Chapter VI

Optimal Control of Variable Rate Coding with Incomplete Observation in ATM Networks

This chapter investigated the design of dynamic voice coding rate control with doubly stochastic cell arrival processes to reduce traffic congestion in an ATM network. Incomplete observation is assumed. A control maximizes the average coding rate of voice cells subject to an average voice/data cell waiting time constraint on a finite horizon with incomplete observation. The optimal control has a simple structure and is a randomization of two feedbacks with input saturation. The general separation principle for the optimal control with incomplete observation, that is, the sufficiency of the posterior distribution for the optimal control, is derived. Since the complexity of the analytical results prevents us from obtaining an explicit control when observation is incomplete, an adaptive suboptimal control is proposed. This is also a random modification of two feedbacks with saturation and is easily implementable. The effectiveness of the suboptimal control is verified by simulations.

The result of this paper is applicable to networks including video traffic. [Saito 90b]

1. Introduction

Recently, there has been considerable interest in integrated communication networks. In particular, the development of ATM based communication networks is widely recognized as a significant step towards achieving a broadband ISDN.

ATM networks require a new congestion control scheme for voice cells. One such scheme that has recently received attention is variable rate coding. This decreases the bit rate of voice coding during an overload [Listanti 83, Bially 80a, Holtzman 85, Seguel 82, Forst 86, Fredericks 86, Gafni 84, Goodman 80].

Variable rate voice coding can be achieved in a variety of ways [Bially 80a]. One method called embedded coding [Bially 80a, Goodman 80] is the following: Voice signals from off-hook users are sampled at regular intervals, typically 125 μsec. Each sample is quantized by, for example, 4 bits. During silence periods of talkers, samples are discarded. A specified number of samples during talkspurt are grouped together, packetized and transmitted. When cells are constructed, the i-th bit of all samples are grouped to make up a cell (packet interleave method). Thus, four cells are constructed simultaneously, when a sample is quantized by 4 bits. When a network is congested, the bits of signals are dropped by discarding cells at the vocoders or multiplexers at the network entry point and at ATM nodes within the network. The cell consisting of the least significant bits is discarded first. Cells are discarded in ascending order of significance.

Traditional voice flow control mechanisms discard voice cells, which contain whole bits of a sampled signal, and none of the information of the sampled voice signal contained in

the discarded cells is transmitted. By contrast, variable rate coding dynamically trades off voice quality and congestion by reducing the voice coding bit rate at the point of congestion or the point of entry. When a network is congested, only higher bits of a sampled voice signal are transmitted, but many cells with less significant bits are discarded under this control. The cell loss probability to satisfy the specified quality standards is less than 1% under traditional voice flow control, and 5% cell loss probability is admissible under the embedded coding scheme. This is why many systems employ embedded coding [Muise 86].

The remaining problem in the congestion control using embedded coding is how to feedback congestion information in order to drop cells. Previous works assumed that cells are dropped if the queue length or the number of active talkers exceeds a specified level [Yin 87]. Control with a queue length threshold is attractive, because it is easy to implement. If the group size of cells constructed simultaneously is two, that is, the least significant cell and the most significant cell, it is the optimal solution in [Saito 89a]. [Saito 89a] analyzed the optimal control which maximizes the average voice coding rate subject to an average queue length constraint on an infinite horizon with uncorrelated arrival processes. If the admissible control class is restricted to the class of randomized stationary controls employing the total number of waiting voice/data cells in the system and the number of cells arriving in the current slot, then the optimal control is randomized feedback with input saturation. In practice, since the voice arrival process is correlated when the number of multiplexed voice sources is small, this result must be extended. However, the method in [Saito 89a] cannot be extended to a multidimensional Markov chain and cannot be applied for a correlated arrival process. Additionally, an optimal control derived under the assumption of the correlation of arrival processes must be compared with optimum derived under the renewal assumption.

This chapter aims to derive a control scheme which makes the voice quality as good as possible while reducing congestion. It is known that voice quality deteriorates as the coding rate decreases [Heffestad 82], though the effect of bit rate on voice coding is quite complicated [Cox 80, Holtzman 85]. Futhermore, it is known [Saito 89a] that maximization of the average coding rate means maximization of cell-wise signal-to-noise ratio and minimization of cell loss probability. Thus the average voice coding rate is used as the criterion for voice quality [Bially 80b, Muise 86]. A Markov chain provides the voice arrival process in this chapter. The states of the chain at t specify the probabilities for all numbers of voice cells arriving at t. This model for an arrival process reflects the facts: It is known that during the talkspurt, individual voice sources generate cells and that the number of active talkers at t determines the arrival rate of voice cells at t. Time evolution of the number of active talkers can be modelled by a Markov chain [Bially 80b, Heffes 86].

This chapter is organized as follows. In Section 2, notation is introduced and the problem is formulated. In Section 3, optimality equations are analyzed. A Lagrange multiplier is introduced and the optimization problem with a constraint is reduced to one without a constraint. It is shown that the optimal control of the Lagrangian is a feedback with input saturation [Vakil 87]. The general separation principle, that is, that the sufficient statistics for the optimal control are the posterior distribution of the phases of arrival processes, is shown. Unfortunately, the optimal control with incomplete observation cannot be implemented easily. Thus, Section 4 proposes an implementable suboptimal

control, that is, feedback with saturation, and its effectiveness is verified by simulations in Section 5. Section 6 presents some conclusion.

2. Problem formulation

Consider an ATM multiplexer (or ATM nodes) and assume that cells arrive at the multiplexer at the beginnings of slots and leave at the ends of slots. Voice cells are assumed to arrive in a batch, since packet interleaving is used and several voice cells are constructed simultaneously. A voice or data cell is assumed to have a fixed length [Heffes 86, Muise 86, Saito 89a]. It is assumed that a slot has a fixed duration equal to the transmission time of a cell. A cell is transmitted in a slot.

The arrival processes of voice/data cells are assumed to be two independent processes. Data cells are assumed to arrive in a renewal process. Let $\lambda_d(l)$ be the probabilty that $A_d(t)$, the number of data cells arriving at the beginning of the t-th slot, is l. The states of a Markov chain at t provides the voice cell arrival process. Let $\lambda_v(t;k)$ be the probability that $A_v(t)$, the number of voice cell batches arriving at the beginning of the t-th slot, is k. When the chain z_v, which denotes the number of talkers in talkspurt and is called the phase of the voice cell arrival process, is in state i ($i = 1, \ldots, K_v$) at t, then $\lambda_v(t;k) = \lambda_v(k|i)$. Denote the probability of the transition from state i to state j in the chain as $r_v(i,j)$.

The size of an arriving voice cell batch is $\tilde{N}_{max}$, that is, a voice cell consists of $\tilde{N}_{max}$ cells when it arrives. After bit dropping, the size of an accepted voice cell batch is $\tilde{N}_v$, where $0 < \tilde{N}_{min} \leq \tilde{N}_v \leq \tilde{N}_{max}$. Here, $\tilde{N}_v$ is selected as a control parameter. Let $N_v(t)$ denote the total accepted voice cells at t. Then, $0 \leq N_{min} \leq N_v \leq N_{max}$, where $N_{min} = A_v(t)\tilde{N}_{min}, N_{max} = A_v(t)\tilde{N}_{max}$. Data cells are assumed not to be discarded.

Thus, $N_d(t)$, the total number of accepted data cells, is equal to $A_d(t)$.

Let $z_v(t)$ and $z_d(t)$ denote the number of waiting voice and data cells in the system just before the arrival of cells at the t-th slot. (Figure 1.)

To determine $N_v(t)$, the total number of accepted cells at t, we can use the information I_t, $I_t \triangleq \{I_1, z_v(1), \ldots, z_v(t), z_d(1), \ldots, z_d(t), A_v(1), \ldots, A_v(t), A_d(1), \ldots, A_d(t)\}$, where I_1=the initial distributions of the phases of the arrival rates.

The control scheme which makes the voice quality (average coding rate) as good as possible while reducing congestion (average waiting time of cells) is the solution of the following problem, under the assumption that a priority is followed for voice cells over data cells as in [Vakil 87]. The criterion is the average coding rate of voice cells arriving during $[0, T]$. The constraint is the average of the total waiting times of voice/data cells during $[0, T]$.

$$\max_{N_v(t)} \quad J = E[\sum_{t=1}^{T} N_v(t)|I_1] \tag{2.1}$$

$$\text{subj. to} \quad C = E[\sum_{t=1}^{T} W(t)|I_1] \leq c \tag{2.2}$$

Here, $W(t)$ is the total number of cells at t. That is,

$$W(t) = [z_v(t) + z_d(t) + N_v(t) + N_d(t) - 1]^+. \tag{2.3}$$

The process of the number of voice cells in the system, $z_v(t)$, follows

$$z_v(t+1) = [z_v(t) + N_v(t) - 1]^+ \tag{2.4}$$

and the process of the number of data cells in the system, $z_d(t)$, follows

$$z_d(t+1) = z_d(t) + N_d(t) - \mathbf{1}(z_v(t) + N_v(t) = 0, z_d(t) + N_d(t) > 0). \tag{2.5}$$

Here, $\mathbf{1}(\cdot)$ denotes an indicator function and

$$[x]^+ = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

3. Optimality equations

In actual systems, observation of the number of users in talkspurt requires too much work for the controller and the control with its observation is not implementable in practice. Thus, z_v, the phase of the voice arrival process, is assumed to be unobserved. Conversely, number of arriving voice cell groups $A_v(t)$ is easy to measure, and the number of active talkers $z_v(t)$ is assumed to be estimated from the measurement of $A_v(t)$.

For a fixed γ, a Lagrangian [Nain 86a,b, Tijms 86] is defined as,

$$J_\gamma = \sum_{t=1}^{T} E[N_v(t) - \gamma W(t)|I_1], \tag{3.1}$$

and we consider the problem

$$\max_{Nv(t)} J_\gamma. \tag{3.2}$$

Define $V_t(z_v(t), z_d(t), A_v(t), A_d(t), I_t)$, the maximal expected reward during $[t, T]$. $V_t(z_v(t), z_d(t), A_v(t), A_d(t), I_t)$ depends on I_t only through $A_v(t)$, and distribution of $A_v(t)$ given I_t is equivalent to that given $\hat{z}_v(t)$, the posterior distributions of $z_v(t)$ given I_t. Thus,

$$V_t(z_v(t), z_d(t), A_v(t), A_d(t), I_t) = V_t(z_v(t), z_d(t), A_v(t), A_d(t), \hat{z}_v(t)).$$

Here, $\hat{z}_v(t)$ can be evaluated using the observation $A_v(t)$;

$$\hat{z}_v(t) = (\hat{z}_v(t; 1), ..., \hat{z}_v(t; K_v)), \tag{3.3}$$

$$\hat{z}_v(t; i) = \Pr[z_v(t) = i|I_t],$$
$$= \frac{\sum_j \hat{z}_v(t-1; j) r_v(j, i) \lambda_v(A_v(t)|i)}{\sum_{j,k} \hat{z}_v(t-1; j) r_v(j, k) \lambda_v(A_v(t)|k)}. \tag{3.4}$$

Then, the optimal reward function is well defined by the recursive optimality equations [Howard 71],

$$V_t(z_v(t), z_d(t), A_v(t), A_d(t), \hat{z}_v(t))$$

$$= \max_{Nv(t)} \{N_v(t) - \gamma W(t) + \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l)$$

$$V_{t+1}(z_v(t+1), z_d(t+1), k, l, \hat{z}_v(t+1))\}, \tag{3.5a}$$

for $t = 1, ..., T$ and

$$V_{T+1} = 0. \tag{3.5b}$$

Here,

$$\hat{\lambda}_v(t+1; k) = \Pr[A_v(t) = k|I_t] = \sum_{i,j} \hat{z}_v(t; i) r_v(i, j) \lambda_v(k|j), \tag{3.6a}$$

$$W(t) = [z_v(t) + z_d(t) + N_v(t) + N_d(t) - 1]^+, \tag{3.6b}$$

$$z_v(t+1) = [z_v(t) + N_v(t) - 1]^+, \tag{3.6c}$$

$$z_d(t+1) = z_d(t) + N_d(t) - \mathbf{1}(z_v(t) + N_v(t) = 0, z_d(t) + N_d(t) > 0), \tag{3.6d}$$

$$N_d(t) = A_d(t). \tag{3.6e}$$

We first prove the following theorem.

Theorem 1.

For all $i, j, m, n, \hat{z}_v, t$,

$$V_t(i+1, j, m, n, \hat{z}_v(t)) = V_t(i, j+1, m, n, \hat{z}_v(t)). \tag{3.7}$$

Proof: The proof is by induction on t. Suppose that Eq. (3.7) holds true for $t+1, t+$

$2, ..., T + 1$. Then,

$$V_t(i + 1, j, m, n, \hat{z}_v(t))$$

$$= \max \quad \{N_v(t) - \gamma[N_v(t) + i + j + N_d(t)]^+$$

$$+ \sum_{k,l} \hat{\lambda}_v(t + 1; k)\lambda_d(l)V_{t+1}(i + N_v(t), j + N_d(t), k, l, \hat{z}_v(t + 1))\},$$

$$V_t(i, j + 1, m, n, \hat{z}_v(t))$$

$$= \max \quad \{N_v(t) - \gamma[N_v(t) + i + j + N_d(t)]^+$$

$$+ \sum_{k,l} \hat{\lambda}_v(t + 1; k)\lambda_d(l)$$

$$V_{t+1}([i + N_v(t) - 1]^+, j + N_d(t) + 1 - \mathbf{1}(i + N_v(t) = 0), k, l, \hat{z}_v(t + 1))\}.$$

From the assumption,

$$V_{t+1}(i + N_v(t), j + N_d(t), k, l, \hat{z}_v(t + 1))$$

$$= V_{t+1}([i + N_v(t) - 1]^+, j + N_d(t) + 1 - \mathbf{1}(i + N_v(t) = 0), k, l, \hat{z}_v(t + 1)).$$

Therefore,

$$V_t(i + 1, j, m, n, \hat{z}_v(t)) = V_t(i, j + 1, m, n, \hat{z}_v(t)). \quad \blacksquare$$

Theorem 1 shows that there is no need to distinguish voice cells in the system from data cells in the system, and we can eliminate one dimension from the optimality equation (3.5). Set $z(t) = z_v(t) + z_d(t)$. The optimal reward function $U_t(z(t), A_v(t), A_d(t), \hat{z}_v(t))$ during $[t, T]$ is defined as

$$U_t(z(t), A_v(t), A_d(t), \hat{z}_v(t))$$

$$= \max \quad \{N_v(t) - \gamma W(t) + \sum_{k,l} \hat{\lambda}_v(t + 1; k)\lambda_d(l)U_{t+1}(z(t + 1), k, l, \hat{z}_v(t + 1))\}, \quad (3.8a)$$

for $t = 1, \dots, T$ and

$$U_{T+1} = 0, \quad (3.8b)$$

$$W(t) = [z(t) + N_v(t) + N_d(t) - 1]^+, \quad (3.9a)$$

$$z(t + 1) = [z(t) + N_v(t) + N_d(t) - 1]^+. \quad (3.9b)$$

Equations (3.8a)-(3.8b) correspond to (3.5a)-(3.5b), and (3.9a) corresponds to (3.6b).

