Dynamic Bandwidth Allocation Scheme for Multiple Real-time VBR Videos over ATM Networks

Chae Y. Lee, Seok J. Koh* and Hee M. Eun

Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, 373-1 Kusung Dong, Taejon 305-701, Korea

E-mail: cylee@heuristic.kaist.ac.kr

* Protocol Engineering Center, ETRI, 161 Kajung Dong, Taejon, 305 -350, Korea

Abstract

This paper considers VBR transmission of multiple real -time videos over ATM networks. Multiple real-time VBR video sources are multiplexed into an ATM switch to transmit cells into the network. Given the ATM switch capacity, the problem is to dynamically allocate the required channel bandwidth for each video source such that the e ncoder buffer occupancy is maintained at a target level. To solve the problem, we present a mathematical formulation and propose an algorithm for the bandwidth allocation. To allocate a suitable bandwidth at a given control period, QoS demand levels and tr affic characteristics of the video sources are considered. The performance of the proposed scheme is examined in terms of the number of encoder rate controls required and the gap between the target and the current buffer occupancy at each control period. N umerical results are analyzed for different QoS environments as well as different levels of target buffer, ATM switch capacity, buffer size and leaky bucket token rate.

Key Words: Real-time VBR video, Bandwidth Allocation, ATM Networks.

I. Introduction

Recent progress on the standardization of high-speed networking and digital video technology has led to active commercial development of various video services such as video conferencing, videophone and TV broadcast. The video output generated by MPEG (Moving Picture Experts Group) coder [1] is intrinsically variable bit rate (VBR) for most practical compression algorithms. The VBR nature of compressed video gives rise to a motivation for establishing networks which allow video transmission at variable bit rate while providing the quality of service (QoS) guarantees such as cell loss, cell delay and cell delay variations. Asynchronous transfer mode (ATM) network [2] is an example of a network architecture which would allow this type of VBR transmission.

We consider VBR transmission of real-time video sources over ATM networks. To provide reliable video quality in ATM networks, it is very important to effectively manage network resources including buffers and channel bandwidth. Given ATM switch capacity, we propose a dynamic bandwidth allocation scheme for multiple real-time VBR videos such that video quality and network QoS constraints are guaranteed.

Recent studies related to the real-time VBR video can be classified into three categories; traffic modeling, constant bit rate (CBR) transport and VBR transport. In the study of traffic modeling [3, 4, 5], traffic models for real-time video source have been proposed. In [3], it is shown that the traffic of video teleconference is accurately characterized by a multistate Markov chain model. In [4], the number of cells per frame is not normally distributed for video teleconference sequence. Instead, it follows a gamma distribution and a stationary stochastic process. The study in [5] proposes to model a single video source as a Markov renewal process whose states represent different bit rates.

In the CBR transport [6, 7, 8, 9], it is assumed that the channel bandwidth for a video source is constant. Current proposals for real-time VBR videos use CBR transport on the ATM. Given a constant channel bandwidth which is negotiated between a user and a network, MPEG encoder adjusts the source coding rate based on the quantization control. In [6], a source rate control scheme is proposed based on the adjustment of the quantization rate (Q). When a buffer occupancy level is in potential overflow region, the MPEG encoder increases Q to decrease the source output rate. In the opposite case, Q is decreased to improve the quality of VBR videos. In [7], the required channel bandwidth for a video source is estimated in the CBR transport. Luo and Zarki [8] present the relationship between picture quality and encoder rate control for different combinations of channel bandwidth, buffer size and quantization rate. In [9], a source rate control algorithm is presented based on the leaky bucket controller. However, CBR transport for VBR video suffers from disadvantages such as variable video quality and relatively high transmission cost.

In the study of VBR transport [10, 11, 12], a joint cont rol of encoder rate and channel bandwidth is considered to maintain the end -to-end delay of transmitted videos in the appropriate level that is suitable to reliable display. In [10], the authors present conditions to ensure that the video encoder and decod er buffers do not overflow or underflow when a channel can transmit a VBR service. In the study, it is assumed that the cell transmission and switching delay in ATM networks are constant. In [11], it is shown that increasing the delay in the video buffers decreases the necessary peak bandwidth and significantly increases the number of calls that can be carried by the network. In [12], the source encoder rate and the channel rate are jointly selected to ensure that real -time video display at the decoding end is possible. However, these studies of VBR transmission have focused on the end-to-end performance of a single video source. Thus the ATM switching capacity and the channel bandwidth allocation have not been considered.

In this paper, we consider VBR tran smission of multiple real -time video sources over ATM networks. To allow uniform picture quality of video sources, the MPEG encoder initially generates frames by using constant quantization rate, but when the excess traffic is generated which violates the negotiated traffic parameters, frames may be degraded due to increased quantization rate. In the system considered in this paper, multiple real -time VBR videos are multiplexed into an ATM switch to transmit cells into the network. Given ATM switch capacity, the problem is to allocate the required channel bandwidth for each video source such that the encoder buffer occupancy is maintained at a target level.

This paper is organized as follows. In Section II, a video transmission system model is described for VBR video transmission over ATM networks. In Section III, the bandwidth allocation problem and the relationship with existing studies are described. In Section IV, we propose a dynamic bandwidth allocation scheme. Section V shows the performance of the proposed scheme through numerical experiments by varying the related system parameters; buffer size, an ATM switch capacity, target buffer occupancy level and leaky bucket token generation rate. Section VI concludes this paper with future works.

II. Video Transmission Model

Fig. 1 shows an end-to-end model for real-time VBR video over ATM networks [6]. The VBR/ATM sender consists of an encoder and an ATM Network Interface Card (NIC). The MPEG encoder processes the raw video source and passes the bitstream to the NIC. The MPEG bitstream consists of units of group of pictures (GOP's) containing an Intra (I) picture, Predictive (P) pictures and bidirectional or interpolated (B) pictures [1]. The realization of a GOP pattern may be implementation-specific. However, a GOP of size 12 is recommended by ISO [1], and the GOP pattern is IBBPBBPBBPBB, and two GOP's are transmitted for a second. In this case, one frame period corresponds to 1/24 second. Typically, the target quantization rate is set to be 3 for I frame, 4 for P frame and 6 for B frame, respectively.

The NIC prepares the bitstream for ATM delivery by segmenting data into cells and adding appropriate ATM adaptation layer. The VBR traffic profile is defined through a set of traffic descriptors such as peak rate, burst length and sustained rate. These parameters are referred as a usage parameter control (UPC) set [2]. The network admits a VBR connection based on its declared UPC. Once the connection is established, it is expected that the terminal device will comply with the declared UPC. The network may enforce the declared UPC using a leaky bucket (LB) based network policing mechanism. The LB mechanism may be implemented to control the peak cell rate, average cell rate and burst length of each source using single or double LB [12]. In real-world deployments, however, a simple version of LB application is commonly used to police the peak cell rate which is negotiated between a user and a network at the connection setup phase [2]. A cell, before entering the network, must obtain a token from the LB token pool. Tokens are generated at a constant rate and placed in the token pool. To police the peak cell rate, it is sufficient to set the token generation rate to the peak cell rate.

The NIC also participates in the control of quantization rate to maintain the buffer occupancy at a specified level which is negotiated between a user and a network. The UPC controller sends a target quantization rate to the MPEG encoder to adjust the source coding rate. In general, a larger quantization rate induces a smaller source output rate by MPEG encoder and a poorer video quality.

The signaling and traffic control messages are exchanged between ATM network and NIC to support QoS of the transmitted video. At the VBR/ATM receiver, NIC reassembles the ATM cells and terminates the transport protocol. Thus the transported real-time video is displayed through MPEG decoder and a user's set-top box (STB).

III. Problem Formulation

In this section, we examine the system environment considered in this paper and provide a mathematical model to allocate the bandwidth required for each source. We also compare the existing studies with the proposed model.

Fig. 2 shows the transmission system for multiple real-time VBR videos. In the system, *n* real-time video sources are multiplexed to an ATM switch. The ATM switch capacity is given to be *C* (cells/second). Each video source is encoded by MPEG coder using constant quantization rate for all I, P and B frames to allow uniform video quality. We define $E^s(t)$ as the encoder rate for source *s* at frame period $t = 1, 2, \ldots T$, where T is the number of total frames. The encoded cell goes into the buffer in NIC. For each source *s*, we define $B^s(t)$ as the buffer occupancy at frame period *t* and $B^s max$ as the buffer size.

Each cell in the buffer is allowed to enter the ATM network by taking a token from the LB token pool. In this paper, a simple LB policing mechanism is employed to police the peak cell rate (PCR) of each video source. The PCR is negotiated between a user and a network at the connection setup phase. Tokens are generated constantly at the PCR and placed in the LB token pool. We define L^s as the LB token generation rate or PCR for video source *s*. LB policer constrains the channel bandwidth for the source *s* not to exceed L^s.

Now we define $\lambda^s(t)$ as the cell transmission rate or channel bandwidth for the source *s* at frame period *t*. Note that the channel bandwidth is allocated based on the ATM switch capacity *C* (cells/second) as well as the buffer occupancy $B^s(t)$ and the LB parameter L^s. Then the problem is to allocate the required bandwidth dynamically for each video source such that the buffer occupancy is maintained at a target level. Let us define $B^s target$ as the target buffer occupancy level, where $B^s target$ is ranged from 0 to $B^s max$. $B^s target$ is determined based on the QoS requirement of video source *s*. For example, $B^s target$ is set to be near to zero for delay-sensitive services, and $B^s max$ for delay-tolerant services. We also consider QoS demand level, QoS(s), for each source *s*. It is assumed that the value of QoS(s) is given as a positive integer by service providers. The source with higher QoS(s) has the priority to the lower one.

