system including the one in service).
3) On all occasions, non-preemption is assumed.
4) Type-2 customers are allowed to occupy the buffer up to the level ($D \le B$).
5) If the number of type-1 customers in the buffer is less than the threshold $T(\le B-D)$ at the end of a service, a type-2 customer is served. Otherwise, a type-1 customer is served. If only one type of customers exist at the end of a service, one of them is taken into service.
6) A type-1 customer who arrives when the buffer is full pushes out the oldest type-2 customer (if any).

Since customers arrive according to Poisson processes, the state probabilities observed by arriving customers are identical to the time-average probabilities (PASTA (Wolff [10])). Since service distributions are exponential, the system is Markovian. Let us define the following notations and probabilities :

λ_1 : arrival rate of type-1 customer

λ_2 : arrival rate of type-2 customer

λ : total arrival rate ($\lambda = \lambda_1 + \lambda_2$)

μ : service rate

$N_1(t)$: queue size (number of customers in the buffer excluding the one in service) of type-1 customers at time t

$N_2(t)$: queue size of type-2 customers at time t

B : buffer size

D : maximum allowed occupancy of type-2 customers ($D \le B$)

T : threshold for the service priority of type-1 customers, ($T \le B-D$)

$$Y(t) = \begin{cases} 1, & \text{if the server is busy at time} \\ 0, & \text{if the server is idle at time} \end{cases}$$

$$P_{ij}(t) = \Pr[N_1(t)=i,\ N_2(t)=j,\ Y(t)=1]$$

$$Q_{0,0}(t) = \Pr[Y(t)=0]$$

$$P_{i,j} = \lim_{t\to\infty} P_{i,j}(t), \quad 0 \le j \le D, \quad 0 \le i \le B-j$$

$$Q_{0,0} = \lim_{t\to\infty} Q_{0,0}$$

State (i, j) : in steady-state, there are i type-1 customers and j type-2 customers in the buffer (excluding the one in service).

The rate flow diagram is seen in <Figure 1>.

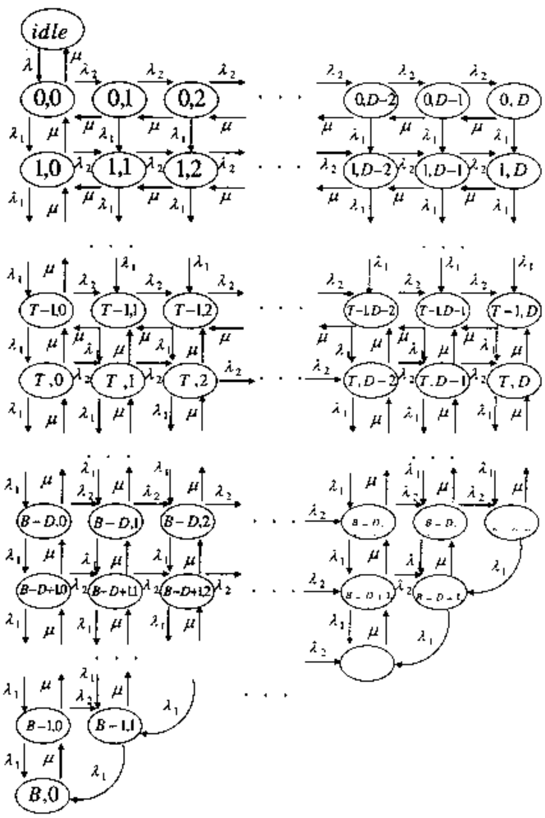

Figure 1. The rate flow diagram.

2.1 The system equations

From the rate diagram we can set up the following steady-state system equations:

$$(\lambda_1+\lambda_2)Q_{0,0} = \mu P_{0,0} \tag{2.1}$$

$$(\mu+\lambda_1+\lambda_2)P_{0,0}=(\lambda_1+\lambda_2)Q_{0,0}+\mu P_{1,0}+\mu P_{0,1} \tag{2.2}$$

$$(\mu+\lambda_1+\lambda_2)P_{i,0}=\lambda_1 P_{i-1,0}+\mu P_{i+1,0}+\mu P_{i,1} \quad (1\le i\le T-1) \tag{2.3}$$

$$(\mu+\lambda_1+\lambda_2)P_{i,0}=\lambda_1 P_{i-1,0}+\mu P_{i+1,0} \quad (T\le i\le B-1) \tag{2.4}$$

$$(\mu+\lambda_1+\lambda_2)P_{0,j}=\lambda_2 P_{0,j-1}+\mu P_{0,i+1} \quad (1\le j\le D-1) \tag{2.5}$$

$$(\mu+\lambda_1+\lambda_2)P_{i,j}=\lambda_1 P_{i-1,j}+\lambda_2 P_{i,j-1}+\mu P_{i,j+1} \quad (1\le i\le T-2,\ 1\le j\le D-1) \tag{2.6}$$

$$(\mu+\lambda_1+\lambda_2)P_{i,j}=\lambda_1 P_{i-1,j}+\lambda_2 P_{i,j-1}+\mu P_{i+1,j} \quad (i=T-1,\ 1\le j\le D-1) \tag{2.7}$$

$$(\mu+\lambda_1+\lambda_2)P_{i,j}=\lambda_1 P_{i-1,j}+\lambda_2 P_{i,j-1}+\mu P_{i,j+1}+\mu P_{i+1,j} \quad (T\le i\le B-j-1,\ 1\le j\le D-1) \tag{2.8}$$