The solution of the optimality equation (3.8) is discussed in the Appendix, which shows that the maximum expected reward is a decreasing and concave function of the number of cells in the system (Theorem 2, Theorem 3), and that the optimal control of the Lagrangian is monotonic, a feedback control with saturation (Theorem 4). The optimal control maximizing Eq. (2.1) with a constraint (2.2) is a randomization of two optimal controls of the Lagrangians. As a result, the optimal control of the average coding rate constrained by the average waiting time with complete information has a simple structure: randomization of two deterministic controls which are the feedback control with saturation (Theorem 5, 6 and 7).

That is, when $z(t) = n, A_v(t) = i, A_d(t) = j, \hat{z}_v(t) = \mathbf{k}$ at t, the optimal control is the control which accepts $N_v^*(t; n, i, j, \mathbf{k}; \gamma)$ voice cells or $N_v^*(t; n, i, j, \mathbf{k}; \gamma - 0)$ voice cells at t with probability r or $1 - r$. Here,

$$N_v^*(t; n, i, j, \mathbf{k}; \gamma) = [K_\gamma^*(t; \mathbf{k}) - n - j]_{N_{min}}^{N_{max}}$$

and $K_\gamma^*(t; \mathbf{k})$ is provided by the optimality equation (3.8) and $y = [x]_b^a$ is a saturation function defined in (A.15).

The generalized separation principle is valid in the following sense. The optimality equation (3.8) depends only on $\hat{z}_v$, which $\hat{z}_v$ are sufficient statistics for the optimal control. We refer to this property as the generalized separation principle following [Segall 77, Vakil 87].

4. Suboptimal control

The previous argument highlights the structure of the optimal control. However, since the computation of the reward function U using (3.8) generally appears to be prohibitive because of the dependence on $\hat{z}_v(t)$, the explicit solution of the optimal control with incomplete observation cannot be determined in practice from the above results. Hence, it is necessary to consider an implementable suboptimal control for incomplete observation.

By replacing $\hat{z}_v(t)$ into $z_v(t)$, we can prove that all the result mentioned above are also valid for complete observation. Here, an implementable suboptimal control for incomplete observation is proposed based on the optimal control with complete observation. The optimal control with complete observation uses $N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t); \gamma)$ or $N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t); \gamma - 0)$ with probability r and $1 - r$. The optimal control with incomplete observation is a randomized modification of $N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), \hat{z}_v(t); \gamma)$ and $N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), \hat{z}_v(t); \gamma - 0)$, and $\hat{z}_v$ are the sufficient statistics for the optimal control. Thus, it is natural to consider the following class of suboptimal controls as in [Vakil 87]. It employs

$$\begin{cases} N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t) = i; \gamma), & \text{with probability } r\hat{z}_v(t; i), \\ N_v^*(t; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t) = i, ; \gamma - 0), & \text{with probability } (1 - r)\hat{z}_v(t; i), \end{cases} \quad (4.1)$$

for $1 \leq i \leq K_v$. To provide control by Eq. (4.1), a controller must hold the data of $\{K_\gamma(t; i), t = 0, \ldots, T\}$. To decrease the complexity of control, for a certain fixed $\tau \in [0, T]$, we use $\{K_\gamma(\tau; i)\}$ instead of $\{K_\gamma(t; i), t = 0, \ldots, T\}$. In other words,

$$\begin{cases} N_v^*(\tau; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t) = i; \gamma), & \text{with probability } r\hat{z}_v(t; i), \\ N_v^*(\tau; \boldsymbol{z}(t), A_v(t), A_d(t), z_v(t) = i, ; \gamma - 0), & \text{with probability } (1 - r)\hat{z}_v(t; i). \end{cases} \quad (4.2)$$

Futhermore, employing the method in [Heffes 86], we can reduce the dimension of voice arrival phase to 2 ($K_v = 2$). As a result, control is possible with only $K_\gamma(\tau; 1), K_\gamma(\tau; 2)$.

There are two methods for determining $K_\gamma(\tau; l)$ or equivalently $N_v^*(\tau; i, j, k, l)$. One method is to solve Eq. (3.8) for complete observation with an approximation that the maximum number of cells in the system is finite L.

$$U_t(\boldsymbol{z}(t), A_v(t), A_d(t), z_v(t))$$
$$= \max \quad \{ N_v(t) - \gamma W(t) $$
$$+ \sum_{i, j, z_v(t+1)} r_v(z_v(t), z_v(t+1))\lambda_v(i|z_v(t+1))U_{t+1}(\boldsymbol{z}(t+1), i, j, z_v(t+1))\}, \quad (4.3a)$$

for $t = 1, \ldots, T$, $0 \leq \boldsymbol{z}(t) \leq L$ and

$$U_{T+1} = 0, \quad (4.3b)$$

$$W(t) = [\boldsymbol{z}(t) + N_v(t) + N_d(t) - 1]^+, \quad (4.4a)$$

$$\boldsymbol{z}(t+1) = [\boldsymbol{z}(t) + N_v(t) + N_d(t) - 1]_0^L. \quad (4.5b)$$

The other method is to use the optimal control with an infinite horizon. For an infinite horizon, the performance measures under the optimal control are directly derived in the following way, because the structure of an optimal control is given. We can omit t or τ for an infinite horizon.

If we assume $\{K_\gamma(i)\}$, then $\lambda(n|i, k)$, the probability that the number of accepted voice and data cells is n when the number of cells in the system is i and the voice arrival phase is k, can be determined. Let $p(i, j)$ be the stationary probability that the number of cells in the system is i and the voice arrival phase is j. Thus, for $i \geq 1$,

$$p(i, j) = \sum_k r_v(k, j)\{\sum_{l=0}^{i+1} \lambda(l|i + 1 - l, k)p(i + 1 - l, k)\}, \quad (4.6a)$$

and

$$p(0,j) = \sum_k \tau_v(k,j)\{\sum_{l=0}^{1} \lambda(l|1-l,k)p(1-l,k) + \lambda(0|0,k)p(0,k)\}. \qquad (4.6b)$$

Set $K = max_i\{K_\gamma(i)\}$. Define $\lambda(n|k)$ as the probability that the number of accepted voice and data cells is n when the size of accepted voice cell group $\tilde{N}_v$ is $\tilde{N}_{min}$ and that the voice arrival phase is k. For $i > K$, Eq. (4.6) is

$$p(i,j) = \sum_k \tau_v(k,j)\{\sum_{l=0}^{i+1} \lambda(l|k)p(i+1-l,k)\}. \qquad (4.7)$$

Define $G_j(z) = \sum_{i=K+1}^{\infty} p(i,j)z^i$. From Eq. (4.7),

$$zG_j(z) = \sum_k \tau_v(k,j)\{\Lambda_k(z)G_k(z) + \Lambda_k(z)\sum_{l=0}^{K} p(l,k)z^l$$
$$- \sum_{l=0}^{K+1}\sum_{m=0}^{K+1-l} \lambda(m|k)p(l,k)z^{l+m}\}. \qquad (4.8)$$

Here,

$$\Lambda_k(z) \triangleq \sum_{n=0}^{\infty} \lambda(n|k)z^n = D(z)\Lambda_v(z^{\tilde{N}_{min}}|k), \qquad (4.9)$$

$$D(z) \triangleq \sum_{l=0}^{\infty} \lambda_d(l)z^l, \qquad (4.10)$$

$$\Lambda_v(z|j) \triangleq \sum_{k=0}^{\infty} \lambda_v(k|j)z^k. \qquad (4.11)$$

Let π_i be the stationary probability that the voice arrival phase is at i, and

$$\pi_1 = \tau_v(2,1)/(\tau_v(1,2) + \tau_v(2,1)), \pi_2 = \tau_v(1,2)/(\tau_v(1,2) + \tau_v(2,1)). \qquad (4.12)$$

Then,

$$G_j(1) + \sum_{l=0}^{K} p(l,j) = \pi_j. \qquad (4.13)$$

The performance measures such as the average coding rate J, the average waiting time C and the Lagrangian J_γ can be obtained from Eq. (4.6) for $0 \le i \le K$, Eq. (4.8) and Eq. (4.13). Consequently, the optimal $\{K_\gamma(i)\}$ can be obtained.

5. Performance of suboptimal control

This section examines the performance of the suboptimal control proposed in the previous section. The effectiveness of the suboptimal control is verified by simulation. Suboptimal control with incomplete observation, optimal control with complete observation and fixed rate control ($\tilde{N}_v$ is constant irrespective of congestion) are compared for the same sample path of cell arrivals.

In the following numerical examples, we consider a multiplexer with a 1.5Mbits/s transmission line. We assume that the cell length is 64 bytes, and that the talkspurts and silence periods are exponentially distributed with means 352 ms and 650 ms [Sriram 86, Heffes 86]. The voice signal in a talkspurt is sampled every 125 μsec and quantized by 4 bits [Dravide 89]. Thus, every 64 ms, four cells are generated. We assume that $\tilde{N}_{min} = 2$, and $\tilde{N}_{max} = 4$ [Dravide 89].

Two cases are considered: 100 off-hook users added to data traffic (Case 1), and 50 off-hook users added to data traffic (Case 2). Offered voice and data traffic intensity totals 0.85 in both cases. Using the method in [Heffes 86], we obtain the continuous-time Markov modulated Poisson process model ($K_v = 2$) for the voice arrival process. The holding time of a server, that is, the transmission time of a cell, is quite small in our model. Thus we neglect the probabilities that there is more than one transition in a voice arrival process during a slot, that more than one voice cell arrives during a slot, and that more than one data cells arrive during a slot. Thus we obtain a discrete-time model.

For Case 1: $\lambda_v(0|1) = 0.78, \lambda_v(1|1) = 0.22, \lambda_v(0|2) = 0.84, \lambda_v(1|2) = 0.16, \lambda_d(0) = 0.90, \lambda_d(1)=0.10$ and $\lambda_v(l|k) = 0, \lambda_d(l) = 0$ for $l \ge 2, k = 1,2$. The transition matrix of

voice arrival process $(r_v(i,j))_{i,j}$ is $\begin{pmatrix} 1 - 0.66 \times 10^{-3} & 0.66 \times 10^{-3} \\ 0.66 \times 10^{-3} & 1 - 0.66 \times 10^{-3} \end{pmatrix}$.

For Case 2: $\lambda_v(0|1) = 0.89, \lambda_v(1|1) = 0.11, \lambda_v(0|2) = 0.923, \lambda_v(1|2) = 0.077, \lambda_d(0) = 0.52, \lambda_d(1) = 0.48$ and $\lambda_v(l|k) = 0, \lambda_d(l) = 0$ for $l \geq 2, k = 1, 2$. The transition matrix of voice arrival process $(r_v(i,j))_{i,j}$ is $\begin{pmatrix} 1 - 0.66 \times 10^{-3} & 0.66 \times 10^{-3} \\ 0.65 \times 10^{-3} & 1 - 0.65 \times 10^{-3} \end{pmatrix}$.

Figure 2 shows the processes of the number of cells in the system under the five controls. Adaptive controls (suboptimal controls with incomplete observation and optimal controls with complete observation) regulate the number of accepted cells when the number of cells in the system is not small. Thus, the number of cells in the system increases steeply under those controls at the beginning of the busy period and is mitigated afterwards. It is shown that the number of cells in the system under the non-adaptive control $\tilde{N}_v = 3$ exceeds those under adaptive controls around the middle of the busy period ($t \in [3, 6]$ in Figure 2.).

Furthermore, adaptive controls regulate the input severely when the number of cells in the system increases. This is because the voice arrival rate is estimated (or observed) to be high if voice cells arrive successively, and the number of accepted cells is small when the arrival rate is high. This mechanism is also useful for reducing congestion and is not observed for the independent arrival process [Saito 89a].

Figure 3 shows that the voice arrival phase is estimated to be 1, is drawn. In this example, since $\lambda_v(1|1) > \lambda_v(1|2)$, there is a jump in estimation $\hat{z}_v(t; 1) = \Pr(z_v(t) = 1|I_t)$ when a voice cell arrives. In other words, the posterior probability that the voice arrival phase is at 1, increase when a voice cell arrives. While no voice cells are arriving, the estimation that the arrival phase is 1 decreases exponentially. In Figure 3, the transition

from phase 1 to phase 2 occurs at 40, and the estimation that the voice arrival phase is 1 is sufficiently small for $t > 80$. Since the mean sourjourn time at phase 2 is more than 1000, the estimation is sufficiently rapid for practical use. The estimation becomes more important if the difference in arrival rate in each arrival phase is larger and the sojourn time in each arrival phase is longer. Fortunately, the estimation becomes easier for that case.

In our numeral examples, the total waiting times under the adaptive controls are larger than under the fixed rate control $\tilde{N}_v = 2$, and smaller than under the fixed rate control $\tilde{N}_v = 4$. The coding rates of voice cells under the adaptive controls are larger than under the control $\tilde{N}_v = 2$ and smaller than under the control $\tilde{N}_v = 4$. Performance measures of adaptive controls are compared with the control $\tilde{N}_v = 3$ to clarify their effectiveness. Figures 4 and 5 show the mean waiting times under the control $\tilde{N}_v = 3$ and under the adaptive controls, and the mean voice cell coding rate. In these figures, if the mean waiting time using an adaptive control is smaller, and the coding rate are larger, the adaptive control is well performed and is more desirable than the control $\tilde{N}_v = 3$, regarding both waiting time and coding rate.

In Figure 4, the result for case 1 is shown. The result for case 2 is shown in Figure 5. Figures 4 and 5 show the effectiveness of adaptive controls. The suboptimal control works as well as the optimal control with complete information, although the suboptimal control tends to accept more cells than the optimal one. Since the confidence interval of waiting time for the suboptimal control is larger than that for the optimal one, the variance of waiting time for suboptimal control seems to be larger than that the for optimal one.