Thus the variables or parameters considered in this paper are summarized for each source *s* at frame period t as follows;

System parameters: *C*, $B^s max$, L^s and $B^s target$,

Input variables: $E^s(t)$, $B^s(t-1)$,

Decision variables: $\lambda^s(t)$.

Then the bandwidth allocation problem is formulated at each frame period *t* as follows;

$$\text{Minimize} \quad \sum_{s=1,\ldots,n} QoS(s) \times |B^s(t) - B^s target| \tag{1}$$

$$\text{Subject to} \quad B^s(t) = B^s(t-1) + E^s(t) - \lambda^s(t) \tag{2}$$

$$0 \leq B^s(t) \leq B^s max \tag{3}$$

$$0 \leq \lambda^s(t) \leq L^s \tag{4}$$

$$\sum_{s=1,\ldots,n} \lambda^s(t) \leq C \tag{5}$$

This model allocates the bandwidth required for each video source *s* at frame period *t* such that the absolute deviation between $B^s(t)$ and B^starget is minimized as shown in (1). The QoS demand levels, QoS(s), must be also reflect ed on the bandwidth allocation. Constraint (2) represents the current buffer occupancy level $B^s(t)$, which is derived from previous buffer occupancy $B^s(t-1)$, the encoded source data $E^s(t)$ and the channel bandwidth $\lambda^s(t)$. In constraint (3), the buffer occupa ncy $B^s(t)$ is constrained by the buffer size, B^smax. In constraint (4), the channel bandwidth $\lambda^s(t)$ is bounded by the LB token generation rate L^s. The constraint (5) represents that the sum of channel bandwidths of all individual sources should not exceed the ATM switch capacity *C*.

This model also depends on the selection of appropriate target buffer occupancy B^starget for each source *s*. If B^starget is set to be zero, then the model is to minimize total cell delay in the buffer. If Bstarget is set to be B^smax, then the buffer occupancy will converge to the B^smax. Thus the model minimizes the bandwidth allocated for each source. Instead, the cell loss probability becomes higher in the buffer.

The QoS can be guaranteed by tuning the cell loss and cell delay var iation parameters. In the model, the objective function minimizes the variation of cell delays. The cell loss is tuned by constraints (3), (4), and (5).

The model considered in this paper is different from existing studies. In CBR transport [6 - 9], the channel bandwidth for a source s is fixed to be the constant peak rate, i. e., $\lambda^s(t) = \lambda^s$ for each frame period *t* and the source encoder rate $E^s(t)$ is adjusted by MPEG encoder. However, the proposed model provides the channel bandwidth required for each source and the encoder rate $E^s(t)$ to allow uniform video quality. The study of VB R transport [10-12] considers the end-to-end performance of a single video transmission. Thus the bandwidth allocation based on ATM switch capacity is not considered. In particular, they have a focus on the non-real time video transmission during overall f rame period.

IV. Dynamic Bandwidth Allocation Scheme

The bandwidth allocation problem is not easy and time -consuming to solve. In the worst case, we may have to examine all possible combination of $\lambda^s(t)$ for all sources. Thus the bandwidth allocation problem is decomposed into subproblems, each of which solves the optimal bandwidth allocation for each individual source. Based on the properties of bandwidth allocation, we propose a dynamic bandwidth a llocation scheme for multiple real-time VBR videos at each frame period.

A. Problem Decomposition

The dynamic bandwidth allocation (DBA) problem considered in this paper is to allocate the bandwidth $\lambda^s(t)$ required for each source *s* at frame period *t*. Given system parameters $B^s max$, L^s, *C* and $B^s target$, we use the input variables $B^s(t-1)$ and $E^s(t)$ at frame period *t*. The objective is to find real-time bandwidth $\lambda^s(t)$ for each source *s* that minimizes the absolute deviation between $B^s(t)$ and $B^s target$, such that the buffer, LB and ATM switch capacity constraints are satisfied.

In the model, constraints (2), (3) and (4) are imposed on each individual source. To decompose the problem into subproblems for each individual source, we define the maximum allowable bandwid th $\lambda^{s*}(t)$ for each source *s* and frame period *t*, based on the capacity constraint (5). The methods to obtain such $\lambda^{s*}(t)$ will be discussed in the next section. Thus the constraint (5) is modified as

$$\lambda^s(t) \leq \lambda^{s*}(t) \quad . \qquad (6)$$

Now, we consider the following indi vidual DBA (IDBA) problem for each source:

Minimize $QoS(s) \times |B^s(t) - B^s target|$, subject to the constraints (2), (3), (4) and (6).

The IDBA problem is to find the bandwidth required for a source *s* such that the buffer, UPC and capacity constraints are sa tisfied. If the maximum allowable bandwidth $\lambda^{s*}(t)$ for each source *s* is known a priori, DBA problem can be decomposed into IDBA problem for each individual source. Thus we solve the IDBA problem by allocating the bandwidth required for each individual source.

The following lemma shows that the objec tive function (1) is linear convex over the channel bandwidth $\lambda^s(t)$ as shown in Fig. 3.

Lemma 1. The objective function $QoS(s) \times |B^s(t) - B^s target|$ is linear convex over $\lambda^s(t)$.

(Proof) Given $E^s(t)$ and $B^s(t-1)$, $B^s(t) = B^s(t-1) + E^s(t) - \lambda^s(t)$. Then the obje ctive function is

$$QoS(s) \times |B^s(t) - B^s target| = QoS(s) \times | B^s(t-1) + E^s(t) - B^s target - \lambda^s(t) |.$$

Note that the value of $B^s(t-1) + E^s(t) - B^s target$ is fixed at the frame period *t*. Thus the objective function can be represented as the linear convex function o f $\lambda^s(t)$ as shown in Fig. 3. In the figure, $\lambda^s target$ is obtained by setting $|B^s(t) - B^s target| = 0$. That is, $\lambda^s target = B^s(t-1) + E^s(t) - B^s target$.

The following lemma determines the range of the bandwidth $\lambda^s(t)$ required for the source *s*. The range of $\lambda^s(t)$ is derived from the constraints (2), (3), (4) and (6).

Lemma 2. In the IDBA problem, the feasible region of $\lambda^s(t)$ is ranged as follows;

$$\max \{0, B^s(t-1) + E^s(t) - B^s max\} \le \lambda^s(t) \le \min \{ B^s(t-1) + E^s(t), L^s, \lambda^{s*}(t) \}. \quad (7)$$

Also, if $B^s(t-1) + E^s(t) - B^s max > \min \{ L^s, \lambda^{s*}(t) \}$, the IDBA problem is infeasible for the given $E^s(t)$ and $B^s(t-1)$.

(Proof) We assume that $B^s max$, L^s, $\lambda^{s*}(t)$, $B^s(t-1)$ and $E^s(t)$ are given to be fixed at the frame period *t*. From the constraints (2) and (3), $0 \le B^s(t-1) + E^s(t) - \lambda^s(t) \le B^s max$. Thus, $B^s(t-1) + E^s(t) - B^s max \le \lambda^s(t) \le B^s(t-1) + E^s(t)$. By adding the constraints (4) and (6), we obtain the result (7). In (7), if $B^s(t-1) + E^s(t) - B^s max > \min \{L^s, \lambda^{s*}(t)\}$, then we cannot obtain any feasible $\lambda^s(t)$ for the given $E^s(t)$ and $B^s(t-1)$.

In Lemma 2, if the problem becomes infeasible, the encoder rate control based on the quantization rate is required to decrease the source output rate $E^s(t)$. This can be done by increasing the quantization rate in the MPEG encoder. Specifically, $E^s(t)$ should be decreased such that $B^s(t-1) + E^s(t) - B^s max \le \min \{L^s, \lambda^{s*}(t) \}$ is satisfied. In this case, the uniform video quality cannot be guaranteed for the source *s* at the frame period *t*. Thus the required number of encoder rate controls at each frame period is considered as a performance measure of the DBA scheme as e xplained in Section V.

Now, we consider an IDBA problem which has a feasible solution. Let us define the following variables;

$$\lambda^s target = B^s(t-1) + E^s(t) - B^s target,$$

$$\lambda^s min = \max \{0, B^s(t-1) + E^s(t) - B^s max\},$$

$$\lambda^s max = \min \{ B^s(t-1) + E^s(t), L^s, \lambda^{s*}(t) \}.$$

Note that λ^starget is the bandwidth for which the objective function becomes zero. λ^smin and λ^smax are respectively the minimum and maximum values in the range of feasible bandwidths in the constraint (7). Then the following lemma characterizes the optimal bandwidth allocation for the source *s* as shown in Fig. 3.

Lemma 3. For the source s at the frame period *t*, the optimal bandwidth $\lambda^s(t)$ is allocated as follows;

Case 1. If $\lambda^s min \leq \lambda^s target \leq \lambda^s max$, then $\lambda^s(t) = \lambda^s target$,

Case 2. If $\lambda^s target < \lambda^s min$, then $\lambda^s(t) = \lambda^s min$,

Case 3. If $\lambda^s target > \lambda^s max$, then $\lambda^s(t) = \lambda^s max$.