$$(\lambda_1+\mu)P_{0,D}=\lambda_2 P_{0,D-1} \tag{2.9}$$

$$(\lambda_1+\mu)P_{i,D}=\lambda_2 P_{i,D-1}+\lambda_1 P_{i-1,D} \quad (1\le i\le T-2) \tag{2.10}$$

$$(\lambda_1+\mu)P_{i,D}=\lambda_2 P_{i,D-1}+\lambda_1 P_{i-1,D}+\mu P_{i+1,D} \quad (i=T-1) \tag{2.11}$$

$$(\lambda_1+\mu)P_{i,D}=\lambda_2 P_{i,D-1}+\lambda_1 P_{i-1,D}+\mu P_{i+1,D} \quad (T\le i\le B-D-1) \tag{2.12}$$

$$\mu P_{B,0}=\lambda_1 P_{B-1,0}+\lambda_1 P_{B-1,1} \tag{2.13}$$

$$(\lambda_1+\mu)P_{B-D,D}=\lambda_1 P_{B-D-1,D}+\lambda_2 P_{B-D,D-1} \tag{2.14}$$

$$(\lambda_1+\mu)P_{i,j}=\lambda_1 P_{i-1,j}+\lambda_2 P_{i,j-1}+\lambda_1 P_{i-1,j+1} \quad (i+j=B,\ 1\le j\le D-1) \tag{2.15}$$

2.2 Recursive solution

Recursive solution seems to be the only way to solve the above system of equations due to their complexity. We first express $Q_{0,0}$ and $P_{i,j}$ in terms of $P_{0,0}$ and $P_{B,0}$. We then express $P_{0,B}$ in terms of $P_{0,0}$. Finally using the normalization condition, we obtain the state probabilities. To this end, we take the following steps.

(Step 1)

From (2.1),

$$Q_{0,0}=\frac{\mu}{\lambda}P_{0,0}. \tag{2.16}$$

(Step 2) ($i=0$, $0\le j\le D$)

From (2.9), we have

$$P_{0,D}=\lambda_2 A_0 P_{0,D-1}, \tag{2.17}$$

where $A_0=\dfrac{1}{\lambda_1+\mu}$.

Using (2.17) in (2.5), we get

$$P_{0,D-1}=\lambda_2 A_1 P_{0,D-2}, \tag{2.18}$$

where $A_1=\dfrac{1}{(\lambda+\mu)-\mu\lambda_2 A_0}$.

In the similar way, we have

$$P_{0,j}=\lambda_2 A_{D-j}P_{0,j-1},\ (1\le j\le D-2) \tag{2.19}$$

where $A_{D-j}=\dfrac{1}{(\lambda+\mu)-\mu\lambda_2 A_{D-j-1}}$.

(Step 3) ($i=1$, $j=0$)

If we sum up (2.1), (2.2), (2.5) and (2.9), we have

$$\lambda_1\Big(\sum_{j=0}^{D}P_{0,j}\Big)=\mu P_{1,0}$$

Thus we have

$$P_{1,0}=\frac{\lambda_1}{\mu}\Big(\sum_{j=0}^{D}P_{0,j}\Big), \tag{2.20}$$

where $\{P_{0,j},\ j=0,1,\ldots,D\}$ were obtained in (Step 2).

(Step 4) ($i=1$, $1\le j\le D$)

From (2.10), we have

$$P_{1,D}=\lambda_2 A_0 P_{1,D-1}+\lambda_1 A_0 P_{0,D} \tag{2.21}$$

Using (2.21) in (2.6),

$$P_{1,D-1}=\lambda_2 A_1 P_{1,D-2}+\lambda_1 A_1 P_{0,D-1}$$

$$+\mu\lambda_1 A_0 A_1 P_{0,D} \qquad (2.22)$$

In the similar way, we have

$$P_{1,j} = \lambda_2 A_{D-j} P_{1,j-1} + \lambda_1 \sum_{n=0}^{D-j}\left[P_{0,D-n}\left(\prod_{m=n}^{D-j} A_m\right)\mu^{D-j-n}\right] \quad (1 \le j \le D-2) \qquad (2.23)$$

where $A_0 = \frac{1}{\lambda_1+\mu}$, and $A_m = \frac{1}{(\lambda+\mu)-\mu\lambda_2 A_{m-1}}$.

(Step 5) ($2 \le i \le T-2,\ 0 \le j \le D$)

In the similar way as in (step 4),

$$P_{i,j} = \lambda_2 A_{D-j} P_{i,j-1} + \lambda_1 \sum_{n=0}^{D-j}\left[P_{i-1,D-n}\left(\prod_{m=n}^{D-j} A_m\right)\mu^{D-j-n}\right] \qquad (2.24)$$

$$(2 \le i \le T-2,\ 1 \le j \le D)$$

From (2.3), (2.6) and (2.11), we have

$$P_{i,0} = \frac{\lambda_1}{\mu}\sum_{j=0}^{D} P_{i-1,j},\ (2 \le i \le T-2) \qquad (2.25)$$

(Step 6) ($T-1 \le i \le B-1,\ j=0$)

From (2.4), we get

$$P_{B-1,0} = \lambda_1 B_0 P_{B-2,0} + \mu B_0 P_{B,0} \qquad (2.26)$$

where $B_0 = \frac{1}{\lambda+\mu}$.