V. Conclusions

This chapter investigated dynamic voice coding rate control in ATM networks was investigated. The embedded coding scheme was considered and the control which maximizes the average voice coding rate subject to an average waiting time constraint was analysed using the Lagrangian multiplier technique and dynamic programming. The optimal control has a simple structure and is a random modification of two deterministic feedback controls with saturation.

The incomplete observation case was considered, since in practice, the arrival process phases are not observed. Therefore, the control with incomplete observation is important in practice. The general separation principle was derived. That is, the sufficient statistics for the optimal control are the posterior distributions of the phases of arrival processes.

Finally, a practical suboptimal control was proposed. With incomplete observation, the optimal control cannot be explicitly specified because of the complexity of the analytical result. The proposed suboptimal control is implementable and has the same structure as the optimal control. The simulation results indicated its effectiveness.

The analysis presented in this chapter is also valid for an average voice cell waiting time constraint and the optimal control is also a random modification of two feedback controls with saturation.

The results of this chapter was assumed that voice and data traffic flows in ATM networks. However, this can be extended to the case such that embedded coded video traffic is added to voice and data traffic.

Appendix.

In this appendix, the structure of the optimal control is shown.

The following theorem shows that the maximal expected reward is a decreasing funcition of z, the total number of cells in the system.

Theorem 2.

For all n, i, j, t,

$$U_t(n, i, j, \hat{\mathbf{z}}_v(t)) \geq U_t(n+1, i, j, \hat{\mathbf{z}}_v(t)). \tag{A.1}$$

Proof:

The proof is by induction on t. Assume $U_{t+1}(n, i, j, \hat{\mathbf{z}}_v(t+1)) \geq U_{t+1}(n+1, i, j, \hat{\mathbf{z}}_v(t+1))$ for all $n, i, j, \hat{\mathbf{z}}_v$. If N is the optimal number of accepted voice cells when $z(t) = n+1, A_v(t) = i, A_d(t) = j, \hat{\mathbf{z}}_v(t) = \mathbf{m}$,

$$U_t(n, i, j, \mathbf{m})$$

$$\geq N - \gamma[N + n + N_d(t) - 1]^+$$

$$+ \sum_{k,l} \hat{\lambda}_v(t+1;k)\lambda_d(l)U_{t+1}([N + n + N_d(t) - 1]^+, k, l, \hat{\mathbf{z}}_v(t+1))$$

$$\geq N - \gamma[N + n + 1 + N_d(t) - 1]^+$$

$$+ \sum_{k,l} \hat{\lambda}_v(t+1;k)\lambda_d(l)U_{t+1}([N + n + 1 + N_d(t) - 1]^+, k, l, \hat{\mathbf{z}}_v(t+1))$$

$$= U_t(n+1, i, j, \mathbf{m}). \qquad \blacksquare$$

Let $\hat{U}_t(N|n, i, j, \mathbf{m})$ denote the expected reward during $[t, T]$ when $N_v(t) = N, z(t) = n, A_v(t) = i, A_d(t) = j, \hat{\mathbf{z}}_v(t) = \mathbf{m}$ with the optimal control during $[t+1, T]$. That is,

$$\hat{U}_t(N|n, i, j, \mathbf{m})$$

$$\triangleq N - \gamma W(t)$$

$$+ \sum_{k,l} \hat{\lambda}_v(t+1;k)\lambda_d(l)U_{t+1}(z(t+1),k,l,\hat{z}_v(t+1)). \qquad (A.2)$$

Here,

$$W(t) = z(t+1) = [n + N + N_d(t) - 1]^+.$$

Let $\beta_t(N|n,i,j,\mathbf{m})$ denote the difference of the expected rewards during $[t,T]$ with $N_v(t) = N+1$ and $N_v(t) = N$. That is,

$$\beta_t(N|n,i,j,\mathbf{m}) \stackrel{\triangle}{=} \hat{U}_t(N+1|n,i,j,\mathbf{m}) - \hat{U}_t(N|n,i,j,\mathbf{m}). \qquad (A.3)$$

Notice that, for all $N,n,i,j,n',i',j',\mathbf{m},t$,

$$\beta_t(N|n,i,j,\mathbf{m})$$

$$= 1 - \gamma\mathbf{1}(n + N + N_d(t) > 0)$$

$$+ \sum_{k,l} \hat{\lambda}_v(t+1;k)\lambda_d(l)\{U_{t+1}(n + N + N_d(t),k,l,\hat{z}_v(t+1))$$

$$\qquad - U_{t+1}([n + N + N_d(t) - 1]^+,k,l,\hat{z}_v(t+1))\}$$

$$= \beta_t(N + (n-n') + (j-j')|n',i',j',\mathbf{m})$$

In particular,

$$\beta_t(N-1|n+1,i,j,\mathbf{m}) = \beta_t(N|n,i,j,\mathbf{m}). \qquad (A.4)$$

In other words, $\beta_t(N|n,i,j,\mathbf{m})$ is a function of $N + n + j$ and $\mathbf{m}$. To emphasize this fact, set

$$\tilde{\beta}_t(N + n + j|\mathbf{m}) = \beta_t(N|n,i,j,\mathbf{m}). \qquad (A.5)$$

The β_t used in this Appendix is employed in the proof of Theorem 3.

The following theorem shows the concavity of the optimal reward function U_t in $z(t)$.

Define

$$\Delta U_t(n,i,j,\mathbf{m}) \stackrel{\triangle}{=} U_t(n+1,i,j,\mathbf{m}) - U_t(n,i,j,\mathbf{m}) \qquad (A.6)$$

$$\Delta^2 U_t(n,i,j,\mathbf{m}) \stackrel{\triangle}{=} \Delta U_t(n+1,i,j,\mathbf{m}) - \Delta U_t(n,i,j,\mathbf{m}) \qquad (A.7)$$

to simplify the notation.

We show that $\beta_t(N|n,i,j,\mathbf{m})$ is a decreasing function of N.

Lemma 1. *If, for all $n,i,j,\mathbf{m}$,*

$$\Delta^2 U_{t+1}(n,i,j,\mathbf{m}) \leq 0,$$

then, for all $N,n,i,j,\mathbf{m}$,

$$\beta_t(N|n,i,j,\mathbf{m}) \geq \beta_t(N+1|n,i,j,\mathbf{m}).$$

Proof: From the above assumption,

$$\beta_t(N+1|n,i,j,\mathbf{m}) - \beta_t(N|n,i,j,\mathbf{m})$$

$$= \gamma(\mathbf{1}(n + N + N_d(t) > 0) - 1)$$

$$\sum_{k,l} \hat{\lambda}_v(t+1;k)\lambda_d(l)$$

$$\{U_{t+1}(n + 1 + N + N_d(t),k,l,\mathbf{m}')$$

$$\qquad - 2U_{t+1}(n + N + N_d(t),k,l,\mathbf{m}')$$

$$\qquad + U_{t+1}([n + N + N_d(t) - 1]^+,k,l,\mathbf{m}')\}$$

$$\leq 0. \qquad \blacksquare$$

Note that if $\Delta^2 U_{t+1}(n,i,j,\mathbf{m}) \leq 0$ and $\beta_t(N|n,i,j,\mathbf{m}) \geq 0$, then $\beta_t(N'|n,i,j,\mathbf{m}) \geq 0$ for $N' \leq N$, using Lemma 1.

The following is assumed for simplifying the result. If maximization is achieved by more than one integer, we adopt the largest one as optimal. That is, if

$$\hat{U}_t(N|n, i, j, \mathbf{m}) = \hat{U}_t(N - 1|n, i, j, \mathbf{m}),$$

$$\hat{U}_t(N|n, i, j, \mathbf{m}) > \hat{U}_t(N'|n, i, j, \mathbf{m}), \quad (N' \neq N - 1)$$

then the optimal number of accepted voice cells is N, not $N - 1$.

It is shown by the following Lemma that the optimal number of accepted cells is a decreasing function of the number of cells in the system.

Lemma 2.

If, for all $n, i, j, \mathbf{m}$,

$$\Delta^2 U_{t+1}(n, i, j, \mathbf{m}) \leq 0,$$

then

$$N_v^*(t; n, i, j, \mathbf{m}) = \begin{cases} N_v^*(t; n + 1, i, j, \mathbf{m}), & \text{or} \\ N_v^*(t; n + 1, i, j, \mathbf{m}) + 1. \end{cases}$$

Here, $N_v^*(t; n, i, j, \mathbf{m})$ is the optimal total number of accepting voice cells at t, when $\boldsymbol{z}(t) = n, A_v(t) = i, A_d(t) = j, \hat{\boldsymbol{z}}_v(t) = \mathbf{m}$.

Proof:

Set

$$N_0 = N_v^*(t; n, i, j, \mathbf{m}),$$

$$N_1 = N_v^*(t; n + 1, i, j, \mathbf{m}),$$

to simplify the notation.

By Lemma 1 and (A.4),

$$0 \geq \beta_t(N|n + 1, i, j, \mathbf{m}) - \beta_t(N|n, i, j, \mathbf{m}). \tag{A.8}$$

Using the optimality of N_0,

$$0 > \beta_t(N_0|n, i, j, \mathbf{m}).$$

By Lemma 1, for all N ($N_0 \leq N < N_{max}$)

$$\beta_t(N_0|n, i, j, \mathbf{m}) \geq \beta_t(N|n, i, j, \mathbf{m}).$$

Thus, using (A.8),

$$0 > \beta_t(N|n, i, j, \mathbf{m}) \geq \beta_t(N|n + 1, i, j, \mathbf{m}).$$

Therefore,

$$N_0 \geq N_1. \tag{A.9}$$

For the remaining, the proof is divided into three disjoint cases:

(i) $N_0 = N_{min}$, (ii) $N_{min} < N_0 < N_{max}$ and (iii) $N_0 = N_{max}$.

(i) $N_1 = N_{min}$, since $N_0 \geq N_1$.

(ii) Assume $N_{min} < N_0 < N_{max}$. By the optimality of N_0,

$$0 > \beta_t(N_0|n, i, j, \mathbf{m}),$$

$$0 \leq \beta_t(N_0 - 1|n, i, j, \mathbf{m}).$$

Using Eq. (A.4),

$$0 > \beta_t(N_0 - 1|n + 1, i, j, \mathbf{m}),$$

$$0 \leq \beta_t(N_0 - 2|n + 1, i, j, \mathbf{m}).$$

Thus, using Lemma 1, $N_1 = N_0 - 1$ when $N_0 > N_1$. Consequently, $N_1 = N_0, N_0 - 1$.

(iii) Assume $N_0 = N_{max}$. By the optimality of N_0,

$$0 \leq \beta_t(N_{max} - 1 | n, i, j, \mathbf{m})$$

$$= \beta_t(N_{max} - 2 | n + 1, i, j, \mathbf{m})$$

Therefore, by Lemma 1 and the above equation, $N_1 = N_{max}$ or $N_{max} - 1$.

This concludes the proof. ■

Theorem 3.

$$\Delta^2 U_t(n, i, j, \mathbf{m}) \leq 0 \qquad (A.10)$$

for all $n, i, j, \mathbf{m}, t$.

Proof:

Let N_0, N_1 and N_2 denote the optimal control at $z(t) = n, n+1$ and $n+2$. That is,

$$N_0 \triangleq N_v^*(t; n, i, j, \mathbf{m})$$

$$N_1 \triangleq N_v^*(t; n+1, i, j, \mathbf{m})$$

$$N_2 \triangleq N_v^*(t; n+2, i, j, \mathbf{m}).$$

We first consider the case $z(t) = A_v(t) = A_d(t) = 0$. Then, $N_0 = N_1 = N_2 = 0$, and

$$\Delta^2 U_t(0, 0, 0, \mathbf{m})$$

$$= -\gamma + \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l)(U_{t+1}(1, k, l, \mathbf{m}') - U_{t+1}(0, k, l, \mathbf{m}')) < 0. \qquad (A.11)$$

Next, we consider the case $z(t) + A_d(t) > 0$. The proof is by induction on t. Assume, for all $n, i, j, \mathbf{m}$,

$$\Delta^2 U_{t+1}(n, i, j, \mathbf{m}) \leq 0.$$

$$\Delta^2 U_t(n, i, j, \mathbf{m}) = (N_2 - 2N_1 + N_0)(1 - \gamma)$$

$$+ \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l)$$

$$\{ U_{t+1}(N_2 + n + N_d(t) + 1, k, l, \mathbf{m}')$$

$$- 2U_{t+1}(N_1 + n + N_d(t), k, l, \mathbf{m}')$$

$$+ U_{t+1}(N_0 + n + N_d(t) - 1, k, l, \mathbf{m}') \}. \qquad (A.12)$$

By Lemma 2, we should consider the four disjoint cases: (i) $N_2 = N_1 = N_0$, (ii) $N_2 = N_1 = N_0 - 1$, (iii) $N_2 = N_1 - 1 = N_0 - 1$ and (iv) $N_2 = N_1 - 1 = N_0 - 2$.

(i) Set $N \triangleq N_0 = N_1 = N_2$. By (3.18),

$$\Delta^2 U_t(n, i, j, \mathbf{m})$$

$$= \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l) \Delta^2 U_{t+1}(N + n + N_d(t) - 1, k, l, \mathbf{m}') \leq 0.$$

(ii) Set $N \triangleq N_2 = N_1$ and $N = N_0 - 1$. By (3.18)

$$\Delta^2 U_t(n, i, j, \mathbf{m})$$

$$= 1 - \gamma + \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l) \Delta U_{t+1}(n + N + N_d(t), k, l, \mathbf{m}')$$

$$= \beta_t(N | n + 1, i, j, \mathbf{m}).$$

By the optimality of $N_1 = N$, $\beta_t(N | n + 1, i, j, \mathbf{m}) < 0$.

(iii) Set $N \triangleq N_2 = N_1 - 1 = N_0 - 1$.

$$\Delta^2 U_t(n, i, j, \mathbf{m})$$

$$= -1 + \gamma - \sum_{k,l} \hat{\lambda}_v(t+1; k) \lambda_d(l) \Delta U_{t+1}(n + N + A_d(t), k, l, \mathbf{m}')$$

$$= -\beta_t(N | n + 1, i, j, \mathbf{m})$$

By the optimality of $N_1 = N + 1$, $\beta_t(N | n + 1, i, j, \mathbf{m}) \geq 0$.

(iv) Assume $N_2 = N_1 - 1 = N_0 - 2$.

$$\Delta^2 U_t(n, i, j, \mathbf{m}) = 0$$

This concludes the proof. ∎

It is shown that the optimal control of the Legrangian (3.1) is monotonic in the following sense and has a simple structure: feedback control with saturation.