(Proof) In Lemma 1, the objective function is linear convex over $\lambda^s(t)$. The range of $\lambda^s(t)$ corresponds to one of the following three cases as shown in Fig. 3. In Case 1, we obtain the optimal $\lambda^s(t)$ such that $B^s(t) = B^s$target. In Case 2 and 3, λ^smin and λ^smax are selected as the optimal solution such that the objective function is minimized, respectively.

The following theorem characterizes some conditions on system parameters to obtain an optimal solution of $\lambda^s(t) = \lambda^s$target.

Theorem 1. An optimal bandwidth allocation of $\lambda^s(t) = \lambda^s$target can be obtained, if the following conditions on system parameters B^smax, L^s, and $\lambda^{s*}(t)$ are satisfied:

(a) $\lambda^s target = B^s(t-1) + E^s(t) - B^s target \geq 0$;

(b) $\lambda^s target = B^s(t-1) + E^s(t) - B^s target \leq \min \{L^s, \lambda^{s*}(t) \}$.

(Proof) From Case 1 of Lemma 3, we obtain the following inequalities;

$$\max\{0,B^s(t-1)+E^s(t) - B^s max\} \leq B^s(t-1)+E^s(t) - B^s target \leq \min\{B^s(t-1)+E^s(t),L^s,\lambda^{s*}(t)\}.$$

From the left inequality, since $B^s max \geq B^s target$, we obtain the following inequality;

$$B^s(t-1) + E^s(t) - B^s target \geq 0.$$

From the right inequality, since $B^s target \geq 0$, we obtain the following inequality;

$$B^s(t-1)+E^s(t) - B^s target \leq \min\{L^s, \lambda^{s*}(t)\}.$$

In the theorem, Case 2 of Lemma 3 occurs if and only if λ^starget < 0. Then, the optimal solution is $\lambda^s(t) = \lambda^s$min = 0. Similarly, Case 3 of Lemma 3 occurs if and only if λ^starget > min $\{L^s, \lambda^{s}*(t)\}$. The optimal solution for this case is $\lambda^s(t) = \lambda^s$max = min $\{L^s, \lambda^{s}*(t)\}$. In fact, Case 1 with the optimal band width allocation of $\lambda^s(t) = \lambda^s$target occurs exactly when 0 ≤ λ^starget ≤ min $\{L^s, \lambda^{s}*(t)\}$.

Note that Case 2 occurs when the current buffer occupancy is very much below the target threshold B^starget, so that the data generated during the current frame are no t enough for increasing the buffer occupancy to the target value. Thus $\lambda^s(t) = \lambda^s$min = 0 is appropriate. On the other hand, Case 3 occurs when the current buffer occupancy has become very much greater than the target threshold. A large amount of bandwidth λ^starget is required to reduce the buffer occupancy to the target value, but this bandwidth cannot be granted because it exceeds either the available capacity $\lambda^{s}*(t)$ or the leaky bucket constraint L^s. The best solution is to allocate the maximum bandwidth that does not violate the constraints; hence $\lambda^s(t) = \lambda^s$max = min $\{L^s, \lambda^{s}*(t)\}$ is appropriate. Finally, Case 1 applies when the conditions of Theorem 1 hold. Then the natural bandwidth allocation $\lambda^s(t) = \lambda^s$target may be granted.

B. Bandwidth Allocation Scheme for Multiple VBR Videos

In the previous section, we described the bandwidth allocation algorithm for each individual video source. In the algorithm, it is assumed that the maximum allowable bandwidth $\lambda^{s}*(t)$ is fixed for each source *s* in the constraint (6). Given $\lambda^{s}*(t)$, the IDBA algorithm produces an optimal bandwidth allocation.

In this section, we present the DBA scheme for multiple VBR video sources. In particular, we propose a scheme to allocate $\lambda^{s}*(t)$ for each source *s*, based on QoS demand levels and source traffic characteristics.

Case 1. Different QoS levels

Consider *n* video sources with ATM switch capacity *C*. Let us assume that each video source *s* has a distinct QoS(s) as a positive integer value. A larger QoS(s) requires higher priority. In fact, QoS(s) may be given to each video source, based on the contract or negotiation between service providers and users.

Based on QoS(s) values, we assign the maximum allowable bandwidth $\lambda^{s}*(t)$ to each source during overall frame period $t = 1, .. T$, as follows:

$$\lambda^{s*}(t) = \frac{QoS(s)}{\sum_{s=1}^{n} QoS(s)} \times C \qquad (8)$$

Thus the maximum allowable bandwidth is given by the portion of relative priority for all sources. Note that the bandwidth allocation policy depends only on the gi ven QoS(s) values, not on the characteristics of source traffic at each frame period. This is because the sources with higher QoS(s) must not be affected by those with smaller ones. However, if some video sources have the same QoS level, they may be given $\lambda^{s}*(t)$ according to the characteristics of source traffic, as described below.

Case 2. Same QoS levels

We now assume that each video source has the same QoS class for all sources. In this case, we assign a priority weight w^s to video source s at frame period t based on the characteristic of source traffic. Then n video sources are rearranged in the non -increasing order of w^s such that the video source $s = 1$ has the highest priority and $s = n$ has the lowest one.

The following policies are proposed to assig n w^s to each source s:

(a) $w^s = 1$,

(b) $w^s = E^s(t)$,

(c) $w^s = B^s(t-1)$,

(d) $w^s = E^s(t) + B^s(t-1)$.

In the policy (a), all sources use the same $\lambda^{s}*(t) = C/n$, which implies no differentiation for all sources. In other cases, the traffic characteristic at frame period t will give an impact on the selection of $\lambda^{s}*(t)$. In (b), the encoder rate $E^s(t)$ is used as w^s to reflect the newly generated video traffic on the bandwidth allocation, while the policy (c) considers the current buffer occupancy $B^s(t-1)$ to represent the bandwidth allocation behavior of the past. In (d), both the encoder rate and the buffer occupancy are considered. The selection of a suitable policy may depend on the traffic engineering requirements of service providers and the availability of information on those data. In any case, the maximum allowable bandwidth $\lambda^{s}*(t)$ can be determined for each source s as follows:

$$\lambda^{s*}(t) = \frac{w^s}{\sum_{s=1}^{n} w^s} \times C \qquad (9)$$

Based on the discussion above, the DBA scheme is summarized as follows;

Dynamic bandwidth allocation at frame period t

Step 1. Initialize $B^s(t\text{-}1)$, $E^s(t)$, L^s, $B^s max$, $B^s target$, $\lambda^s(t) = 0$ for each source *s*.

Step 2. Assign the priority weight w^s to each source *s* by one of the four policies.

Step 3. Using (8) or (9), calculate $\lambda^{s*}(t)$ for each source *s*.

Step 4. Arrange all the video sources $s = 1, .., n$ in the non-decreasing order of $\lambda^{s*}(t)$.

Step 5. $s = 1$.

Step 6. For the source *s*, do the following procedures:

If $B^s(t\text{-}1) + E^s(t) - B^s max > \min \{ L^s, \lambda^{s*}(t) \}$,

Then decrease the encoder rate $E^s(t)$ by increasing the quantization rate

such that $E^s(t) = B^s max - B^s(t\text{-}1) + \min \{ L^s, \lambda^{s*}(t) \}$, and then go to Step 7.

Else go to Step 7.

Step 7. Obtain the following values:

$\lambda^s target = B^s(t\text{-}1) + E^s(t) - B^s target$,

$\lambda^s min = \max \{0, B^s(t\text{-}1) + E^s(t) - B^s max\}$,

$\lambda^s max = \min \{ B^s(t\text{-}1) + E^s(t), L^s, \lambda^{s*}(t) \}$.

Step 8. Allocate the required bandwidth $\lambda^s(t)$ as follows;

If $\lambda^s min \leq \lambda^s target \leq \lambda^s max$, then $\lambda^s(t) = \lambda^s target$,

If $\lambda^s min > \lambda^s target$, then $\lambda^s(t) = \lambda^s min$,

If $\lambda^s target > \lambda^s max$, then $\lambda^s(t) = \lambda^s max$.

Step 9. $B^s(t) = B^s(t\text{-}1) + E^s(t) - \lambda^s(t)$.

Step 10. If $\lambda^s(t) < \lambda^{s*}(t)$, then $\lambda^{s+1*}(t) = \lambda^{s+1*}(t) + \{\lambda^{s*}(t) - \lambda^s(t)\}$

Step 11. Set $s = s + 1$, and then go to Step 6.

In Step4, all video sources are arranged in non -decreasing order of w^s. Thus the lowest-weighted source will first perform the bandwidth allocation. In Step 10, if the allocated bandwidth $\lambda^s(t)$ is less than the given maximum allowable bandwidth $\lambda^{s*}(t)$, then the residual capacity $\{\lambda^{s*}(t) - \lambda^s(t)\}$ is assigned to the next source s+1. In this way, it is likely that the highest-weight source utilizes the more capacity.

V. Numerical Results

A. Test VBR videos

To test the performance of the DBA scheme proposed in this paper, we obtained an MPEG encoded video source, "Star -Wars", from the anonymous ftp [13]. The test data set represents the encoded rate E(t) by MPEG -1 standard. The source material contains quite a diverse mixture of material ranging from low complexity/motion scenes to those with very high action. The data file consists of 174,138 integers, representing the number of bits per video frame. The original film is coded with 24 frames per second. Thus a frame period corresponds to 1/24 second. The sequence of MPEG I, P and B frames used in a GOP is IBBPBBPBBPBB. The length of the movie is approximately 2 hours.