In the similar way, we have, from (2.4),

$$P_{i,0} = \lambda_1 B_{B-i-1} P_{i-1,0} + \mu^{B-i}\left(\prod_{m=0}^{B-i-1} B_m\right) P_{B,0}$$

$$(T \le i \le B-2) \qquad (2.27)$$

where $B_m = \frac{1}{(\lambda+\mu)-\mu\lambda_1 B_{m-1}}$.

From (2.3), (2.6) and (2.11), we have

$$P_{T-1,0} = \frac{\lambda_1}{\mu}\sum_{j=0}^{D} P_{T-2,j} \qquad (2.28)$$

By careful examination, we see that $P_{T-1,0}$ can be expressed in terms of $P_{0,0}$ alone. Let $X_{T-1,0}$ be the coefficient of $P_{0,0}$ such that

$$P_{T-1,0} = X_{T-1,0} P_{0,0} \qquad (2.29a)$$

$X_{T-1,0}$ can be easily calculated from (2.24). Now, let $X_{i,0}$ and $Z_{i,0}$ be the coefficients of $P_{0,0}$ and $P_{B,0}$ respectively when $P_{i,0}$ is expressed in terms of $P_{0,0}$ and $P_{B,0}$ such that

$$P_{i,0} = X_{i,0} P_{0,0} + Z_{i,0} P_{B,0} \qquad (2.29b)$$

Then from (2.4), we get, for $T \le i \le B-1$,

$$P_{i,0} = \lambda_1 B_{B-i-1} X_{i-1,0} P_{0,0} + \left[\lambda_1 B_{B-i-1} Z_{i-1,0} + \mu^{B-i}\prod_{m=0}^{B-i-1} B_m\right] P_{B,0}$$

Thus we have recursions

$$X_{i,0} = \lambda_1 B_{B-i-1} X_{i-1,0} \qquad (2.30a)$$

$$Z_{i,0} = \lambda_1 B_{B-i-1} Z_{i-1,0} + \mu^{B-i}\left(\prod_{m=0}^{B-i-1} B_m\right) \qquad (2.30b)$$

Thus starting from $X_{T-1,0}$ and $Z_{T,0}$ (note $Z_{T-1,0}=0$ from (2.29a)), we can obtain the coefficients $X_{i,0}$ and $Z_{i,0}$ recursively. In the sequel, we will use $X_{i,j}$ and $Z_{i,j}$ as the coefficients of $P_{0,0}$ and $P_{B,0}$ respectively such that

$$P_{i,j} = X_{i,j} P_{0,0} + Z_{i,j} P_{B,0} \qquad (2.31)$$

(Step 7) ($i = B-1,\ j=1$)

From (2.13), we have

$$\begin{aligned} P_{B-1,1} &= \frac{\mu}{\lambda_1} P_{B,0} - P_{B-1,0} \\ &= \frac{\mu}{\lambda_1} P_{B,0} - [X_{B-1,0} P_{0,0} + Z_{B-1,0} P_{B,0}] \\ &= -X_{B-1,0} P_{0,0} + \left[\frac{\mu}{\lambda_1} - Z_{B-1,0}\right] P_{B,0} \end{aligned} \qquad (2.32a)$$

Thus we see that

$$X_{B-1,1} = -X_{B-1,0},$$

$$Z_{B-1,1} = \frac{\mu}{\lambda_1} - Z_{B-1,0} \qquad (2.32b)$$

(Step 8) ($i = T-1,\ j=1$)

From (2.3), we obtain

$$\begin{aligned} P_{T-1,1} &= \frac{1}{\mu}[(\lambda+\mu) P_{T-1,0} - \lambda_1 P_{T-2,0} - \mu P_{T,0}] \qquad (2.33) \\ &= \left[\frac{(\lambda+\mu)}{\mu} X_{T-1,0} P_{0,0} - \frac{\lambda_1}{\mu} P_{T-2,0} - X_{T,0} P_{0,0}\right] \\ &\quad - Z_{T,0} P_{B,0} \end{aligned}$$

where $P_{T-2,0}$ was obtained in (step 5).

(Step 9) ($T \le i \le B-2,\ j=1$)
Similarly from (2.8), we get

$$\begin{aligned} P_{i,1} &= \lambda_1 B_{B-i-2} P_{i-1,1} + \lambda_2 \sum_{n=0}^{B-i-2}\left[P_{B-2-n,0}\left(\prod_{m=n}^{B-i-2} B_m\right)\mu^{(B-i-2)-n}\right] + \mu^{B-i-1}\left(\prod_{m=0}^{B-i-2} B_m\right) P_{B-1,1} \\ &= \left[\lambda_1 B_{B-i-2} X_{i-1,1} + \lambda_2 \sum_{n=0}^{B-i-2} X_{B-2-n,0}\left(\prod_{m=n}^{B-i-2} B_m\right)\mu^{B-i-2-n} + \mu^{B-i-1}\right. \\ &\quad \left.\left(\prod_{m=0}^{B-i-2} B_m\right) X_{B-1,1}\right] P_{0,0} + [\lambda_1 B_{B-i-2} Z_{T-1,1} + \\ &\quad \lambda_2 \sum_{n=0}^{B-i-2} Z_{B-2-n,0}\left(\prod_{m=n}^{B-T-2} B_m\right)\mu^{B-T-2-n} + \mu^{B-T-1}\left(\prod_{m=0}^{B-T-2} B_m\right) Z_{B-1,1}\Big] P_{B,0} \end{aligned} \tag{2.34}$$

where $B_m = \dfrac{1}{(\lambda+\mu) - \mu\lambda_1 B_{m-1}}$ and $B_0 = \dfrac{1}{\lambda+\mu}$.