Theorem 4.

For all $n, i, j, \mathbf{m}, t$, there exists an integer $K_\gamma^*(t; \mathbf{m})$ and

$$N_v^*(t; n, i, j, \mathbf{m})$$

$$= N_v^*(t; n+1, i, j, \mathbf{m}) \ or \ N_v^*(t; n+1, i, j, \mathbf{m}) + 1, \qquad (A.13)$$

$$= [K_\gamma^*(t; \mathbf{m}) - n - N_d(t)]_{N_{min}}^{N_{max}}. \qquad (A.14)$$

Here, $y = [x]_a^b$ denotes a saturation function,

$$y = \begin{cases} b, & \text{for } b < x, \\ x, & \text{for } a \le x \le b, \\ a, & \text{for } x < a. \end{cases} \qquad (A.15)$$

Proof:

Employing Lemma 2 and Theorem 3, Eq. (A.13) showing monotonicity of N_v^*, can be proved. Using Lemma 1 and Theorem 3, for all $k, \mathbf{m}, t$,

$$\tilde{\beta}_t(k|\mathbf{m}) \ge \tilde{\beta}_t(k+1|\mathbf{m}).$$

Therefore, there exists an integer $K_\gamma^*(t; \mathbf{m}) \ge 0$, for fixed $t, \mathbf{m}$ such that

$$\tilde{\beta}_t(k|\mathbf{m}) \begin{cases} < 0, & \text{for } K_\gamma^*(t; \mathbf{m}) \le k, \\ \ge 0, & \text{for } K_\gamma^*(t; \mathbf{m}) > k, \end{cases}$$

or for all $k \ge 0$,

$$\tilde{\beta}_t(k|\mathbf{m}) \ge 0,$$

or for all $k \ge 0$,

$$\tilde{\beta}_t(k|\mathbf{m}) < 0.$$

Therefore, if $K_\gamma^*(t; \mathbf{m}) \in [n + i\tilde{N}_{min} + j, n + i\tilde{N}_{max} + j]$, then there exists an integer N ($N_{min} \le N \le N_{max}$) for fixed i, j such that

$$\beta_t(N|n, i, j, \mathbf{m}) \begin{cases} < 0, & \text{if } K_\gamma^*(t; \mathbf{m}) \le n + N + j \\ \ge 0, & \text{if } K_\gamma^*(t; \mathbf{m}) > n + N + j, \end{cases}$$

that is, $N_v^*(t; n, i, j, \mathbf{m}) = K_\gamma^*(t; \mathbf{m}) - n - j$.

If $K_\gamma^*(t; \mathbf{m}) < n + i\tilde{N}_{min} + j$, $\beta_t(N|n, i, j, \mathbf{m}) < 0$ for $N \in [i\tilde{N}_{min}, i\tilde{N}_{max}]$, that is, $N_v^*(t; n, i, j, \mathbf{m}) = i\tilde{N}_{min}$.

If $K_\gamma^*(t; \mathbf{m}) > n + i\tilde{N}_{max} + j$, $\beta_t(N|n, i, j, \mathbf{m}) \ge 0$ for $N \in [i\tilde{N}_{min}, i\tilde{N}_{max}]$, that is, $N_v^*(t; n, i, j, \mathbf{m}) = i\tilde{N}_{max}$.

Consequently, $N_v^*(t; n, i, j, \mathbf{m}) = [K_\gamma^*(t; \mathbf{m}) - n - N_d(t)]_{N_{min}}^{N_{max}}$. ∎

Theorem 4 shows that the optimal number of accepted cells is a decreasing function of the number of cells in the system.

By Theorem 4, the structure of the optimal control for the Lagrangian is clarified and we proceed to analysis of the original problem (2.1) and (2.2).

Theorem 5.

Assume $\gamma_1 > \gamma_2$. Let u_1 and u_2 denote the optimal controls for the Lagrangians J_{γ_1} and J_{γ_2}.

Then, for all $\gamma_1 > \gamma_2$, the criterion (2.1) and the constraint (2.2) follow the inequalities:

$$J(u_1) \leq J(u_2), \qquad (A.16)$$

$$C(u_1) \leq C(u_2). \qquad (A.17)$$

Proof: By the optimality of u_1 at $\gamma = \gamma_1$,

$$J(u_1) - \gamma_1 C(u_1) \geq J(u_2) - \gamma_1 C(u_2).$$

Therefore, if $C(u_1) > (=)C(u_2)$, then $J(u_1) > (\geq)J(u_2)$, because $\gamma_1 > 0$.

If $J(u_1) < (=)J(u_2)$, $C(u_1) < (\leq)C(u_2)$.

Similarly, if $J(u_1) > (=)J(u_2)$, then $C(u_1) > (\geq)C(u_2)$, because of the optimality of u_2 at $\gamma = \gamma_2$ and $\gamma_2 \geq 0$.

If $C(u_1) < (=)C(u_2)$, $J(u_1) < (\leq)J(u_2)$.

Therefore,

$$J(u_1) \leq J(u_2), C(u_1) \leq C(u_2),$$

or

$$J(u_1) \geq J(u_2), C(u_1) \geq C(u_2),$$

for all $\gamma_1 > \gamma_2$.

If $\gamma = 0$, then maximization of the Lagrangian (3.1) reduces to the maximization of the criterion (2.1) without the constraint (2.2). Conversely, when $\gamma \to \infty$, the maximization of the Lagrangian (3.1) becomes minimization of (2.2) without considering (2.1). Therefore, $J(u_1) < J(u_2)$, when $\gamma_1 \to \infty$ and $\gamma_2 = 0$. Consequently,

$$J(u_1) \leq J(u_2), C(u_1) \leq C(u_2),$$

for all $\gamma_1 > \gamma_2$. ∎

Denote u_γ as the optimal control of the Lagrangian J_γ.

In particular, let u_∞ be the control by which $N_v(t; n, i, j, \mathbf{m}) = N_{min}$ for all $t, n, i, j, \mathbf{m}$, and let u_0 be the control by which $N_v(t; n, i, j, \mathbf{m}) = N_{max}$ for all $t, n, i, j, \mathbf{m}$.

Theorem 6.

If $C(u_\infty) \leq c < C(u_0)$, there exist an $r \in [0, 1]$ and a $\gamma \geq 0$ such that $C(v(r, \gamma)) = c$ and $v(r, \gamma)$ is the optimal control of the Lagrangian J_γ. Here, $v(r, \gamma)$ is the control which employs either u_γ or $u_{\gamma-0}$ with probabilities r and $1 - r$.

Proof:

Employing Theorem 5, there exists a γ such that $C(u_{\gamma-0}) \leq c \leq C(u_\gamma)$, if $C(u_\infty) \leq c < C(u_0)$. $C(v(r, \gamma))$ is a continuous function of r, and $C(v(0, \gamma)) = C(u_{\gamma-0})$ and $C(v(1, \gamma)) = C(u_\gamma)$.

Therefore, there exists an $r \in [0, 1]$ such that $C(v(r, \gamma)) = c$.

If $J_\gamma(u_\gamma) < J_\gamma(u_{\gamma-0})$, $J_{\gamma-0}(u_{\gamma-0}) = J_\gamma(u_{\gamma-0}) > J_\gamma(u_\gamma)$. This contradicts the optimality of u_γ. Therefore, $J_\gamma(u_\gamma) = J_\gamma(u_{\gamma-0})$.

In addition, for all controls u,

$$J_\gamma(v(r, \gamma)) = r J_\gamma(u_\gamma) + (1 - r) J_\gamma(u_{\gamma-0})$$

$$= J_\gamma(u_\gamma)$$

$$\geq J_\gamma(u).$$

Thus, $v(r, \gamma)$ is an optimal control of the Lagrangian J_γ.

This completes the proof. ∎

Theorem 7.

If $C(u_\infty) \leq c < C(u_0)$, there exist an $r \in [0,1]$ and a $\gamma \geq 0$ such that $v(r, \gamma)$ is the optimal control of (2.1) with the constraint (2.2).

If $C(u_\infty) > c$, there is no control which meets the constraint (2.2).

If $C(u_0) \leq c$, u_0 is the optimal control of (2.1), which meets the constraint (2.2).

Proof:

First, assume $C(u_\infty) \leq c < C(u_0)$. Then, there exist an $r \in [0,1]$ and a $\gamma \geq 0$ such that $v(r, \gamma)$ is the optimal control of J_γ with $C(v(r, \gamma)) = c$, employing Theorem 6. Thus, for any control u which meets a constraint (2.2),

$$J(v(r, \gamma)) - \gamma C(v(r, \gamma))$$

$$= J_\gamma(v(r, \gamma))$$

$$\geq J_\gamma(u)$$

$$= J(u) - \gamma C(u)$$

Therefore,

$$J(v(r, \gamma)) - J(u) \geq \gamma(C(v(r, \gamma)) - C(u)) = \gamma(c - C(u)) \geq 0.$$

If $C(u_\infty) > c$, then, for any control u, $C(u) > c$. If $C(u_0) \leq c$, then for any control u, $J(u_0) \geq J(u)$. ■
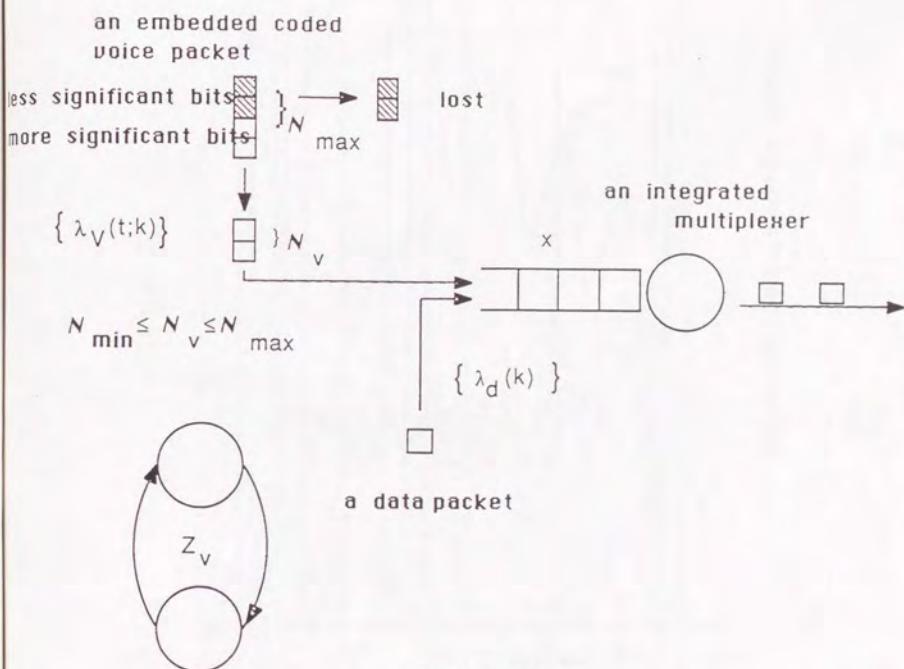
Figure 1. Model
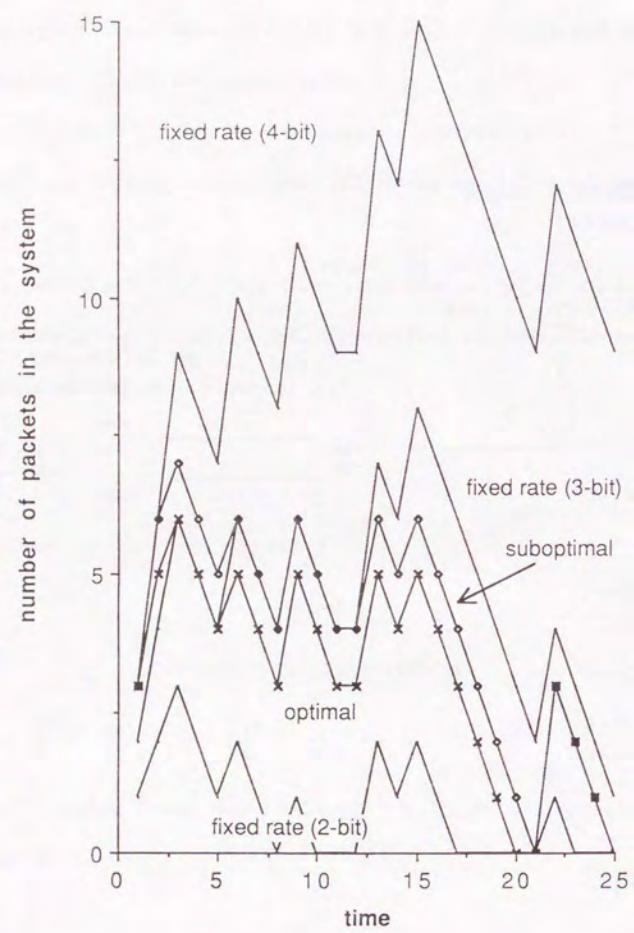
Figure 2. Number of packets in the system
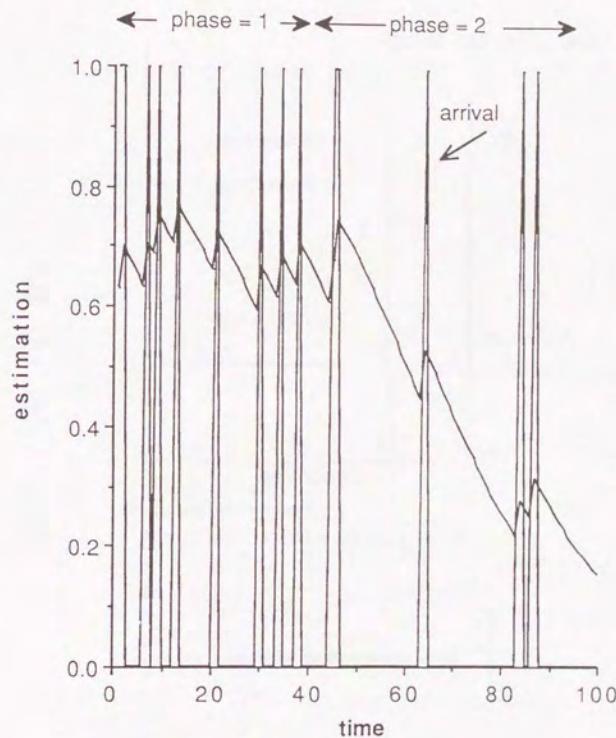
Figure 3. Phase of voice arrival process