To test the proposed DBA algorithm for multiple videos, we divide the "Star -Wars" video into 15 individual video sources. This is done by taking first 150,000 frames from the original data and each individual source is generated with 10,000 frames. Table 1 shows the traffic characterization of 15 test video sources. The table represents the traffic statistics for test videos; average cell rate, peak cell rate, peak to average ratio. In the table, a cell includes information of 48 bytes (= 48 * 8 bits). Average cell rate and peak cell rate represent the average and peak value of the number of cells per frame thro ugh total 10,000 frames, respectively. It is shown that the peak to average ratio is ranged from 6.4 to 13.4.

B. Performance Metrics

To test the performance of DBA scheme proposed in this paper, we use two performance metrics; the number of required enco der rate controls and the gap from B^starget from $B^s(t)$. In Step 6 of DBA algorithm, we define the number of encoder rate controls required for the source *s* during the frame period *t* as $ERC^s(t)$. If the encoder rate control is required for the source *s* at the frame period *t*, then $ERC^s(t)$ is 1, otherwise, 0. Thus the average number of encoder rate controls required over frame interval (1, ..., t) is defined as

$$\frac{\sum_{i=1,\dots,t} \sum_{s=1,\dots,n} ERC^s(i)}{t \times n}$$

Similarly, we also define the average gap between B^starget and $B^s(t)$ over frame interval (1, ..., t) as

$$\frac{\sum_{i=1,\dots,t} \sum_{s=1,\dots,n} | B^s(i) - B^s{}_{target} |}{t \times n}$$

The number of encoder rate controls represents a quantitative measure of video quality. Increasing the quantization rate of the MPEG coder at a frame period has an effect of degrading the quality of video. In related studies [10 -12], this measure is represented as the peak signal to noise ratio (PSNR). Thus high PSNR is guaranteed when the number of encoder rate controls is small. Average gap between B^starget and $B^s(t)$ represents the variation of cell delays in the encoder buffer which is a critica l measure of QoS requirements. Generally a small gap is required for reliable video display.

We analyze the effects of DBA scheme for various system parameters such as the ATM switch capacity *C*, buffer size B^smax, LB token generation rate L^s and target buffer occupancy level B^starget. These four parameters are employed as follows:

1) C = VBR peak cell rate * α, where $0 \le \alpha \le 1$, and the VBR peak cell rate represents the peak rate of the aggregate cell stream of 15 video sources;

2) B^smax = peak cell rate of source *s* (in Table 1) * β, where $0 \le \beta \le 1$;

3) L^s = peak cell rate of source *s* (in Table 1) * γ, where $0 \le \gamma \le 1$;

4) B^starget = B^smax * δ, where 0 <= δ <= 1.

C. Experimental Results

Table 2 shows the performance with four different weight assignment policies in case of the same QoS demand level which is explained in Section IV -B. For the problem with 15 test video sou rces, the following coefficients are employed: $\alpha = \beta = \gamma = \delta =$ 0.5. In the table, it seems that the fourth policy of $w^s = E^s(t) + B^s(t\text{-}1)$ is a reasonable choice. By giving priority to sources with higher encoder rate and buffer occupancy, we could reduce t he number of encoder rate controls and the gap between target and actual buffer occupancy. Thus we use the fourth policy for all the experiments to follow.

The performance with different QoS demand levels is shown in Table 3. 15 test video sources are clas sified into 3 groups; 1 to 5 (group 1), 6 to 10 (group 2), and 11 to 15 (group 3). Sources in a group have the same QoS(s) as in the table. Thus total ATM switch capacity *C* is divided into *C*/6, *C*/3, and *C*/2 for each group. Given such a group capacity, $\lambda^{s*}(t)$ is assigned to each of five video sources in a group based on the weight assignment policy using $w^s = E^s(t) + B^s(t\text{-}1)$. From the table, it is clear that the performance of each group largely depends on its QoS level. Also, each source within a group provides almost the same performance.

Now that we have demonstrated the effect of different QoS, we are now interested in the effect of different system parameters. The effects of target buffer, ATM switch capacity, buffer size and leaky bucket token rate are examined as in Figure 4 – 11. In the experiments all video sources are assumed to have the same QoS demand level with weight assignment policy of $w^s = E^s(t) + B^s(t\text{-}1)$.

Figure 4 and 5 show the effect of B^starget for the frame interval from 1 to 10,000. In the experiments, 15 video sources are multiplexed and the ratios of α, β and γ are fixed to 0.5 for all video sources. Three different B^starget's are considered with δ = 0.3, 0.5 and 0.7. In the figures, it is shown that the required number of encoder rate controls and the gap between B^starget and $B^s(t)$ converge to a certain constant level as the frame number increases. From the figures, we see that the larger the B^starget is, the more frequent encoder rate controls are required. The gap between B^starget and $B^s(t)$ is decreased as B^starget increases. These results are consistent with our expectation that large B^starget increases the risk of cell loss in the buffer and induces the degradation of video quality. However, it maintains stable variation of cell delays.

Figure 6 and 7 show the effect of the ATM switch capacity C over various B^starget. In the figures, 15 video sources are multiplexed with the ratios of $\beta = \gamma = 0.5$ for all video sources. Three different ATM switch capacities are tested with α = 0.5, 0.7 and 1.0. Figure 6 demonstrates that larger capacity requires less encoder rate controls. In particular, the required number of encoder rate controls is zero for α = 1.0. From Figure 7, we see that the gap with α = 0.5 is larger than that with α = 0.7. However, the gap with α = 0.7 and 1.0 are not significantly different. From the two figures, it is clear that larger ATM switch capacity guarantees better video quality and smaller cell delay variations.

Figure 8 and 9 show the effect of buffer size B^smax. In the figures, the ratios of $\alpha = \gamma$ = 0.5 are employed for all video sources. Tests are performed for different buffer size with β = 0.3, 0.5 and 0.7. The figures show that larger buffer size requires less encoder rate controls and larger variation of c ell delays.

The effect of LB token generation rate is shown in Figure 10 and 11. In the figures, the ratios of $\alpha = \beta$ = 0.5 are employed for all video sources. The different LB token rates are experimented with γ = 0.3, 0.5 and 0.7. From the figures, it is shown that larger LB token generation rate requires less encoder rate controls and smaller variation of cell delays.

VI. Conclusions

We considered VBR transmission of multiple real-time video sources over ATM networks. Multiple real-time VBR videos are multiplexed into an ATM switch to transmit cells over ATM network. Given ATM switch capacity, the problem is to dynamically allocate the required channel bandwidth for each source such that the encoder buffer occupancy is maintained at a target level. We proposed a mathematical formulation and an algorithm for the dynamic bandwidth allocation. To allocate a suitable bandwidth at a given control period, QoS demand levels and traffic characteristics of the video sources are considered.

To test the performance of the proposed scheme, we used the following metrics: the number of encoder rate controls required and the gap between the target buffer occupancy and the current buffer occupancy at each frame period. Numerical experiments show that the quantitative QoS levels and source traffic characteristics can be used to improve fairness of the bandwidth allocation scheme. It is also shown that the required number of encoder rate controls at each frame period becomes smaller as the target buffer occupancy gets smaller. The required number is inversely proportional to the ATM switch capacity, the buffer size and the leaky bucket token rate. The gap between the target buffer and the current buffer occupancy becomes smaller as the target buffer occupancy gets larger. The gap is inversely proportional to the ATM switch capacity and the leaky bucket token rate. However, it is proportional to the buffer size.

In this paper, it is assumed that the DBA algorithm is performed at each frame period. Thus the calculation of $\lambda^s(t)$ should be completed during the frame interval (t-1, t). This approach may not be realistic from the viewpoint that one frame period is too short to perform the necessary computation, at least under the current technologies. This problem may get even worse when a large number of video streams are multiplexed together. However, the proposed algorithm is not limited to a specific control period for calculation of the channel rate. For example, one GOP interval may be used as the control period instead of one frame interval [14].

This paper also considers only one multiplexing stage at the first ATM switch receiving video traffic. However, in some cases, to ensure the end-to-end performance, all switching stages in the transmission path may be taken into account. In fact, the main focus of this paper is to characterize the behavior of channel rate allocations for a given ATM switching capability. More complicated scenarios on ATM networking capability need to be examined for further study.

References

[1] International Organization for Standardization, Generic Coding of Moving Pictures and Associated Audio, ISO/IEC/JTC1/SC29/WG11, Mar., 1993.

[2] R. O. Onvural, Asynchronous transfer mode networks: performance issues, 2nd edition, Artech House, 1995.

[3] D. P. Heyman and T. V. Lakshman, "Source Models for VBR Broadcast-Video Traffic", IEEE/ACM Transactions on Networking, Vol. 4, No. 1, pp. 40-48, 1996.

[4] D. P. Heyman, A. Tabatabai and T. V. Lakshman, "Statistical Analysis and Simulation Study of Video Teleconference Traffic in ATM Networks", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 2, No. 1, pp. 49-59, 1992.

[5] D. M. Lucantoni, M. F. Neuts and A. R. Reibman, "Methods for Performance Evaluation of VBR Video Traffic Models", IEEE/ACM Transactions on Networking, Vol. 2, No. 2, pp. 176-180, 1994.