(Step 10) ($i = T-1,\ 2 \le j \le D-1$)
From (2.8), we get

$$\begin{aligned} P_{T-1,j} &= \frac{1}{\mu}\left[(\lambda+\mu)P_{T-1,j-1} - \lambda_1 P_{T-2,j-1} - \lambda_2 P_{T-1,j-2} - \mu P_{T,j-1}\right] \\ &= \left[\frac{(\lambda+\mu)}{\mu} X_{T-1,j-1} - \frac{\lambda_2}{\mu} X_{T-1,j-2} - X_{T,j-1}\right] P_{0,0} - \frac{\lambda_1}{\mu} P_{T-2,j-1} \\ &\quad - \left[\frac{(\lambda+\mu)}{\mu} Z_{T-1,j-1} - \frac{\lambda_2}{\mu} Z_{T-1,j-2} - Z_{T,j-1}\right] P_{B,0} \end{aligned} \tag{2.35}$$

where $P_{T-2,j-1}$ was obtained in (step 5).

(Step 11) ($T \le i \le B-j,\ 2 \le j \le D-1$)
From (2.15), we get

$$\begin{aligned} P_{B-j,j} &= \frac{(\lambda_1+\mu)}{\lambda_1} P_{B-j+1,j-1} - P_{B-j,j-1} \\ &= -\frac{\lambda_2}{\lambda_1} P_{B-j+1,j-2}\left[\frac{(\lambda_1+\mu)}{\lambda_1} X_{B-j+1,j-1} - - X_{B-j,j-1}\frac{\lambda_2}{\lambda_1} X_{B-j+1,j-2}\right] P_{0,0} \\ &\quad - + \left[\frac{(\lambda_1+\mu)}{\lambda_1} Z_{B-j+1,j-1} Z_{B-j,j-1} - - \frac{\lambda_2}{\lambda_1} Z_{B-j+1,j-2}\right] P_{B,0} \end{aligned}$$

For $T \le i \le B-j-1$, we have

$$\begin{aligned} P_{ij} &= \lambda_1 B_{B-i-j-1} P_{i-1,j} + \lambda_2 \sum_{n=0}^{B-i-j-1}\left[P_{B-j-1-n,j-1}\left(\prod_{m=n}^{B-i-j-1} B_m\right)\mu^{(B-i-j-1)-n}\right] + \mu^{B-i-j}\left(\prod_{m=0}^{B-i-j-1} B_m\right) P_{B-j,j} \\ &= \left[\lambda_1 B_{B-i-j-1} X_{i-1,j} + \lambda_2 \sum_{n=0}^{B-i-j-1} X_{B-j-n-1,j-1}\left(\prod_{m=n}^{B-i-j-1} B_m\right)\mu^{(B-i-j-1)-n} + \mu^{B-i-j}\left(\prod_{m=0}^{B-i-j-1} B_m\right) X_{B-j,j}\right] \\ &\quad P_{0,0} + \left[\lambda_1 B_{B-i-j-1} Z_{i-1,j} + \lambda_2 \sum_{n=0}^{B-i-j-1} Z_{B-j-n-1,j-1}\left(\prod_{m=n}^{B-T-j-1} B_m\right)\mu^{(B-i-j-1)-n} + \right. \\ &\quad \left.\mu^{B-i-j}\left(\prod_{m=0}^{B-i-j-1} B_m\right) Z_{B-j,j}\right] P_{B,0} \end{aligned} \tag{2.37}$$

(Step 12) ($i=T-1,\ j=D$)

From (2.7), we have

$$\begin{aligned} P_{T-1,D} &= \frac{1}{\mu}[(\lambda+\mu)P_{T-1,D-1}-\lambda_1 P_{T-2,D-1}-\lambda_2 P_{T-1,D-2}-\mu P_{T,D-1}] \\ &= \Big[\frac{(\lambda+\mu)}{\mu}X_{T-1,D-1}-\frac{\lambda_2}{\mu}X_{T-1,D-2}-X_{T,D-1}\Big]P_{0,0}-\frac{\lambda_1}{\mu}P_{T-2,D-1} \\ &\quad -\Big[\frac{(\lambda+\mu)}{\mu}Z_{T-1,D-1}-\frac{\lambda_2}{\mu}Z_{T-1,D-2}-Z_{T,D-1}\Big]P_{B,0} \end{aligned} \tag{2.38}$$

where $P_{T-2,D-1}$ were obtained in (Step 5).