**mean waiting time (ms)** (y-axis, 0.5 to 1.5)

mean coding rate (x-axis, 2.75 to 3.50)

fixed rate (3-bit) control

suboptimal
(incomplete observation)

optimal
(complete observation)

95% confidence interval
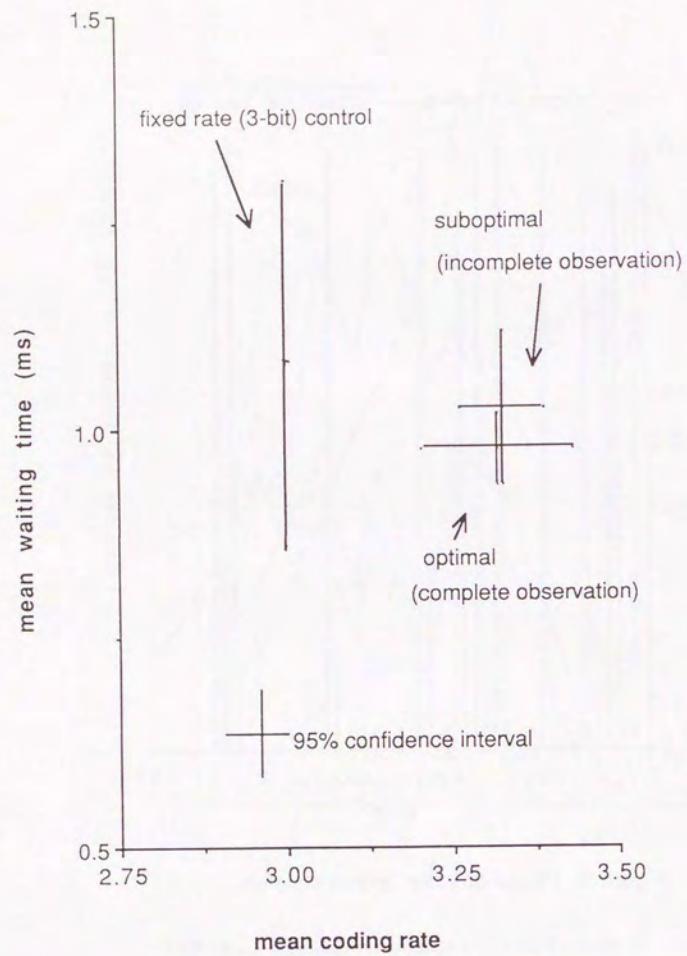
Figure 4. Simulation results (1)

mean waiting time (ms) (y-axis, 0.0 to 1.0)

mean coding rate (x-axis, 2.5 to 3.5)

fixed rate (3-bit) control

suboptimal
(incomplete observation)

optimal
(complete observation)

95% confidence interval
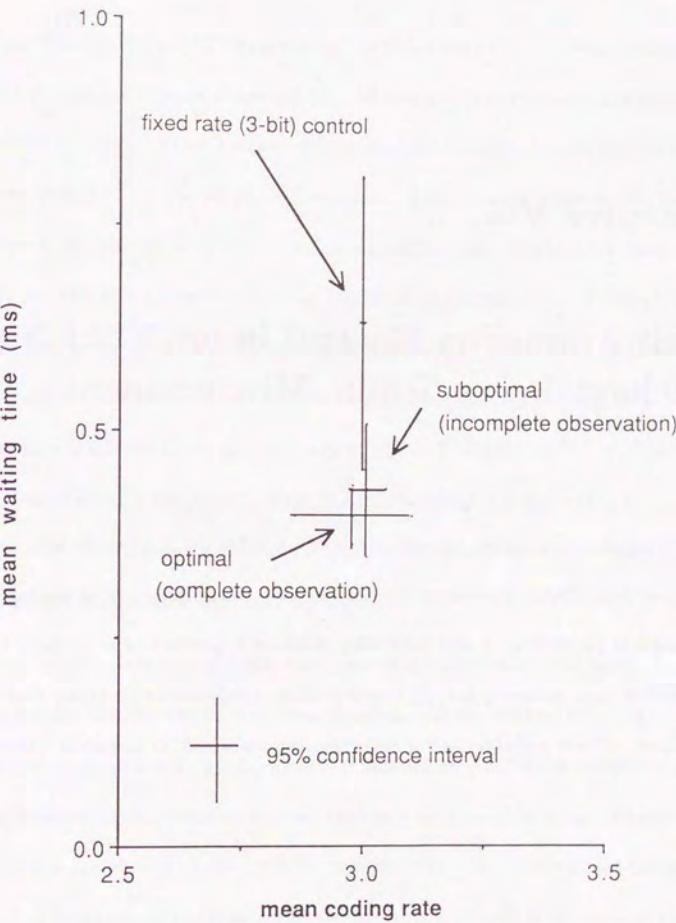
Figure 5. Simulation results (2)

Chapter VII

Call Admission Control in an ATM Network without Using Traffic Measurement

This chapter investigates call admission control in ATM networks without monitoring network load. Traffic parameters specified by users are employed to obtain the upper bound of cell loss probability. A loss probability standard is guaranteed to be satisfied under this control without assuming the cell arrival process. Implementation of this control to quickly evaluate cell loss probability after acceptance of a new call, is discussed. [Saito 90e]

1. Introduction

Asynchronous Transfer Mode (ATM) is expected to be a target technology for Broadband ISDN (B-ISDN) [Turner 86b], [Kawarazaki 88], [Händel 89]. An ATM network is a high-speed multimedia network which handles various classes of traffic with different bit-rates and different quality of service (QOS) requirements. For example, voice traffic has the speed of several kilobits per second and is delay sensitive, while high-speed data traffic used for file transfer or LAN interconnection has the speed of hundreds of mega-bits per second and is loss sensitive. To achieve high-speed multimedia transport, an ATM network adopts a simplified transport protocol based on hardware cell switching with no flow control or retransmission inside the network. Necessary numbers of cells are allocated on demand, and bursty information is statistically multiplexed. Statistical multiplexing may lead to efficient use of network resources, but may require new kinds of bandwidth management and traffic control [Kawashima 89], [Eckberg 89].

Call admission control is one of the most important issues under discussion. In STM networks, when the bandwidth of a new connection exceeds the residual capacity of links, the call request is rejected. On the other hand, in ATM networks, the bandwidth of a call is not clear, since all the information is segmented into fixed sized cells and the necessary number of cells is generated and conveyed through the networks. However, as the number of connected calls increases, the grade-of-service (GOS) of the cell level, such as the cell loss probability, deteriorates. Thus, it is impossible to accept an unlimited number of calls under GOS requirements. It is also necessary to limit the number of accepted calls in ATM networks.

To attain high utilization of links under the grade-of-service (GOS) standards, call admission controls must decide appropriately whether to accept a new connection based on the new connection's anticipated traffic characteristics, the quality requirements of connected calls including the new call, and (if possible) the current network load measurement [Woodruff 88]. Here, the new connection's anticipated traffic characteristics are estimated from the traffic parameters specified by the user.

The traffic parameters should be easy for the user to specify and its conformity with the real load should be easy for a network provider to monitor. In addition, it should involve enough information to guess network performance after acceptance of the new connection. Candidates of traffic parameters are: peak-bit-rate (PBR) and average-bit-rate (ABR) [Kowalk 88]; PBR, ABR and burstiness (=PBR/ABR) [Woodruff 88]; PBR, ABR and bit rate variance [Verbiest 88]; maximum call throughput in short and medium durations [Ohnishi 88]; and burstiness, ABR, and average peak duration [Gallassi 89].

Call admission controls which determine call admission based on a burst model (Figure 1) without considering output buffers are proposed by [Schoute 88], [Verbiest 88], [Burgin 89b] and [Murase 89]. These methods evaluate the instantaneous total load, and control a call request such that the instantaneous load does not exceed the specified level. They are practical, but the relationship between buffer size and cell loss probability is not clear. Thus, the cell loss probability standard may not be satisfied. (The specified level does not depend on buffer size. Intuitively, the proportion of the instantaneous load exceeding the link capacity seems to be the upper bound of cell loss probability, but in fact it is not.) Difficulties occur during implementation of these methods when the number of call classes is large, since the tail of the binomial distribution is necessary. An implementation shown in this paper is also useful for the above methods. One method whereby virtual bandwidth for inhomogeneous traffic is obtained from that for homogeneous traffic, is proposed in [Filipiak 89]. This method is also practical, but it cannot gurantee that the cell loss probability standards will satisfied.

In [Gallassi 89], the queueing effect in the output buffers is taken into account and the ratio between the assigned bandwidth and the peak bandwidth is determined by the number of multiplexed sources, burstiness, and average peak duration from the simulation results, when traffic is homogeneous. For inhomogeneous traffic, the assigned bandwidth is given by the sum of the assigned bandwidths of individual traffic classes or is the bandwidth obtained under the assumption that all multiplexed sources belong to the class with the largest burstiness (worst case). This method is practical and interesting, but the difficulties lie in the facts that users must specify the average peak duration and that utilization decreases when the number of call classes increases.

This paper proposes call admission control which guarantees cell loss probability standards by considering the output buffer size and the queueing effect there. It is based only on traffic parameters specified by a user and does not require a cell arrival process model. The result in this paper is also applicable to ATM network dimensioning based on traffic parameters.

2. Preliminaries

Of the performance deterioration factors, cell segmentation delay and propagation delay over transmission links are fixed, and are independent of traffic characteristics and call admission control. Cell loss and misdelivery due to header field error in transmission are also independent of the traffic. Cell loss and delay in a switch are negligible. Thus, the key factors in ATM network performance deterioration are cell loss and delay in the output queue to a transmission link in ATM nodes, and our study focuses on these factors.

This paper focuses on a single ATM node, and proposes a call admission control which enables the cell loss probability standard to be satisfied in the node. In all nodes along the route of a call, the call admission control functions, and if at least one of the intermediate nodes cannot accept the call, it is rejected.

Consider a transmission link of C Mb/s and its output buffer. The buffer size is assumed to be K. Let T ms be the maximum admissible delay in the buffer. Assume that the buffer size K is dimensioned such that the maximum delay in the buffer is less than T under an FIFO discipline [Saito 89c]:

$$K = 1000TC/L, \tag{2.1}$$

where L is the cell length in bits. Thus, the cell delay standard is always satisfied, and call admission control concentrates on the fulfilling the cell loss requirement.

Remark: Even if another dimensioning method is used, the discussions for the call admission control proposed in this paper are valid and the control gurantees the cell loss probability standards. However, additional efforts are necessary for satisfying the delay requirement.

To take account of feasibility and simplicity of user specification for the traffic parameters, the sets of traffic parameters specified by users are assumed to be the average number and the maximum number of cells arriving during a fixed interval, and the average number and the variance of the number of cells arriving during a fixed interval. We assume that the length of the interval is equal to the time at which $K/2$ cells are transmitted, that is $T/2$ ms. The reason why the interval length should be $K/2$ will be stated later. (We assume $K/2$ is an integer in the remainder of this paper, for simplicity.) We call the average number of cells arriving during this interval, ANA, the maximum number, MNA, and the variance, VNA. Here, MNA is an interger.

Here, the notation for the remainder of this paper is introduced. $\lfloor x \rfloor$ is the largest integer less than or equal to x, $\lceil x \rceil$ is the smallest integer greater than or equal to x, and
$$[x]^+ = \begin{cases} x, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

3. Upper bound of cell loss probability from MNA and ANA

In this section, we consider the case where the traffic parameters specified by users are the average number of cells arriving during a fixed time interval (ANA) and the maximum number of cells arriving during a fixed time interval (MNA), and derive a cell loss probability upper bound based on that using the distribution of the number of cells arriving during a fixed time interval [Saito 89c, 90e].

Let us assume that cells of N calls are transmitted by a transmission link, and that the probability that j cells of the i-th call arrive during $T/2$ ms is $p_i^*(j)$, $i = 1, \ldots, N; j = 0, 1, \ldots$. Then, the cell loss probability in the buffer, CLP, is upper-bounded by $B(p_1^*, \ldots, p_N^*; K/2)$ defined below (Proof is in Appendix 1).

$$CLP \leq B(p_1^*, \ldots, p_N^*; K/2) \triangleq \frac{\sum_{k=0}^{\infty}[k - K/2]^+ p_1^* \star \cdots \star p_N^*(k)}{\sum_{k=0}^{\infty} k p_1^* \star \cdots \star p_N^*(k)}, \tag{3.1}$$

where $\star$ denotes the convolution.

Let $\theta_i = \{\theta_i(j)\}$ be the following Bernoulli-like distribution provided by the i-th call specified parameters (a_i, R_i). Here, R_i is the MNA and a_i is the ANA of the i-th call.

$$\theta_i(j) = \begin{cases} a_i/R_i, & j = R_i, \\ 1 - a_i/R_i, & j = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

That is, θ_i is the distribution that denotes the maximum number of cells arriving or no cell arriving and its average number of cells arriving is a_i.

By employing the result shown in Appendix 2, the upper bound of $B(p_1^*, \ldots, p_N^*; K/2)$ can be obtained.

$$CLP \leq B(p_1^*, \ldots, p_N^*; K/2)$$

$$\leq B(p_1^*, \ldots, p_{N-1}^*, \theta_N; K/2)$$

$$\leq B(p_1^*, \ldots, \theta_{N-1}, \theta_N; K/2)$$

$$\vdots$$

$$\leq B(\theta_1, \ldots, \theta_N; K/2) = \frac{\sum_{k=0}^{\infty}[k - K/2]^+ \theta_1 \star \cdots \star \theta_N(k)}{\sum_{k=0}^{\infty} k \theta_1 \star \cdots \star \theta_N(k)} \tag{3.3}$$

If for all i, R_i cells arrive in a batch at every $T/2$ ms interval during active periods and no cell arrives during silence periods from the i-th call, the durations of these periods are quite long, and the ratio of the mean active period length to the sum of the mean active and silence period lengths is equal to a_i/R_i, then the cell loss probability is equal to the right-hand-side of Eq. (3.3) (Figure 2). That is, this case gives the most bursty superposed process among the cases with specified MNAs and ANAs. This upper bound, which corresponds to the most bursty case, cannot be improved, since there exists a process to achieve this upper bound (Eq. (3.3)). In other words, if we want the estimate of CLP to become tight and the link utilization to be better, an additional traffic parameter is necessary.

This paper propose a call admission strategy based on Eq. (3.3): the $N+1$-th call, which requests admission, is to be accepted, if

$$B(\theta_1, \ldots, \theta_{N+1}; K/2) \leq \tilde{B}, \tag{3.4}$$

and rejected, if

$$B(\theta_1, \ldots, \theta_{N+1}; K/2) > \tilde{B}, \tag{3.5}$$

when call-1, ..., call-N are connected. Here, $\tilde{B}$ is the cell loss probability standard.