[6] D. J. Reininger, D. Raychaudhuri and J. Y. Hui, "Bandwidth Renegotiation for VBR Video over ATM Networks", IEEE Journal on Selected Areas in Communications, Vol. 14, No. 6, pp. 1076-1086, 1996.

[7] P. Pancha and M. E. Zarki, "Bandwidth Allocation Schemes for Variable-Bit-Rate MPEG Sources in ATM Networks", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 3, No. 3, pp. 190-198, 1993.

[8] W. Luo and M. E. Zarki, "Quality Control for VBR Video over ATM Networks", IEEE Journal on Selected Areas in Communication, Vol. 15, No. 6, pp. 1029-1039, 1997.

[9] M. Hamdi, J. W. Roberts and P. Rolin, "Rate Control for VBR Video Coders in Broad-Band Networks", IEEE Journal on Selected Areas in Communication, Vol. 15, No. 6, pp. 1040-1051, 1997.

[10] A. R. Reibman and B. G. Haskell, "Constraints on Variable Bit-Rate Video for ATM Networks", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 2, No. 4, pp. 361-372, 1992.

[11] A. R. Reibman and A. W. Berger, "Traffic Descriptors for VBR Video Teleconferencing Over ATM Networks", IEEE/ACM Transactions on Networking, Vol. 3, No. 3, pp. 329-339, 1995.

[12] C. Y. Hsu, A. Ortega and A. R. Reibman, "Joint Selection of Source and Channel Rate for VBR Video Transmission Under ATM Policing Constraints", IEEE Journal on Selected Areas in Communication, Vol. 15, No. 6, pp. 1016-1028, 1997.

[13] MPEG.data, Trace available via anonymous FTP from ftp.bellcore.com/pub/vbr.video.trace.

[14] C. Y. Lee and S. J. Koh, "A Control of Channel Rate for Real-time VBR Video Transmission", The Korean OR/MS Society, Vol. 24, No. 3, pp. 63 – 72, 1999.

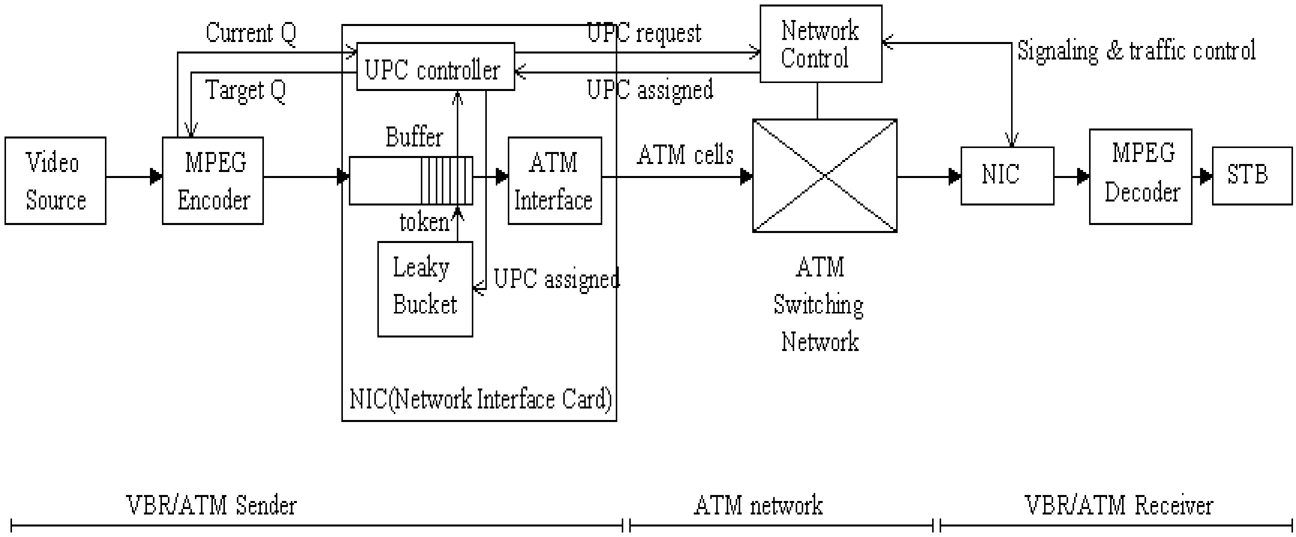


Figure 1. End-to-end model for real-time VBR video transmission over ATM networks

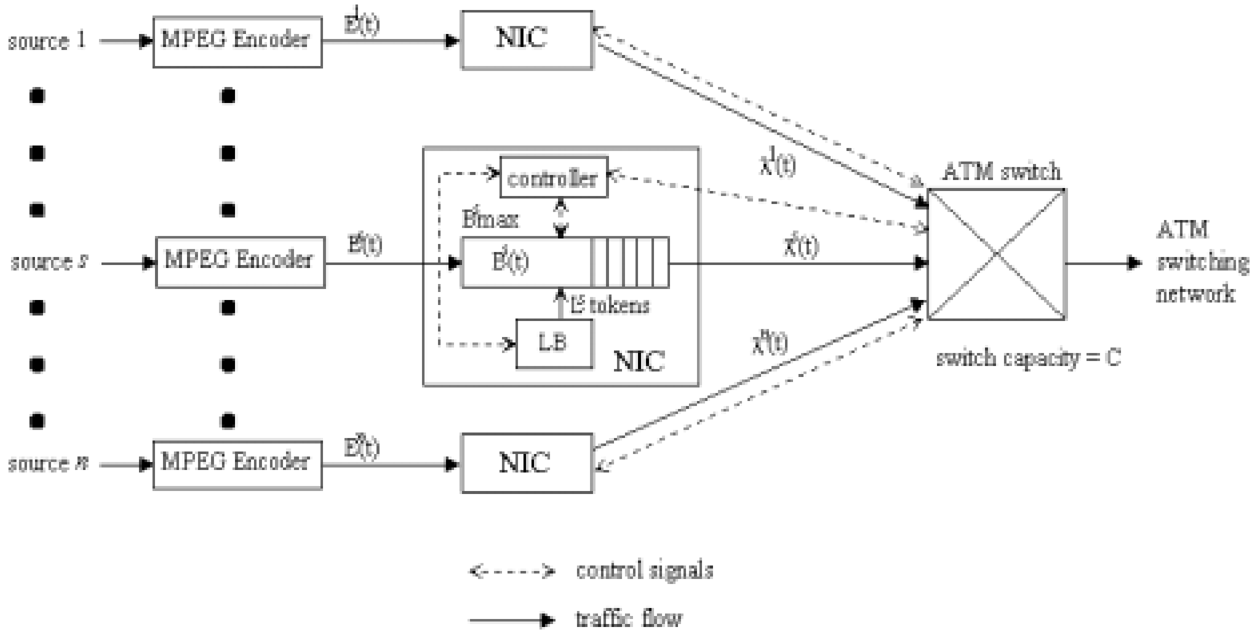


Figure 2. Transmission model for multiple VBR videos over ATM network

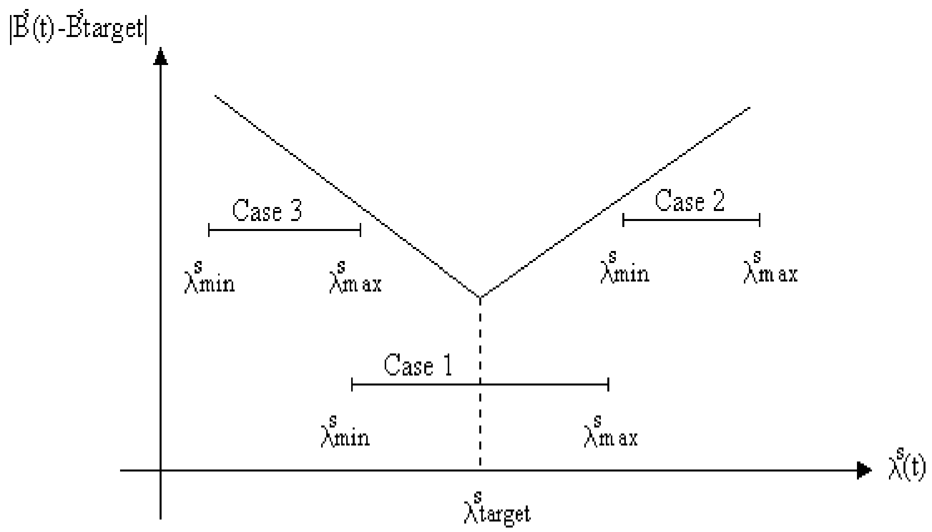


Figure 3. Dynamic bandwidth allocation for an individual VBR video source

Table 1. Traffic Statistics for Test Video Sources

Source Number	Average Cell Rate (cells/frame)	Peak Cell Rate (cells/frame)	Peak to Average Ratio
1	43	421	9.8
2	33	272	8.2
3	41	309	7.5
4	34	290	8.5
5	31	252	8.1
6	31	306	9.9
7	32	254	7.9
8	37	444	12.0
9	36	482	13.4
10	37	256	6.9
11	41	271	6.6
12	44	298	6.8
13	43	276	6.4
14	41	276	6.7
15	45	308	6.8

Each source consists of 10,000 frames.