(Step 13) ($T\le i\le B-D,\ j=D$)

From (2.12), we have

$$\begin{aligned} P_{i,D} &= \lambda_1 C_{B-D-i}P_{i-1,D}+\lambda_2\sum_{n=0}^{B-D-i}\Big[P_{B-D-n,D-1}\Big(\prod_{m=n}^{B-D-i}C_m\Big)\mu^{B-D-i-n}\Big] \\ &= \Big\{\lambda_1 C_{B-D-i}X_{i-1,D}+\lambda_2\sum_{n=0}^{B-D-i}\Big[X_{B-D-n,D-1}\Big(\prod_{m=n}^{B-D-i}C_m\Big)\mu^{B-D-i-n}\Big]\Big\}P_{0,0} \\ &\quad +\Big\{\lambda_1 C_{B-D-i}Z_{i-1,D}+\lambda_2\sum_{n=0}^{B-D-i}\Big[Z_{B-D-n,D-1}\Big(\prod_{m=n}^{B-D-i}C_m\Big)\mu^{B-D-i-n}\Big]\Big\}P_{B,0} \end{aligned} \tag{2.39}$$

where $C_m=\dfrac{1}{(\lambda_1+\mu)-\mu\lambda_1 C_{m-1}}$ and $C_0=\dfrac{1}{\lambda_1+\mu}$.

(Step 14)

Using $P_{T-1,D}$, $P_{T-1,D-1}$, $P_{T-2,D}$ and $P_{T,D}$ in (2.7), we have

$$P_{B,0} = \frac{\lambda_1 P_{T-2,D}+[\lambda_2 X_{T-1,D-1}+\mu X_{T,D}-(\lambda_1+\mu)X_{T-1,D}]P_{0,0}}{(\lambda_1+\mu)Z_{T-1,D}-\mu Z_{T,D}-\lambda_2 Z_{T-1,D-1}} \tag{2.40}$$

(Step 15)

Now we use (2.40) in all state expressions obtained in each step to express all state probabilities in terms of $P_{0,0}$. Then using the normalization condition

$$Q_{0,0}+\sum_{j=0}^{D}\sum_{i=0}^{B-j}P_{i,j}=1$$

we can get all the state probabilities.

3. Performance Measures

In this section, we derive the mean queue waiting time and loss probabilities for each type of customers.

3.1 Type-1 customers

Using the state probabilities obtained in section 2, we can derive the mean queue size of type-1 customers as

$$L_1=\sum_{n=0}^{B}\Big[n\cdot\sum_{j=0}^{\min(B-n,D)}P_{n,j}\Big] \tag{3.1}$$

Since the type-1 customers who enter the system are always served before they leave the system, we use the Little's formula to obtain the mean queue waiting time,

$$W_1=\frac{L_1}{\lambda_1'} \tag{3.2}$$

where $\lambda_1=\lambda_1(1-P_{B,0})$ is the effective arrival rate.

If the buffer is full and no type-2 customers exist in the buffer, an arriving type-1 customer is blocked and lost. So the loss probability of type-1 customers becomes.

$$P_{\text{LOSS1}} = P_{B,0} \tag{3.3}$$

3.2 Type-2 customers

3.2.1 Mean queue waiting time

Even if a particular type-2 customer has entered the buffer and taken an occupancy, he may be pushed out by a type-1 customer who arrives and sees the full system. Therefore, to analyze the performance characteristics of a type-2 customer, we need some information on the position of the type-2 customer. The analysis incorporating the information on the position of the customers can be found in Doshi & Heffes [4]. In this section, we extend their method to calculate the mean queue waiting time of an arbitrary type-2 customer who is served.

Suppose an arbitrarily chosen type-2 customer (we will call him 'test type-2 customer') arrives and joins the system in position j, $j=0,\cdots,D$ (position 0 denotes the customer currently being served). We define the following state at an 'event completion point' which is either an arrival point or a service completion point. We also define the remaining queue waiting time as the time duration from the current time point till the test type-2 customer begins to be served. We define the state (i,j,k) as

State (i,j,k) : after an event completion occurs, there are i type-1 customers in the queue, the test type-2 customer is in position j among the type-2 customers and there are k type-2 customers behind him.

If $j=0$, then the test type-2 customer is already in service and its remaining queue wating time is zero. If $j>0$, the test type-2 customer is still in the buffer. Followings are the possible cases:

1. (For $i \le T-1$)
   i) The test type-2 customer in position j moves to position $j-1$ if the next event is a service completion.
   ii) If the next event is the arrival of a type-1 customer, i increases to $i+1$.
   iii) ($j+k<D$)
   If the next event is the arrival of a type-2 customer, k increases to $k+1$. ($j+k=D$)
   If the next event is the arrival of a type-2 customer, he is lost and the state is unchanged.

2. (For $i \ge T$)
   i) If the next event is a service completion, i decreases to $i-1$.
   ii) If the buffer is not full and the next event is the arrival of a type-1 customer, i increases to $i+1$.
   iii) (For $j+k<D$)
   If the buffer is not full and the next event is the arrival of a type-2 customer, k increases to $k+1$. (For $j+k=D$)
   If the buffer is not full and the next event is the arrival of a type-2 customer, he is lost and the state is unchanged.
   iv) (In case the buffer is full)
   If the next event is the arrival of a type-2 customer, he is lost.
   If the next event is the arrival of a type-1 customer, he pushes out the type-2 customer from the position 1 and joins the buffer.
   If the test type-2 customer is in position 1, he is pushed out. Otherwise he moves to position $j-1$.