4. Upper bound of cell loss probability from ANA and VNA

This section investigates the upper bound of cell loss probability, when the average and variance of the number of cells arriving during $T/2$ ms (that is, ANA and VNA) are specified by users.

Assume that the number of multiplexed sources is N, and that the ANA and VNA specified by the i-th call are a_i and σ_i^2.

If we set $p = p_1^* \star \cdots \star p_N^*$ in Appendix 3, we can obtain an upper bound,

$$
\begin{aligned}
CLP &\leq B(p_1^*, \cdots, p_N^*; K/2) \\
&\leq B(q_N; K/2) \triangleq \frac{(i^* - K/2)\{l(l-1) - a(2l-1) + \sigma^2 + a^2\}}{-a(l - i^*)(i^* - l + 1)},
\end{aligned} \tag{4.1}
$$

where $a = \sum_{i=1}^N a_i$, $\sigma^2 = \sum_{i=1}^N \sigma_i^2$, $l = \lceil a \rceil$ and i^* is defined by Eqs. (C.11) and (C.12) with $r = K/2$.

Thus, when the $N + 1$-th call requests admission, $a(= \sum_{i=1}^{N+1} a_i)$, $\sigma^2(= \sum_{i=1}^{N+1} \sigma_i^2)$, l, and i^* are evaluated, including the specified traffic parameters a_{N+1} and σ_{N+1}^2. Then, $B(q_{N+1}; K/2)$ is obtained and the $N + 1$-th call is accepted if and only if

$$
B(q_{N+1}; K/2) \leq \tilde{B}. \tag{4.2}
$$

5. Implementation of call admission control

The upper bound of cell loss probability, after accepting a new requested connection, can be obtained by $B(\theta_1, \ldots, \theta_{N+1}; K/2)$ and $B(q_{N+1}; K/2)$ using only the specified parameters, and the call admission control can judge whether the new connection should be accepted or not through $B(\theta_1, \ldots, \theta_{N+1}; K/2)$ and $B(q_{N+1}; K/2)$. However, to realize call admission control based on $B(\theta_1, \ldots, \theta_{N+1}; K/2)$ and $B(q_{N+1}; K/2)$, an on-line evaluation of them is necessary.

Of these bounds, $B(q_{N+1}; K/2)$ is easy to evaluate. Thus, the implementation for on-line evaluation of $B(\theta_1, \ldots, \theta_{N+1}; K/2)$, which is expected to have superior performance to $B(q_{N+1}; K/2)$ under ordinary traffic conditions (see 6. Numerical examples), should be considered.

The first method of evaluating $B(\theta_1, \ldots, \theta_N; K/2)$ is to calculate it directly. In this case, we can use the fact that $B(\theta_1, \ldots, \theta_N; K/2)$ depends only on $\theta_1 \star \cdots \star \theta_N$. Thus, for example, if $\theta_1, \ldots, \theta_i$ have the same specific parameters (a, R), $\theta_1 \star \cdots \star \theta_i$ reduces the binomial distribution with parameters $(i, a/R)$. If the number of multiplexed sources N becomes large, $\theta_1 \star \cdots \star \theta_N$ can be approximated by a normal distribution, and this approximation reduces the computational load. Hui's result [Hui 88] is also applicable to the evaluation of $B(\theta_1, \ldots, \theta_N; K/2)$. The development of an approximation for $\theta_1 \star \cdots \star \theta_N$ including Hui's result is one of a promising method for on-line evaluation.

We discuss the second method here. The principle of the second method is as follows. The effect of connected calls on cell loss probability should be evaluated beforehand, and the effect of call requesting connection on the cell loss probability is supplemented later.

This method uses the current load state vector $S = \{S(m)\}$ defined below, and to renew it at call request and completion.

$$S(m) \stackrel{\triangle}{=} \sum_{k=m}^{\infty} Q(k), \quad m \geq 0. \tag{5.1}$$

Here, $Q = \{Q(k)\}$ is the complementary distribution of the number of cells arriving during $T/2$ ms from connected calls. That is, when call-1, ..., call-N are currently connected, it is defined by,

$$Q(k) = \sum_{i=k}^{\infty} \theta_1 \star \cdots \star \theta_N(i). \tag{5.2}$$

The call admission control maintains the current load state vector $\{S(m)\}$ and the average number of cells arriving from all connected calls $A = \sum_{i=1}^{N} a_i$. The current load state vector contains the information of the load yielded by currently connected calls, and can express the upper bounds of cell loss probability:

$$B(\theta_1, \ldots, \theta_N; K/2) = \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=K/2+1}^{\infty} Q(k) = \frac{1}{A} S(K/2 + 1) \tag{5.3}$$

When the $(N+1)$-th call with parameters (a_{N+1}, R_{N+1}) requests connection, calculate $\hat{S}(K/2 + 1) \stackrel{\triangle}{=} F_{K/2+1}(S, R_{N+1}, a_{N+1}/R_{N+1})$, and evaluate

$$B(\theta_1, \ldots, \theta_{N+1}; K/2) = \frac{1}{A + a_{N+1}} \hat{S}(K/2 + 1), \tag{5.4}$$

(Figure 3). Here, $F_i(S, j, v)$ is defined as

$$F_i(S, j, v) = (1 - v)S(i) + \begin{cases} vS(i - j), & i \geq j \\ v((j - i) + S(0)), & i < j \end{cases} \tag{5.5}$$

and represents that the load state after the load yielded by the requested calls is added. The new requested call is accepted, if and only if $B(\theta_1, \ldots, \theta_{N+1}; K/2) \leq \tilde{B}$.

If the $(N+1)$-th call is accepted, $S(m)$ substitutes $F_m(S, R_{N+1}, a_{N+1}/R_{N+1})$ for all m. The average A changes into $A + a_{N+1}$ (Figure 3). These operations indicate that the load yielded by the newly connected call is added to the load of connected calls.

When the service of call-$(N+1)$ is completed, the current load state vector $\{S(m)\}$, and the total average bit rate A are renewed (Figure 3).

$$S(m) \leftarrow F_m^{-1}(S, R_{N+1}, a_{N+1}/R_{N+1}) \tag{5.6}$$

$$A \leftarrow A - a_{N+1} \tag{5.7}$$

In short, by storing the vector $\{S(m), m = 0, 1, \cdots\}$, it is sufficient to calculate $F_{K/2+1}(S, R_{N+1}, a_{N+1}/R_{N+1})$ based on Eq. (5.5) and to evaluate the cell loss probability upper bound $B = F_{K/2+1}(S, R_{N+1}, a_{N+1}/R_{N+1})/(A + a_{N+1})$, when there is a call connection request. This mechanism is easily implementable and reduces the computation load of a switching node at a connection request.

6. Numerical examples

Figure 4 shows the number of calls which can be multiplexed in a 150 Mb/s transmission link with $T = 1$ ms and cell loss probability standard $\tilde{B} = 10^{-2}$. Two cases for MNA and ANA specification are considered in Figure 4: One is the case where calls of burstiness (=MNA/ANA) $b=5$, and the other is the case where $b=2.5$. There are also two cases for the VNA and ANA specification: VNA/(ANA)$^2 = 1$ and 4. In Figure 4, call admission control using MNA and ANA is defined by Eqs. (3.4) and (3.5), and that using VNA and ANA is defined by Eq. (4.3). The number of calls determined by these controls gives the lower bound of the number of calls which can be multiplexed. In particular, the number

of calls determined by the control using MNA and ANA is the tightest (achievable) lower bound, since the control is based on the tightest (achievable) upper bound of the cell loss probability. In addition, to show the tightness of the upper bound defined by Eq. (4.1) using VNA and ANA, the number of calls that can actually be mulitplexed is investigated. As stated at the bottom of Eq. (A.12), if the distribution of the number of cells arriving is specified, Eq. (A.13) can give the actual loss probability for a certain process. Using this fact, we assume the distribution defined in Eqs. (D.1) and (D.2), evaluate Eq. (A.13) for various parameter values in these distributions, and plot the bound obtained for the various parameter values (We call this bound the estimated bound here).

Figure 4 shows the following facts:

(1) If the estimated bound, derived by the actual cell loss probability based on the assumptions of Eqs. (D.1) and (D.2), is a true bound, the upper bound using Eq. (4.3) reduces the utilization by $10\% \sim 50\%$. This reduction becomes larger as VNA/ANA^2 becomes larger. This corresponds to the case where the VNA specification is 4 times larger than the true VNA, since the estimated bound for $VNA/ANA^2=4$ is nearly equal to the upper bound for $VNA/ANA^2=1$.

(2) The upper bound and the estimated bound for $VNA/ANA^2=4$ make the utilization $10\% \sim 50\%$ lower than that for $VNA/ANA^2=1$. Since the accuracy of these various specifications seems difficult, this sensitivity of the VNA results is a disadvantage to the VNA specification.

(3) The upper bound for MNA specification $b = 5$ is nearly equal to that for VNA specification $(VNA/ANA^2=1)$ and the estimated bound of VNA specification $(VNA/ANA^2=4)$.

The upper bound for MNA specification $b = 2.5$ is nearly equal to the estimated bound of VNA specification $(VNA/ANA^2=1)$.

(4) The utilization of $b = 5$ is $10\% \sim 50\%$ smaller than that of $b = 1$.

In Figures 5 and 6, comparisons are made between the upper bound using MNA and ANA and the queueing analysis. Video and voice traffic are assumed to share resources. A 32 kb/s Adaptive Differential PCM is assumed for voice, and the mean duration of a talkspurt is 352 ms and that of a silence period is 650 ms [Heffes 86].

In Figure 5, the MNA of the video is 30 Mb/s or 15 Mb/s, and the ANA is 5 Mb/s. A single video source is assumed to be multiplexed with many voice sources. The actual cell loss probability is obtained in [Saito 91]. The result shows that by using the upper bound the utilization is decreased by about $20\% \sim 30\%$. In other words, the MNA and ANA specification causes a utilization loss of about $20\% \sim 30\%$, compared with the complete cell arrival process specification.

In Figure 6, the results are compared with those in [Yamada 90]. The maximum bit rate of a frame in video traffic is 33.0 Mb/s and the average bit rate is 13.5 Mb/s. We assume that MNA is 33.0 Mb/s and ANA is 13.5 Mb/s. Two or four video sources are multiplexed with voice sources. Utilization based on the upper bound decreases more than 30% compared with that based on the actual cell loss probability.

The loss in this case is larger than that in the case of Figure 5 or [Saito 91] since the process in Figure 6 or [Yamada 90] is more bursty than that in Figure 5 or [Saito 90e]. Figures 5 and 6 show that only the MNA and ANA specification causes a utilization loss of about, for example, 30% compared with the complete cell arrival process specification

under practical traffic conditions.

7. Conclusions

This paper investigated a call admission control in ATM networks. The traffic parameters specified by users are assumed to be the average and maximum numbers of cells arriving during a fixed interval, and the average and the variance of the number of cells arriving during a fixed interval. The cell loss probability in an individual node is guaranteed to be satisfied under this control without assuming a cell arrival process. It was shown that the call admission control with the maximum and average numbers of cells can attain the limit of utilization under the quality requirement, despite the maximum and average number specification giving a fairly considerable amount of utilization loss compared with the complete cell arrival process specification. If a higher utilization is required, an additional or other traffic parameters are necessary.

Implementation methods were discussed with regard to their on-line evaluation of cell loss probability. The result in this paper is also applicable to ATM network dimensioning based on traffic parameters.

Appendix 1

We assume that the time axis is divided into slots, whose length is equal to that of cell transmission time h, where $h = 10^{-3}L/C$. Let r be an integer and assume that $r \leq K + 1$. Consider the queueing behavior at every r slots (Figure A1). We call this interval of length r, r-interval. A cell arrives at the beginning of a slot, and is transmitted at the end of a slot. Let n_k be the number of cells in the system including the one at the server just before the k-th r-interval, that is, just after the transmission of the $(k-1)r$-th slot and just before the arrival of the $(k-1)r + 1$-th slot. Let $a_{i,j}$ be the number of cells arriving and $l_{i,j}(n_i)$ be the number of cells lost given n_i, both during the i-th through j-th r-interval, that is, during the slots of $[(i-1)r + 1, jr]$. Let $i_k(n_k)$ be the number of slots in which the transmission link is idle, that is, the system is empty, during the k-th r-interval given n_k. We introduce the function $f_r(z) \triangleq \begin{cases} z, & \text{if } z \geq r \\ r, & \text{if } z < r \end{cases}$, which denotes the operation that if the number of cells in the system is less than r, the number of cells is increased to r.