Table 2. Performance with Different Weight Assignment Policies

	Number of encoder rate control required (per source and frame period)	Gap between target and actual buffer occupancy (cells per source and frame period)
$w^s = 1$	0.11	18.78
$w^s = E^s(t)$	0.07	13.23
$w^s = B^s(t-1)$	0.07	13.43
$w^s = E^s(t) + B^s(t-1)$	0.06	11.74

All sources have the same QoS demand level.

Table 3. Performance with Different QoS Levels

Source Number	QoS level QoS(s)	Allocated Capacity	Number of encoder rate control required	Gap between target and actual buffer occupancy
1	1		0.24	23.78
2	1		0.18	20.76
3	1	$C/6$	0.21	21.23
4	1		0.19	20.93
5	1		0.17	19.89
6	2		0.06	12.32
7	2		0.05	10.54
8	2	$C/3$	0.06	11.38
9	2		0.06	12.32
10	2		0.06	11.28
11	3		0.01	3.36
12	3		0.02	4.51
13	3	$C/2$	0.02	3.89
14	3		0.01	2.96
15	3		0.02	4.72

C represents total ATM switch capacity.

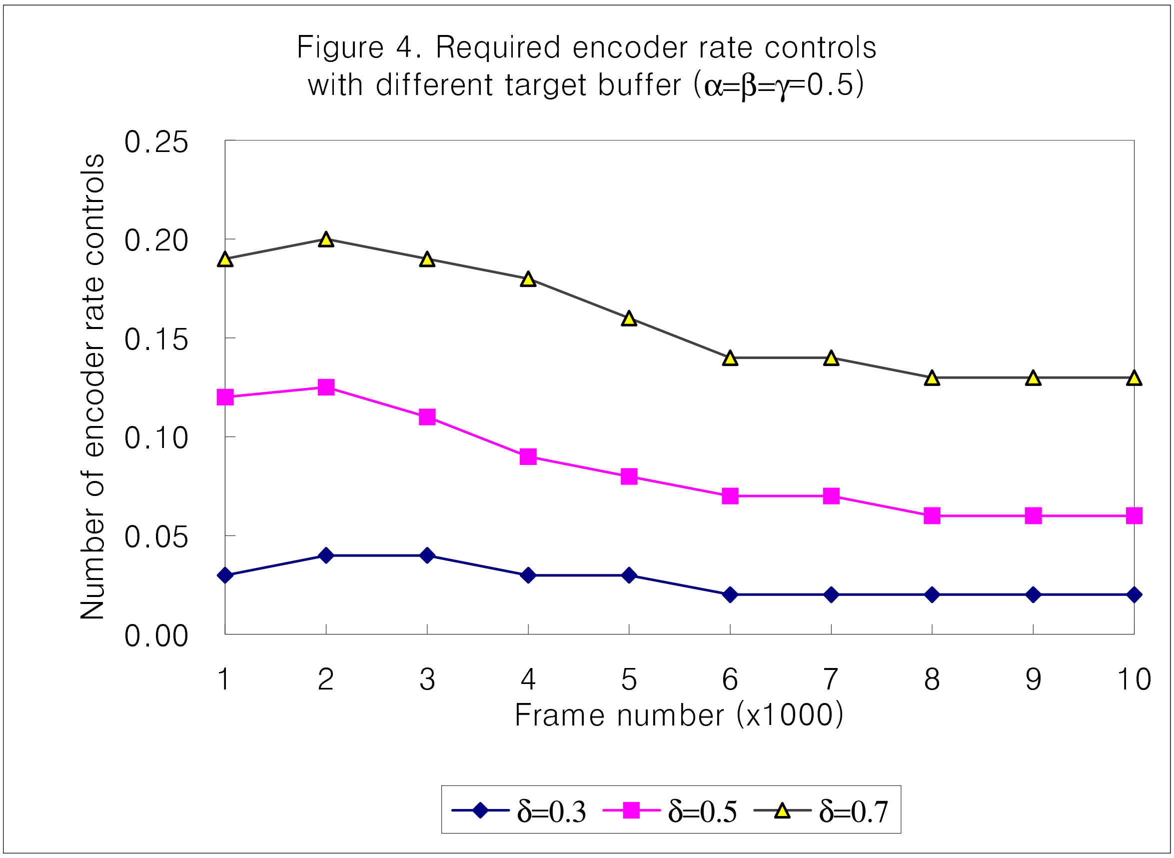


Figure 4. Required encoder rate controls
with different target buffer (α=β=γ=0.5)
Number of encoder rate controls
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9
10
Frame number (x1000)
δ=0.3
δ=0.5
δ=0.7

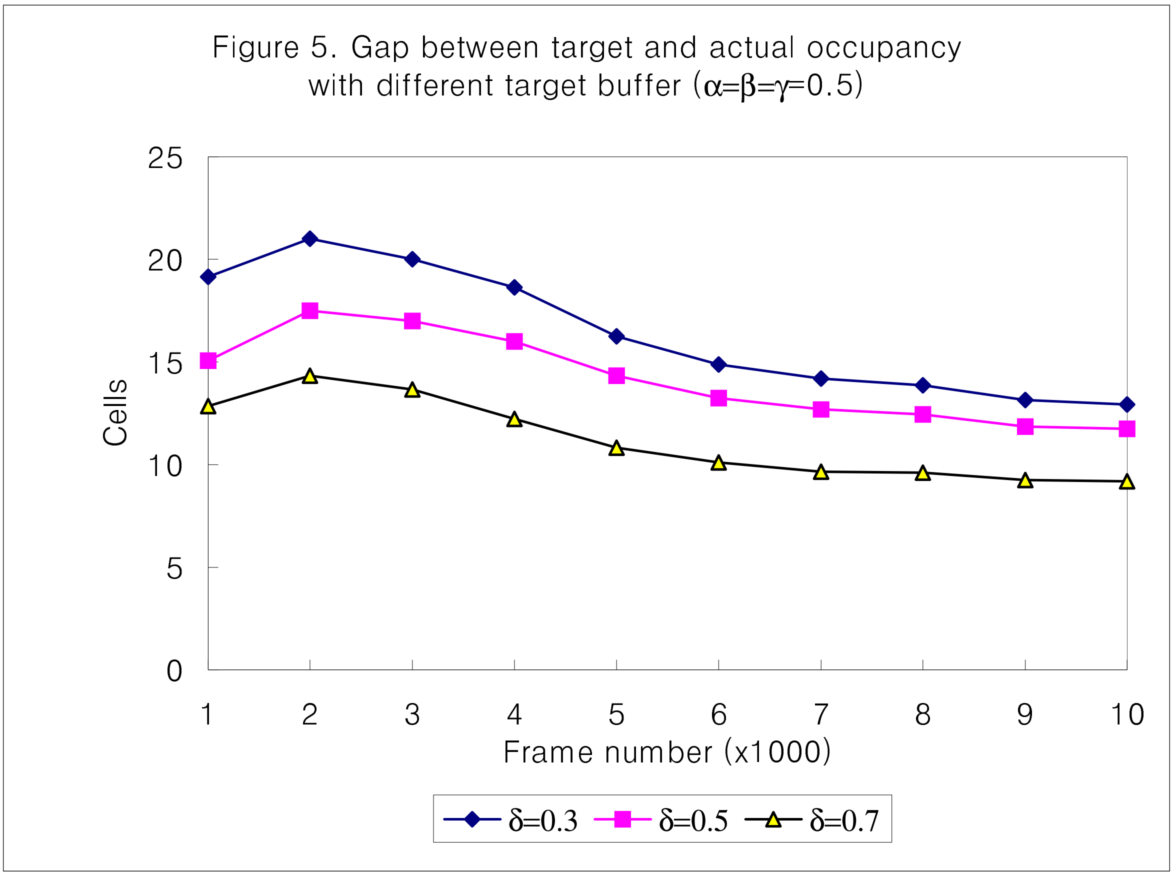


Figure 5. Gap between target and actual occupancy
with different target buffer (α=β=γ=0.5)
25
20
15
10
5
0
Cells
1
2
3
4
5
6
7
8
9
10
Frame number (x1000)
δ=0.3
δ=0.5
δ=0.7