Since the state after the next event depends on the current position of the test type-2 customer, the number of type-1 customers and the number of type-2 customers behind the test type-2 customer, we define the following probabilities and the Laplace-Stieltjes transform(LST),

$\alpha(i,j,k) = Pr$ [the test type-2 customer in state will get served eventually]

$\beta(i,j,k,t) = Pr$ [the test type-2 customer in state (i,j,k) will get served eventually and its remaining queue waiting time will not exceed t]

Note that

$$\beta^*(i,j,k,\theta) = \int_0^\infty e^{-\theta t} d\beta(i,j,k,t)$$

$$\alpha(i,j,k) = \lim_{t\to\infty}\beta(i,j,k,t).$$

Then, we can set up the following recursive equations:

$$\alpha(i,0,k) = 1, \text{ for all } i,\ k$$

$$\alpha(i,j,k) = \frac{\mu}{\lambda+\mu}\alpha(i,j-1,k) + \frac{\lambda_1}{\lambda+\mu}\alpha(i+1,j,k) + \frac{\lambda_2}{\lambda+\mu}\alpha(i,j,k+1)\ (i \le T-1,\ j+k \le D-1)$$

$$\alpha(i,j,D-j) = \frac{\mu}{\lambda_1+\mu}\alpha(i,j-1,D-j) + \frac{\lambda_1}{\lambda_1+\mu}\alpha(i+1,j,D-j)\quad (i \le T-1,\quad j+k = D)$$

$$\alpha(i,j,k) = \frac{\mu}{\lambda+\mu}\alpha(i-1,j,k) + \frac{\lambda_2}{\lambda+\mu}\alpha(i,j,k+1) + \frac{\lambda_1}{\lambda+\mu}\alpha(i+1,j,k)$$

$$(i \ge T,\quad j+k \le D-1)\quad (i \ge T,\quad j+k = D)$$

$$\alpha(i,j,k) = \begin{cases} \dfrac{\mu}{\lambda_1+\mu}\alpha(i-1,j,k) + \dfrac{\lambda_1}{\lambda_1+\mu}\alpha(i+1,j-1,k), & (i \ge T,\ i+j+k = B,\ j \ne 1) \\ \dfrac{\mu}{\lambda_1+\mu}\alpha(i-1,j,k), & (i \ge T,\ i+j+k = B,\ j = 1) \end{cases}$$

Above equations are easily understood once it is understood that $\frac{\lambda_1}{\lambda+\mu}$, for example, is the probability that a type-1 arrival occurs before a service completion or a type-2 arrival occurs.

Defining the LST,

$$\beta^*(i,j,k,\theta) = \int_0^\infty e^{-\theta t} d\beta(i,j,k,t)$$

we can set up the recursive equations with respect to the LST,

$$\beta^*(i,0,k,\theta) = 1 \quad \text{for all } i,\ k \tag{3.4}$$

$$\beta^*(i,j,k,\theta) = \frac{\mu}{\lambda+\mu+\theta}\beta^*(i,j-1,k,\theta) + \frac{\lambda_1}{\lambda+\mu+\theta}\beta^*(i+1,j,k,\theta)\ (i \le T-1,\quad j+k \le D-1) \\ + \frac{\lambda_2}{\lambda+\mu+\theta}\beta^*(i,j,k+1,\theta) \tag{3.5}$$

$$\beta^*(i,j,D-j,\theta) = \frac{\mu}{\lambda_1+\mu+\theta}\beta^*(i,j-1,D-j,\theta) + \frac{\lambda_1}{\lambda_1+\mu+\theta}\beta^*(i+1,j,D-j,\theta) \\ (i \le T-1,\quad j+k = D) \tag{3.6}$$

$$\beta^*(i,j,k,\theta) = \frac{\mu}{\lambda+\mu+\theta}\beta^*(i-1,j,k,\theta) + \frac{\lambda_1}{\lambda} + \mu + \theta\beta^*(i+1,j,k,\theta) + \frac{\lambda_2}{\lambda+\mu+\theta}\beta^*(i,j,k+1,\theta)$$

$$\alpha(i,j,D-j) = \frac{\mu}{\lambda_1+\mu}\alpha(i-1,j,D-j) + \frac{\lambda_1}{\lambda_1+\mu}\alpha(i+1,j,D-j)\ (i \ge T,\quad j+k \le D-1) \tag{3.7}$$

$$\beta^*(i,j,D-j,\theta) = \frac{\mu}{\lambda_1+\mu+\theta}\beta^*(i-1,j,D-j,\theta) + \frac{\lambda_1}{\lambda_1+\mu+\theta}\beta^*(i+1,j,D-j,\theta) \\ (i \ge T,\quad j+k = D) \tag{3.8}$$

$$\beta^*(i,j,k,\theta) = \begin{cases} \dfrac{\mu}{\lambda_1+\mu+\theta}\beta^*(i-1,j,k,\theta) + \dfrac{\lambda_1}{\lambda_1+\mu+\theta}\beta^*(i+1,,j-1,k,\theta) \\ \qquad (i \ge T,\ i+j+k = B,\ j \ne 1) \\ \dfrac{\mu}{\lambda_1+\mu+\theta}\beta^*(i-1,j,k,\theta), \qquad (i \ge T,\ i+j+k = B,\ j = 1) \end{cases} \tag{3.9}$$

Now we can obtain the mean remaining queue waiting time from

$$W(i,j,k) = -\lim_{\theta\to 0}\frac{d}{d\theta}\beta^*(i,j,k,\theta)$$

where $\alpha(i,j,k)=\lim_{\theta\to 0}\beta^*(i,j,k,\theta)$ from the Tauberian theorem.