Consider the number of cells lost during $[1, kr]$ given $a_{1,1}, \cdots, a_{k,k}$. We first note that

$$\text{maximize } l_{k,k}(n_k) \leq [f_r(n_k) + a_{k,k} - (K+1)]^+. \tag{A.1}$$

Then, from the optimality equation [Howard 71],

$$\text{maximize } l_{k-1,k}(n_{k-1})$$

$$=\text{maximize } \{l_{k-1,k-1}(n_{k-1}) + \text{maximize } l_{k,k}(n_k)\}$$

$$\leq\text{maximize } \{l_{k-1,k-1}(n_{k-1}) + [f_r(n_k) + a_{k,k} - (K+1)]^+\}$$

$$\leq\text{maximize } \{l_{k-1,k-1}(f_r(n_{k-1})) + [f_r(n_k) + a_{k,k} - (K+1)]^+\}. \tag{A.2}$$

Note that

$$n_k = n_{k-1} + a_{k-1,k-1} - r + i_{k-1}(n_{k-1}) - l_{k-1,k-1}(n_{k-1}), \qquad (A.3)$$

$$n_k = f_r(n_{k-1}) + a_{k-1,k-1} - r - l_{k-1,k-1}(f_r(n_{k-1})). \qquad (A.4)$$

We first consider the case where $f_r(n_{k-1}) + a_{k-1,k-1} \leq (K+1)$. For this case, there is no loss during the $(k-1)$-th r-interval, that is, $l_{k-1,k-1}(f_r(n_{k-1})) = 0$, and $n_k = f_r(n_{k-1}) + a_{k-1,k-1} - r \leq K + 1 - r$ from Eq. (A.4). Thus, from Eq. (A.2),

$$\text{maximize } l_{k-1,k}(n_{k-1}) \leq \text{maximize } [f_r(n_k) + a_{k,k} - (K+1)]^+$$

$$\leq \text{maximize } [\max(r, K+1-r) + a_{k,k} - (K+1)]^+$$

$$= [a_{k,k} - \min(K+1-r, r)]^+. \qquad (A.5)$$

We next consider the case where $f_r(n_{k-1}) + a_{k-1,k-1} > (K+1)$. For this case, $l_{k-1,k-1}(f_r(n_{k-1})) \leq [f_r(n_{k-1}) + a_{k-1,k-1} - (K+1)]^+$. From Eq. (A.2), we note

$$\text{maximize } l_{k,k-1}(n_{k-1})$$

$$\leq \text{maximize} \begin{cases} l_{k-1,k-1}(f_r(n_{k-1})) \\ l_{k-1,k-1}(f_r(n_{k-1})) + (r + a_{k,k} - (K+1)) \\ l_{k-1,k-1}(f_r(n_{k-1})) \\ \qquad + (f_r(n_{k-1}) + a_{k-1,k-1} - r - l_{k-1,k-1}(f_r(n_{k-1})) + a_{k,k} - (K+1)) \end{cases}$$

$$\leq \begin{cases} [f_r(n_{k-1}) + a_{k-1,k-1} - (K+1)]^+ \\ [f_r(n_{k-1}) + a_{k-1,k-1} - (K+1)]^+ + (r + a_{k,k} - (K+1)) \\ f_r(n_{k-1}) + a_{k-1,k-1} - r + a_{k,k} - (K+1) \end{cases} \qquad (A.6)$$

and hence we obtain

$$\text{maximize } l_{k,k-1}(n_{k-1})$$

$$\leq [f_r(n_{k-1}) + a_{k-1,k-1} - (K+1)]^+ + [a_{k,k} - \min(r, K+1-r)]^+. \qquad (A.7)$$

Consequently, for both cases,

$$\text{maximize } l_{k-1,k}(n_{k-1}) \leq [f_r(n_{k-1}) + a_{k-1,k-1} - (K+1)]^+$$

$$+ [a_{k,k} - \min(r, K+1-r)]^+. \qquad (A.8)$$

Comparing Eq. (A.8) with Eq. (A.2), we obtain

$$\text{maximize } l_{k-2,k}(n_{k-2})$$

$$\leq [f_r(n_{k-2}) + a_{k-2,k-2} - (K+1)]^+$$

$$+ [a_{k-1,k-1} - \min(r, K+1-r)]^+$$

$$+ [a_{k,k} - \min(r, K+1-r)]^+. \qquad (A.9)$$

In general,

$$\text{maximize } l_{i,k}(n_i) \leq [f_r(n_i) + a_{i,i} - (K+1)]^+ + \sum_{j=i+1}^{k} [a_{j,j} - \min(r, K+1-r)]^+. \quad (A.10)$$

Therefore, assuming the ergodicity on $a_{i,i}$, the cell loss probability (CLP) is bounded by,

$$CLP \stackrel{\triangle}{=} \lim_{k\to\infty} l_{1,k}/a_{1,k}$$

$$= \lim_{k\to\infty} [f_r(n_1) + a_{1,1} - (K+1)]^+ / a_{1,k} + \frac{\frac{1}{k}\sum_{j=1}^{k} [a_{j,j} - \min(r, K+1-r)]^+}{\frac{1}{k} a_{1,k}}$$

$$= B(p; \min(r, K+1-r)) \stackrel{\triangle}{=} \sum_{i=0}^{\infty} [i - \min(r, K+1-r)]^+ p(i) / \sum_{i=0}^{\infty} i p(i). \qquad (A.11)$$

Here, $p(i)$ denotes the probability that the number of cells arrive in an r-interval.

In particular, when $r = (K+1)/2$, $\min(r, K+1-r)$ becomes minimum. When we apply the buffer dimensioning method given by Eq. (2.1) and $r = K/2$, the length of the interval which defines the maximum and average numbers of cells arriving (MNA and

ANA) becomes $T/2$ ms and independent of the link capacity. This fact is convenient in practical situations, and the difference of min $(\tau, K+1-\tau)$ at $\tau = (K+1)/2$ and $\tau = K/2$ is small. Thus, we adopt $\tau = K/2$ as the remainder. Then the cell loss probability CLP is bounded by,

$$CLP \leq \sum_{i=0}^{\infty}[i - K/2]^{+}p(i) / \sum_{i=0}^{\infty} ip(i). \qquad (A.12)$$

For the case in which only $p(i)$, the distribution of the number of cells arriving during $T/2$ ms is known, this upper bound cannot be improved, since the CLP of the following process is equal to the right-hand side of Eq. (A.12). The process achieving the upper bound is represented by the interval during which i cells arrive every $T/2$ ms continuing sufficiently long for each i, and the proportion of the interval is equal to $p(i)$.

If N calls are multiplexed and the probability that j cells of the k-th call arrive during $T/2$ ms is $p_k^*(j)$, then $p(i) = p_1^* \star \cdots \star p_N^*(i)$. Thus,

$$CLP \leq \sum_{i=0}^{\infty}[i - K/2]^{+} p_1^* \star \cdots \star p_N^*(i) / \sum_{i=0}^{\infty} ip_1^* \star \cdots \star p_N^*(i). \qquad (A.13)$$

Appendix 2

This appendix shows that $B(p_1, \ldots, p_i, \ldots, p_N; \tau) \leq B(p_1, \ldots, \theta_i, \ldots, p_N; \tau)$ for any distribution $p_1, \ldots, p_N$. Here, θ_i is the distribution provided by the i-th call specified parameters and is defined in Eq. (3.2), and the distribution $p_j = \{p_j(k)\}$, $j = 1, \ldots, N$ is assumed to satisfy the j-th call specified parameters. That is, $p_j(k) = 0$ for $k > R_j$ since the MNA is R_j, and $\sum_{k=0}^{\infty} kp_j(k) = a_j$ since the ANA is a_j.

Let us define a distribution $p_i^{(l)} = \{p_i^{(l)}(j)\}$, $0 < l < R_i$,

$$p_i^{(l)}(j) = \begin{cases} 0, & j = l, \\ p_i(j) + \frac{l}{R_i}p_i(l), & j = R_i, \\ p_i(j) + (1 - \frac{l}{R_i})p_i(l), & j = 0, \\ p_i(j), & \text{otherwise}, \end{cases}$$

which maintains that MNA$=R_j$ and $ANA = a_j$ (Figure B1), and

$$q \stackrel{\triangle}{=} p_1 \star \cdots \star p_{i-1} \star p_{i+1} \star \cdots \star p_N.$$

Then,

$$B(p_1, \ldots, p_i, \ldots, p_N; \tau) \stackrel{\triangle}{=} \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty}[k - \tau]^{+} p_1 \star \cdots \star p_i \star \cdots \star p_N(k)$$

$$= \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty}[k - \tau]^{+} p_i \star q(k), \qquad (B.1)$$

$$B(p_1, \ldots, p_i^{(l)}, \ldots, p_N; \tau) \stackrel{\triangle}{=} \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty}[k - \tau]^{+} p_1 \star \cdots \star p_i^{(l)} \star \cdots \star p_N(k)$$

$$= \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty}[k - \tau]^{+} p_i^{(l)} \star q(k). \qquad (B.2)$$

Employing Eqs. (B.1) and (B.2),

$$B(p_1, \ldots, p_i, \ldots, p_N; \tau) - B(p_1, \ldots, p_i^{(l)}, \ldots, p_N; \tau)$$
$$= \frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty} \sum_{j=0}^{k}[k - \tau]^{+}(p_i(j) - p_i^{(l)}(j))q(k - j)$$

$$=\frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty} \{-[k-r]^+(1-\frac{l}{R_i})p_i(l) + [l+k-r]^+p_i(l) - [R_i+k-r]^+\frac{l}{R_i}p_i(l)\}q(k).$$

$$(B.3)$$

Set $f(k) = -[k-r]^+(1-l/R_i) + [l+k-r]^+ - [R_i+k-r]^+l/R_i$ where $0 < l < R_i$, and consider four disjoint cases (i) $k \leq r - R_i$, (ii) $r - R_i < k \leq r - l$, (iii) $r - l < k \leq r$, and (iv) $r < k$ to show that $f(k)$ is non-positive.

(i) Assume $k \leq r - R_i$. Then, $f(k) = 0$.

(ii) Assume $r - R_i < k \leq r - l$. Then, $f(k) = -(R_i + k - r)l/R_i < 0$.

(iii) Assume $r - l < k \leq r$. Then, $f(k) = (k-r)(1-l/R_i) \leq 0$.

(iv) Assume $r < k$. Then, $f(k) = 0$.

Thus,

$$B(p_1, \ldots, p_i, \ldots, p_N; r) - B(p_1, \ldots, p_i^{(l)}, \ldots, p_N; r)$$

$$=\frac{1}{\sum_{i=1}^{N} a_i} \sum_{k=0}^{\infty} f(k)p_i(l)q(k) \leq 0.$$

$$(B.4)$$

Similarly,

$$B(p_1, \ldots, p_i^{(l)}, \ldots, p_N; r) - B(p_1, \ldots, p_i^{(l,k)}, \ldots, p_N; r) \leq 0,$$

$$(B.5)$$

where

$$p_i^{(l,k)}(j) = \begin{cases} 0, & j = k, \\ p_i^{(l)}(j) + \frac{k}{R_i}p_i^{(l)}(k), & j = R_i, \\ p_i^{(l)}(j) + (1-\frac{k}{R_i})p_i^{(l)}(k), & j = 0, \\ p_i^{(l)}(j), & \text{otherwise.} \end{cases}$$

Consequently,

$$B(p_1, \ldots, p_i, \ldots, p_N; r) \leq B(p_1, \ldots, p_i^{(l)}, \ldots, p_N; r)$$

$$\leq B(p_1, \ldots, p_i^{(l,k)}, \ldots, p_N; r)$$

$$\vdots$$

$$\leq B(p_1, \ldots, \theta_i, \ldots, p_N; r). \qquad (B.6)$$

Appendix 3

Let $\{p(k), k = 0, 1, \cdots\}$ be an arbitrary distribution of the number of cells arriving during $T/2$ ms, and assume that its average variances are a and σ^2. That is, $a = \sum_{k=0}^{\infty} kp(k)$ and $\sigma^2 = \sum_{k=0}^{\infty}(k-a)^2p(k)$. Let us assume that $l - 1 < a \leq l$.

We define a function $\{p^{(j)}(k), k = 0, 1, \cdots\}$, which eliminates the j-th element of $p(k)$.

$$p^{(j)}(k) = \begin{cases} 0, & \text{for } k = j, \\ p(k), & \text{for } k \neq l-1, l, j, i^*, \end{cases} \qquad (C.1)$$

where i^* is an integer defined later. In addition, we assume that $a = \sum_{k=0}^{\infty} kp^{(j)}(k)$ and $\sigma^2 = \sum_{k=0}^{\infty}(k-a)^2p^{(j)}(k)$. Thus, $p^{(j)}(l-1), p^{(j)}(l), p^{(j)}(i^*)$ are defined by the following equations.

$$p^{(j)}(l-1) + p^{(j)}(l) + p^{(j)}(i^*) = p(l-1) + p(l) + p(i^*) + p(j) \qquad (C.2)$$

$$(l-1)p^{(j)}(l-1) + lp^{(j)}(l) + i^*p^{(j)}(i^*) = (l-1)p(l-1) + lp(l) + i^*p(i^*) + jp(j) \quad (C.3)$$

$$(l-1)^2p^{(j)}(l-1) + l^2p^{(j)}(l) + i^{*2}p^{(j)}(i^*) = (l-1)^2p(l-1) + l^2p(l) + i^{*2}p(i^*) + j^2p(j)$$

$$(C.4)$$

From the above equations,

$$p^{(j)}(l-1) = p(l-1) + \frac{(j-l)(j-i^*)}{i^*-l+1}p(j), \qquad (C.5)$$

$$p^{(j)}(l) = p(l) + \frac{(j-l+1)(j-i^*)}{l-i^*}p(j), \qquad (C.6)$$

$$p^{(j)}(i^*) = p(i^*) + \frac{-(j-l)(j-l+1)}{(l-i^*)(i^*-l+1)}p(j). \qquad (C.7)$$

Then, under the assumption of $l \leq r \leq i^*$,

$$B(p^{(j)}; r) - B(p; r)$$
$$= \begin{cases} \frac{(i^*-r)(p^{(j)}(i^*)-p(i^*))+(j-r)(-p(j))}{a}, & \text{if } j > r, \\ \frac{(i^*-r)(p^{(j)}(i^*)-p(i^*))}{a}, & \text{if } j \leq r. \end{cases} \quad (C.8)$$

Note that $p^{(j)}(i^*) - p(i^*) \geq 0$ from Eq. (C.7). Thus, if $j \leq r$,

$$B(p^{(j)}; r) \geq B(p; r). \quad (C.9)$$

From Eq. (C.7), for $j > r$,

$$B(p^{(j)}; r) - B(p; r) = \frac{p(j)}{(i^* - l)(i^* - l + 1)(i^* - r)(j - r)a}(g(j) - g(i^*)), \quad (C.10)$$

where $g(z) = \frac{(z-l)(z-l+1)}{z-r}$. Therefore, if we set

$$i^* = \begin{cases} \lceil z^* \rceil, & \text{if } g(\lceil z^* \rceil) < g(\lceil z^* \rceil - 1), \\ \lceil z^* \rceil - 1, & \text{if } g(\lceil z^* \rceil) \geq g(\lceil z^* \rceil - 1), \end{cases} \quad (C.11)$$

where z^* is the minimum point of $g(z)$ larger than r, that is, i^* is set to the integer which is larger than r and makes $g(\cdot)$ minimum, then $B(p^{(j)}; r) \geq B(p; r)$. Here, z^* is given by

$$z^* = r + \sqrt{(r - l)(r - l + 1)}. \quad (C.12)$$

Consequently, for both $j \leq r$ and $j > r$, $B(p^{(j)}; r) \geq B(p; r)$.