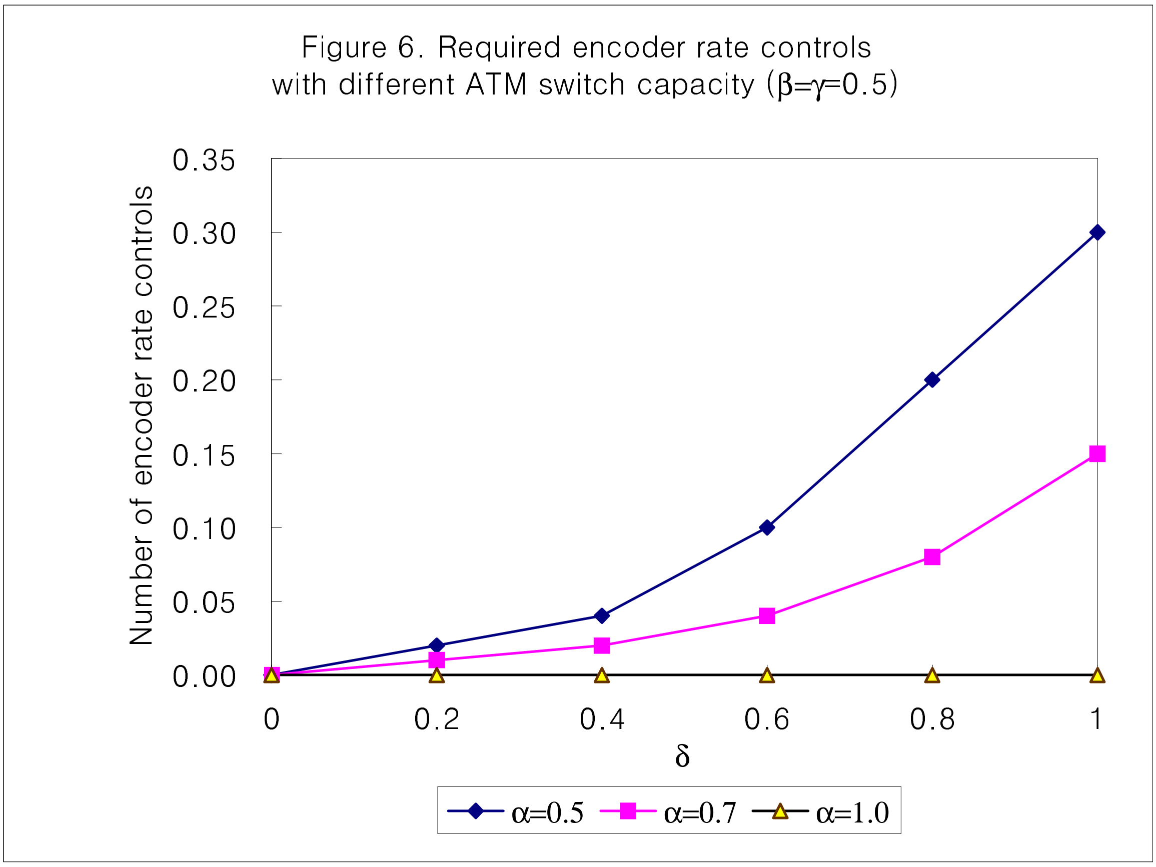


Figure 6. Required encoder rate controls
with different ATM switch capacity (β=γ=0.5)
Number of encoder rate controls
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
0.2
0.4
0.6
0.8
1
δ
α=0.5
α=0.7
α=1.0

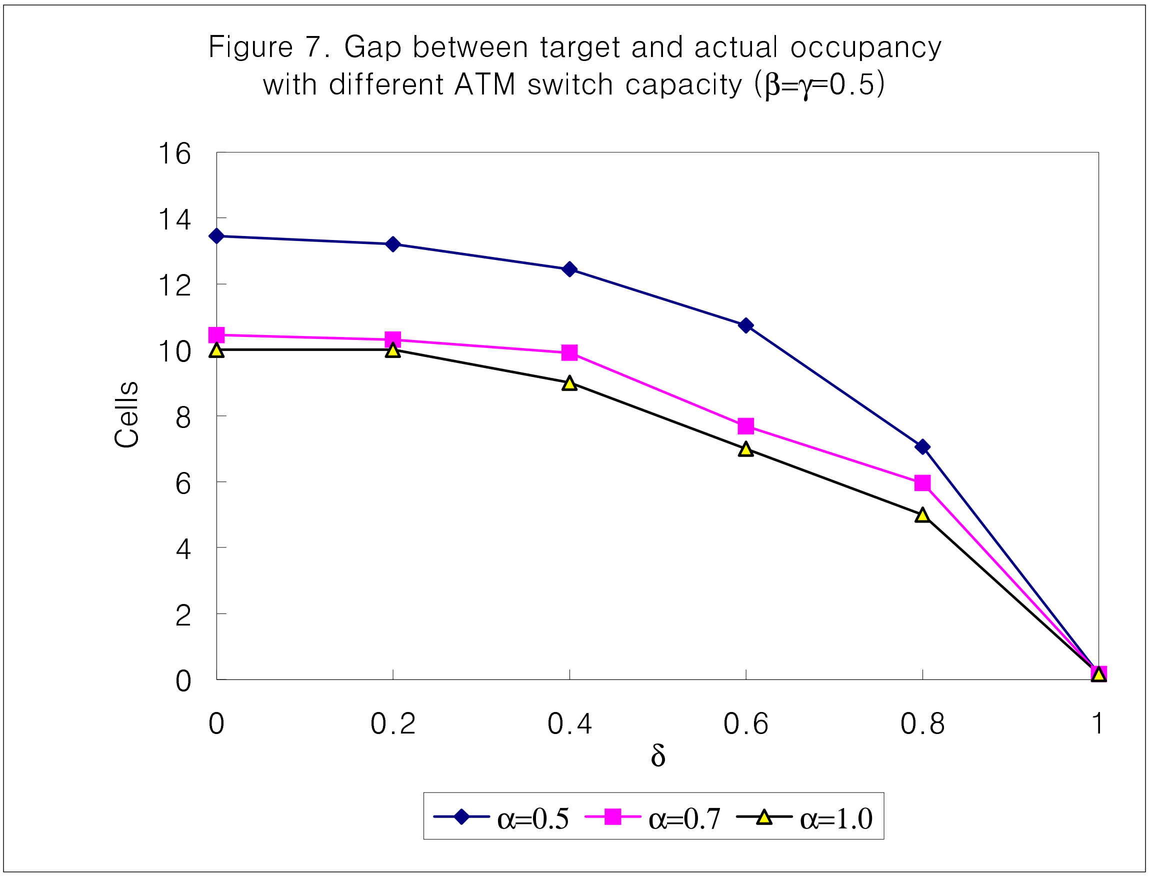


Figure 7. Gap between target and actual occupancy
with different ATM switch capacity (β=γ=0.5)
16
14
12
10
8
6
4
2
0
Cells
0
0.2
0.4
0.6
0.8
1
δ
α=0.5
α=0.7
α=1.0

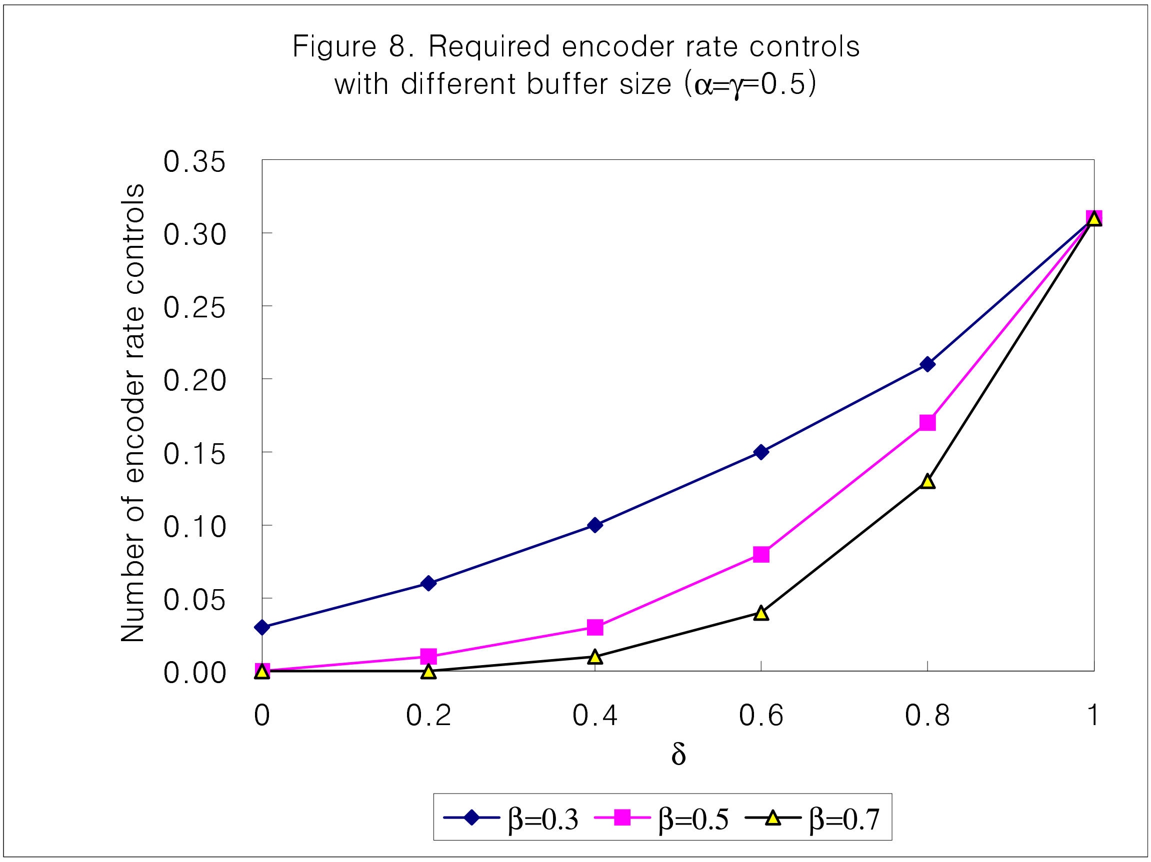
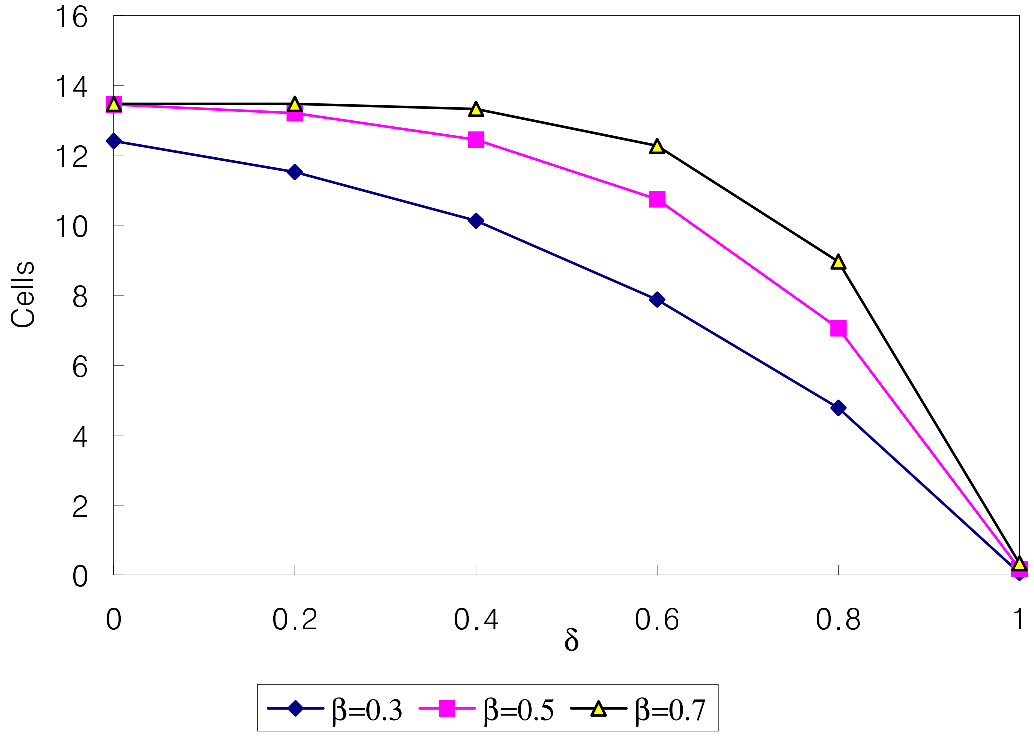