From (3.4)-(3.9), we get

$$W(i,0,k)=0 \quad \text{for all } i,k$$

$$W(i,j,k) = \frac{1}{\lambda+\mu}\alpha(i,j,k)+\frac{\mu}{\lambda+\mu}W(i,j-1,k) + \frac{\lambda_1}{\lambda+\mu}W(i+1,j,k)+\frac{\lambda_2}{\lambda+\mu}W(i,j,k+1) \qquad (i\le T-1,\quad j+k\le D-1)$$

$$W(i,j,D-j) = \frac{1}{\lambda_1+\mu}\alpha(i,j,D-j)+\frac{\mu}{\lambda_1+\mu}W(i,j-1,D-j) + \frac{\lambda_1}{\lambda_1+\mu}W(i+1,j,D-j) \qquad (i\le T-1,\quad j+k=D)$$

$$W(i,j,k) = \frac{1}{\lambda+\mu}\alpha(i,j,k)+\frac{\mu}{\lambda+\mu}W(i-1,j,k) + \frac{\lambda_1}{\lambda+\mu}W(i+1,j,k)+\frac{\lambda_2}{\lambda+\mu}W(i,j,k+1) \qquad (i\ge T,\quad j+k\le D-1)$$

$$W(i,j,D-j) = \frac{1}{\lambda_1+\mu}\alpha(i,j,D-j)+\frac{\mu}{\lambda_1+\mu}W(i-1,j,D-j) + \frac{\lambda_1}{\lambda_1+\mu}W(i+1,j,D-j) \qquad (i\ge T,\quad j+k=D)$$

$$W(i,j,k) = \begin{cases} \frac{1}{\lambda_1+\mu}\alpha(i,j,k)+\frac{\mu}{\lambda_1+\mu}W(i-1,j,k)+\frac{\lambda_1}{\lambda_1+\mu}W(i+1,j-1,k) & (i\ge T,\ i+j+k=B,\ j\ne 1) \\ \frac{1}{\lambda_1+\mu}\alpha(i,j,k)+\frac{\mu}{\lambda_1+\mu}W(i-1,j,k), & (i\ge T,\ i+j+k=B,\ j=1) \end{cases}$$

$W(i,j,k)$ for all states can be obtained recursively. Remember that $W(i,j,k)$ is the mean remaining queue waiting time for the test type-2 customer who has entered the buffer and gets served eventually. We observe that just after an arrival of the test type-2 customer there are no type-2 customers behind him. Thus $W(i,j+1,0)$ is the mean remaining waiting time just after his arrival. From PASTA (Wolff[10]), this occurs with probability $P_{i,j}$. Thus we have the mean queue waiting time of the type-2 customer who is served as

$$W_2=\frac{\sum_{j=0}^{D-1}\sum_{i=0}^{B-1-j}P_{i,j}W(i,j+1,0)}{P_s} \qquad (3.10)$$

where

$Pr=\sum Pr$ [a type-2 customer gets served | he is allowed to enter the system] $\cdot Pr$ [a type-2 customer is allowed to enter the system]

$$=\sum_{j=0}^{D-1}\sum_{i=0}^{B-1-j}\alpha(i,j+1,0)P_{i,j}+Q_{0,0} \qquad (3.11)$$

If the server is idle when the test type-2 customer arrives, he immediately goes into service.

3.2.2. Loss probability

The loss probability of the test type-2 customer is the sum of the blocking probability and the probability that he is pushed out. The blocking probability becomes

$$P_B=\sum_{i=0}^{B-D}P_{i,D}+\sum_{\substack{i+j=B,\\ j\ne D}}P_{i,j}$$

and the pushed-out probability becomes

$$\sum_{j=0}^{D-1}\sum_{i=0}^{B-1-j} P_{i,j}\cdot[1-\alpha(i,j+1,0)]$$

So the loss probability of type-2 customers becomes

$$P_{\text{Loss2}} = \sum_{j=0}^{D}\sum_{i=0}^{B-j} P_{i,j}\cdot[1-\alpha(i,j+1,0)] \quad (3.12)$$

$$W_2 = \frac{\sum_{j=0}^{D-1}\sum_{i=0}^{B-1-j} P_{i,j}W(i,j+1,0)}{P_s}$$

4. Performance analysis

In this section, we show some numerical results and present the optimal buffer control policy that satisfies the quality of service requirements. For computational purposes, we assume that service rate, μ, is fixed at 1 for all cases. From the fixed service rate, the offered load($\rho=\lambda/\mu$) is determined by total service rate($\lambda=\lambda_1+\lambda_2$).

We compare the loss probabilities and the mean queue waiting times of both type of customers for different load ratios of type-2 customers (λ_2/λ) and offered loads ($\rho=\lambda/\mu$). We assume $B=30$, $T=15$ and $D=8$.