Figure 8. Required encoder rate controls
with different buffer size (α=γ=0.5)
Number of encoder rate controls
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
0.2
0.4
0.6
0.8
1
δ
β=0.3
β=0.5
β=0.7

Figure 9. Gap between target and actual occupancy with different buffer size (α=γ=0.5)



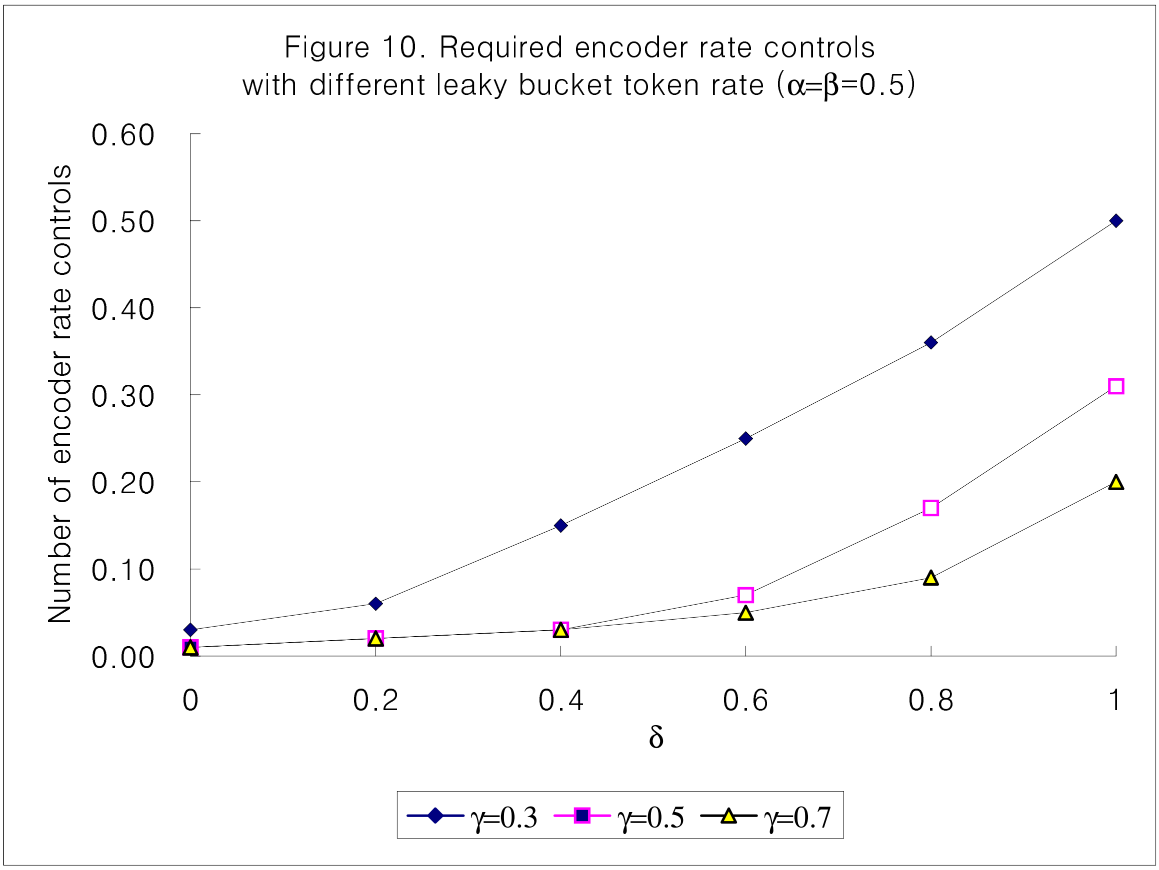


Figure 10. Required encoder rate controls
with different leaky bucket token rate (α=β=0.5)
Number of encoder rate controls
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
0.2
0.4
0.6
0.8
1
δ
γ=0.3
γ=0.5
γ=0.7

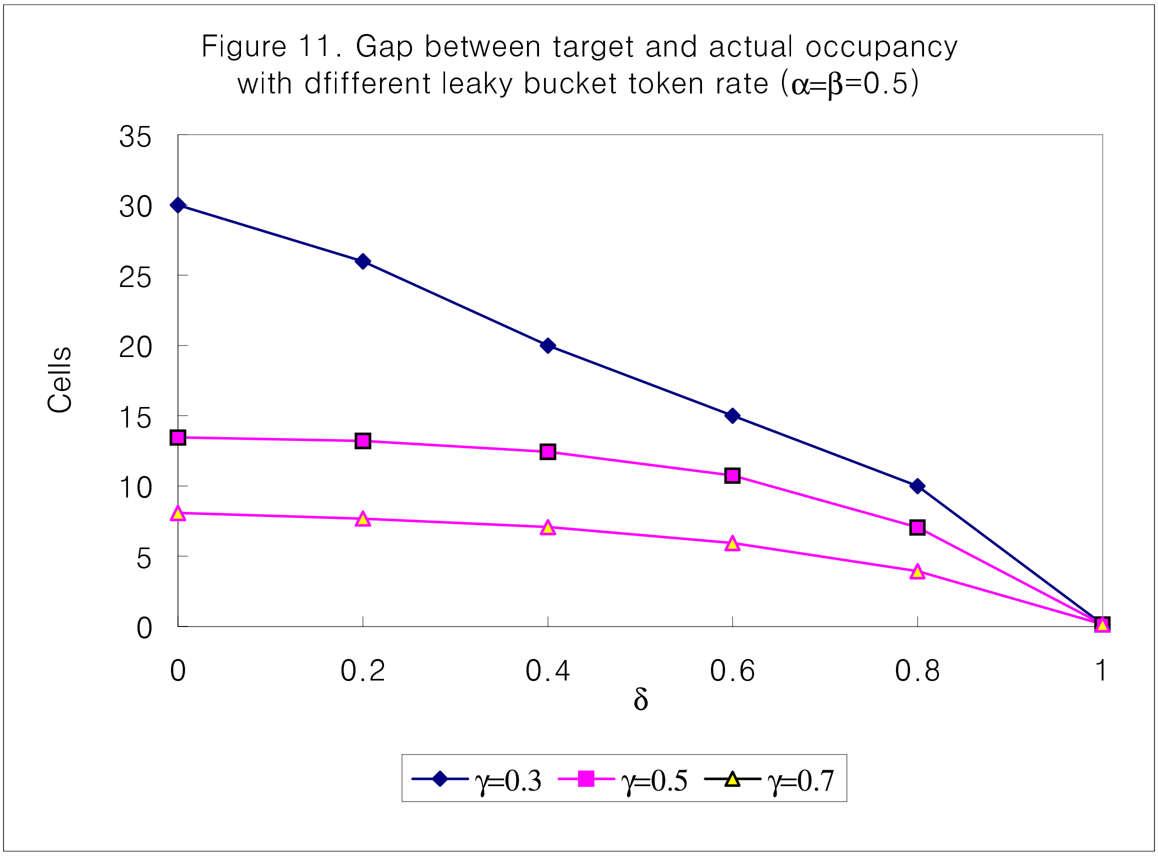


Figure 11. Gap between target and actual occupancy
with dfifferent leaky bucket token rate (α=β=0.5)
Cells
35
30
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
δ
γ=0.3
γ=0.5
γ=0.7