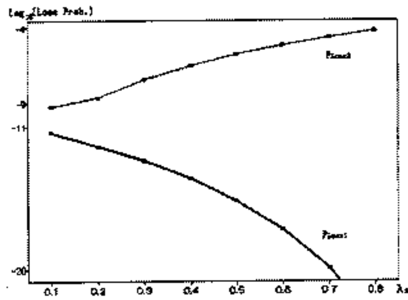

Figure 2. Loss probabilities ($\rho=0.5, B=30, T=15, D=8$)

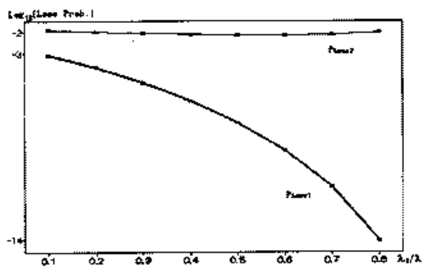

Figure 3. Loss probabilities ($\rho=0.9, B=30, T=15, D=8$)

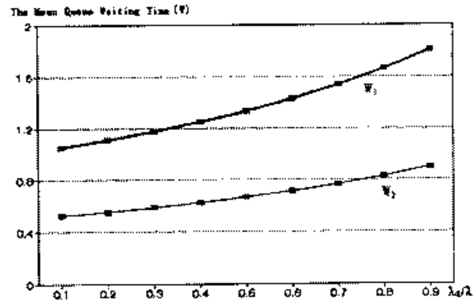

Figure 4. Mean queue waiting times ($\rho=0.5, B=30, T=15, D=8$)

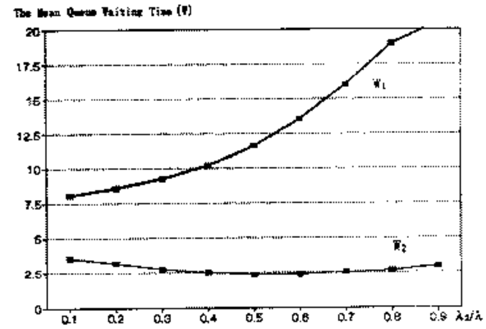

Figure 5. Mean queue waiting times ($\rho=0.9, B=30, T=15, D=8$)

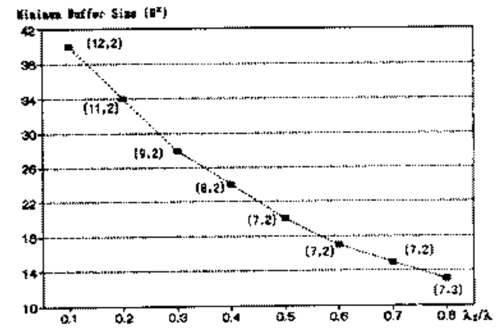

Figure 6. optimal buffer size, (T^*, D^*) ($W_1/W_2 \geq 3$, $\rho=0.7$)

<Figures 2> and 3 show the loss probabilities P_{LOSS1} and P_{LOSS2} for different load ratios of type-2 customers. <Figures 4> and 5 show the change of mean queue waiting time of each type of customers. The loss probabilities of type-1 customers vary from 10^{-25} to 10^{-3} for different values of load ratios. We also see that mean queue waiting times of both types of customers increase as the load ratio of type-2 customer increases. This occurs because as λ_2 increases (with λ fixed), the relative

arrival rate x decreases. Then more λ_1 frequently the system occupancy is below T and the chance of getting service decreases.

Next we derive the minimum buffer control policy which satisfies the quality of service requirements as the load ratio of type-2 customers increases (with the offered load is fixed at $\rho = 0.7$). We assume that the required loss probability is 10^{-8} for type-1 customers and 10^{-3} for type-2 customers. Taking the mean waiting times into consideration, we consider the ratio W_1/W_2. The reason why we consider the ratio is that even though the waiting times may be negligible due to the small service time of ATM network, the ratio may not be so either for the fairness of service or for the discrimination of service. <Figure 6> show the minimum buffer size as the load ratio aries.

References

1. Ahn, B. Y. and Lee, H. W., "A queueing model for optimal control of partial buffer sharing in ATM," submitted for publication, 1997.
2. Ahn, B. Y. and Lee, H. W., "A queueing model for erro: control of partial buffer sharing in ATM," working paper, Dept. of Industrial Engineering, Sung Kyun Kwan University, Korea, 1996.
3. Cha, H. C. and Han, K. C., "Performance analysis of ATM switch with HOL-T priority," *JCCI* '92, pp. 197-202, 1992.
4. Doshi, B. T. and Heffes, H., "Overload performance of several processor queueing disciplines for the $M/M/1$ Queue," *IEEE Trans. Comm.*, Vol. 34, No. 6, pp. 538-546, 1986.
5. Gravey, A. and Hebuterne, G., "Analysis of a priority queue with delay and/or loss sensitive customers," *in Proc. 7th ITC Sem.*, New Jersey, Oct. 1990, paper 13.2.
6. Kröner, H., "Comparative performance study of space priority mechanisms for ATM networks," *IEEE INFOCOM* '90, pp. 1136-1143, 1990.
7. Kröner, H., Theimer, T. H. and Briem, U., "Queueing models for atm systems-a comparison," *in Proc. 7th ITC Sem.*, New Jersey, Oct. 1990, paper 9.1.
8. Neuts, M. F., "A queueing model for a storage buffer in which the arrival rate is controlled by a switch with a random delay," *Perfm. Eval.*, Vol. 5, pp. 243-256, 1985.
9. Rothermel K., "Priority mechanisms in ATM networks" *IEEE GLOBECOM '90*, pp. 847-851, 1990.
10. Saito, H., "Optimal discipline for real-time traffic at ATM switching nodes," *IEEE Trans. Commun.*, Vol. 38, No. 12, 1990.
11. Wolff, R. W., "Poisson arrivals see time averages," *Oper. Res.*, Vol. 30, pp. 223-231, 1982.

Received October 1996, accepted December 1998