If we set $p = p^{(j)}$ and repeat the same argument mentioned above, we obtain $B(p^{(j)(k)}; r) \geq B(p^{(j)}; r)$. Here, $p^{(j)(k)}$, which eliminates the k-th elemnt of $p^{(j)}$, that is, the j and k-th elements of p, can be defined similarly to Eq. (C.1). Finally, we obtain a function $\{q(k), k = 0, 1, \cdots\}$, which eliminates all the elements of p except for the $(l-1), l$ and i^*-th

elements.

$$q(k) = 0 \qquad \text{for } k \neq l, l - 1, i^* \quad (C.13)$$

$$q(l - 1) + q(l) + q(i^*) = 1 \quad (C.14)$$

$$(l - 1)q(l - 1) + lq(l) + i^*q(i^*) = a \quad (C.15)$$

$$(l - 1)^2 q(l - 1) + l^2 q(l) + i^{*2} q(i^*) = a^2 + \sigma^2 \quad (C.16)$$

From the above equations,

$$q(l - 1) = \frac{i^* l - a(l + i^*) + \sigma^2 + a^2}{i^* - l + 1}, \quad (C.17)$$

$$q(l) = \frac{i^*(l - 1) - a(l + i^* - 1) + \sigma^2 + a^2}{l - i^*}, \quad (C.18)$$

$$q(i^*) = \frac{l(l - 1) - a(2l - 1) + \sigma^2 + a^2}{-(l - i^*)(i^* - l + 1)}, \quad (C.19)$$

and

$$B(p; r) \leq B(q; r) = \frac{(i^* - r)\{l(l - 1) - a(2l - 1) + \sigma^2 + a^2\}}{-a(l - i^*)(i^* - l + 1)}, \quad (C.20)$$

where i^* is defined by Eq. (C.11).

Appendix 4

$$p_i^*(k) = \begin{cases} p, & \text{for } k = j \\ q, & \text{for } k = m \\ 1 - p - q, & \text{for } k = n \\ 0, & \text{otherwise} \end{cases} \quad (D.1)$$

j, m, n, p, q are determined to satisfy the average and variance specification.

$$p_i^*(k) = \begin{cases} x + y, & \text{for } k = m \\ x + z, & \text{for } k = j \\ x, & \text{for } m < k < j \\ 0, & \text{otherwise} \end{cases} \quad (D.2)$$

j, m, x, y, z are also determined to satisfy the average and variance specification.
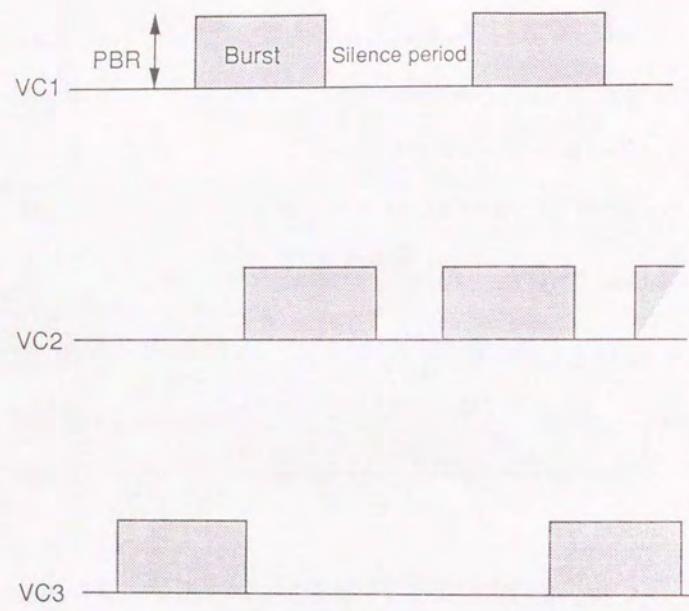
Figure 1. Burst model


$$a_1 = \frac{E[x]}{E[x]+E[y]} R_1$$
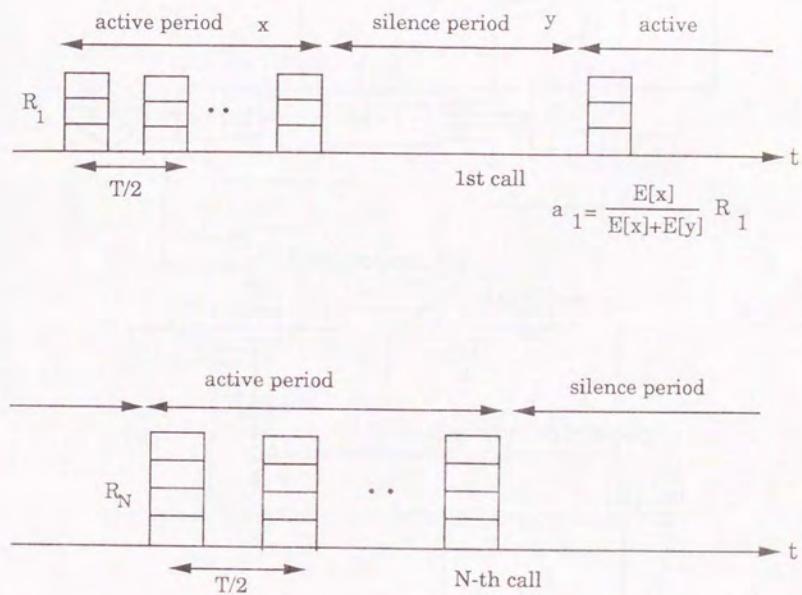
Figure 2. Most bursty process with MNA
and ANA specification
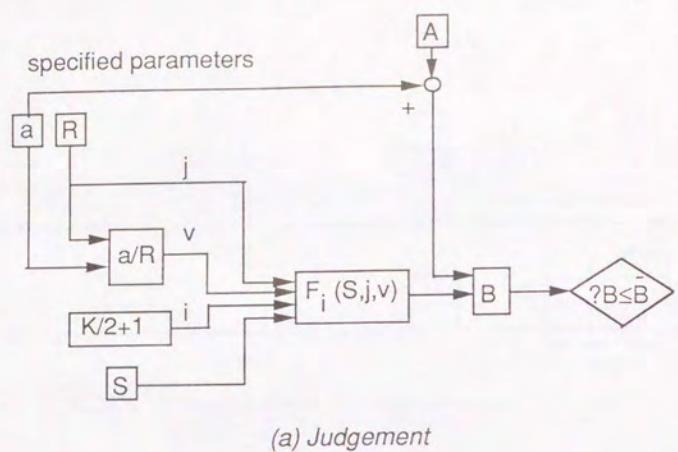
(a) Judgement
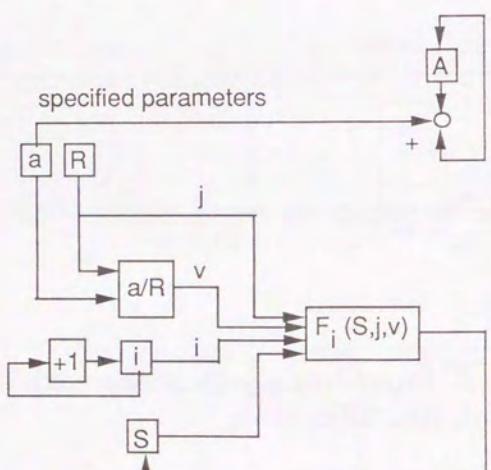
(b) Renewal of load state vector at call acceptance

Figure 3. Load state vector method
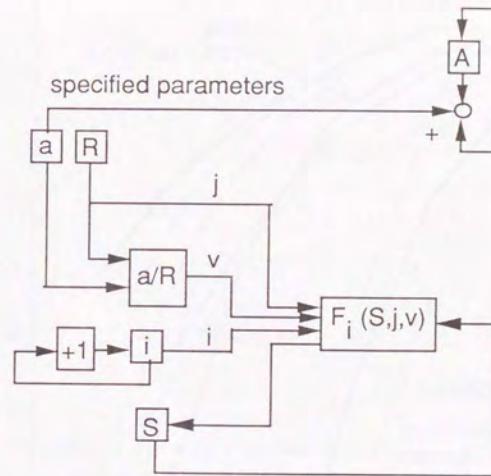
(c) Renewal of load state vector at service completion

Figure 3. Load state vector method
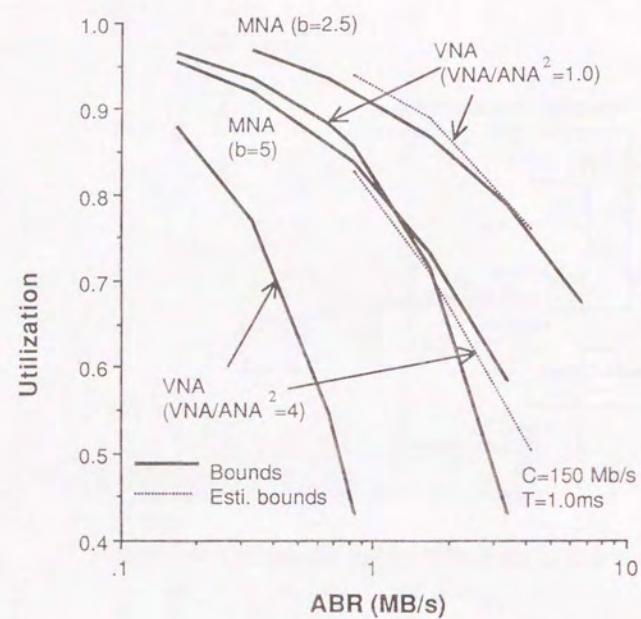
Figure 4. Utilization
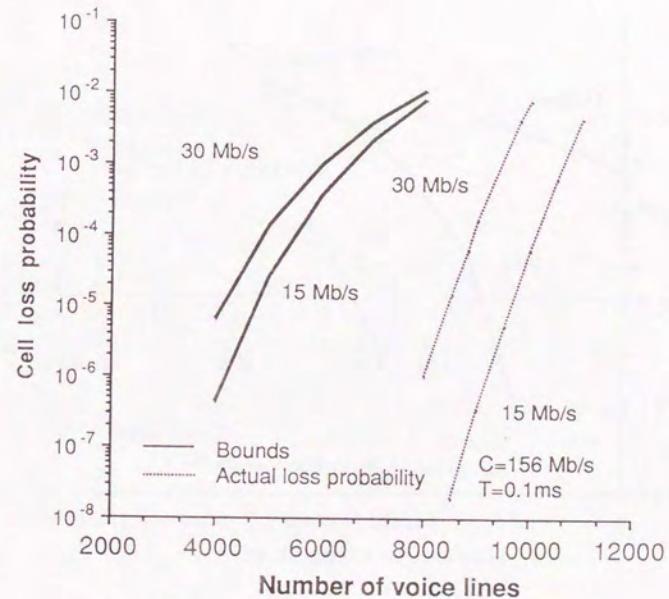


Figure 5. Cell loss probability (Case1)
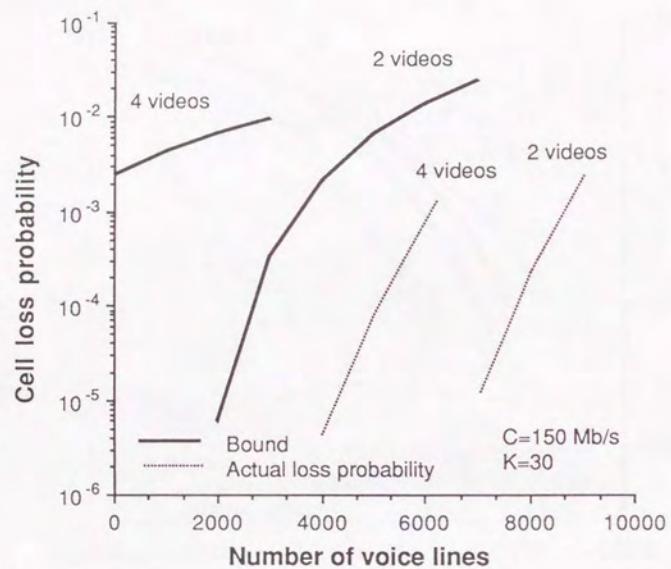
Figure 6. Cell loss probability (Case 2)
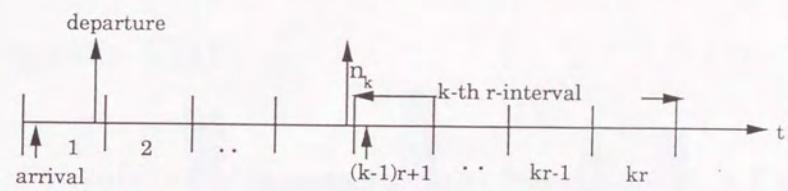
Figure A1.
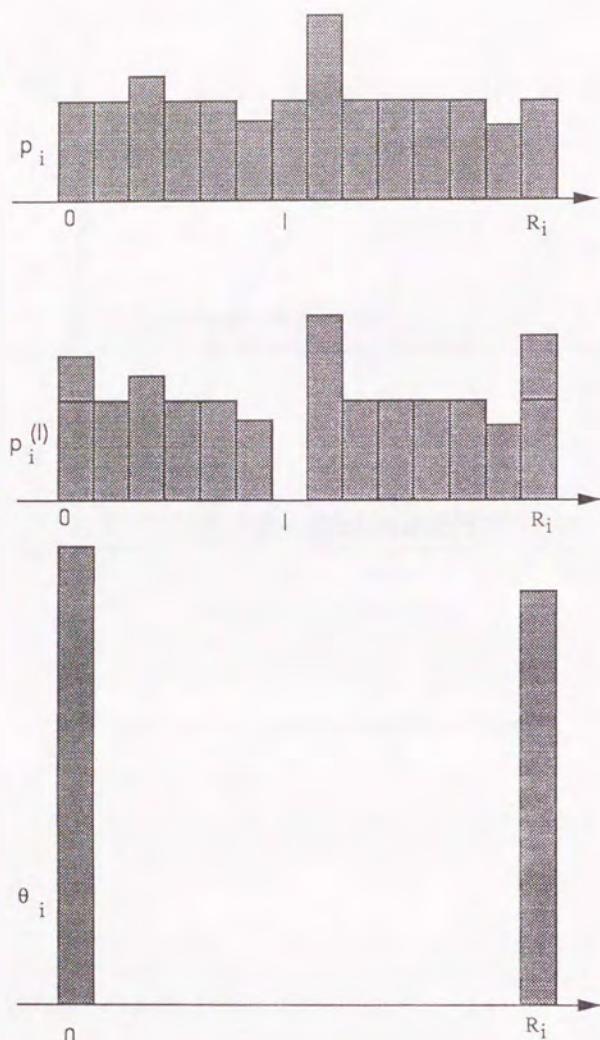
Figure B1.

Chapter VIII

A Simplified Dimensioning Method of ATM Networks

A practical dimensioning method is developed for ATM networks in which call loss probability, cell loss probability and cell delay are considered as GOS standards. This method uses the upper bound of cell loss probability, which is evaluated by the distribution of the number of arriving cells during a fixed length interval. The advantages of this method are that it is unnecessary to analyze a queueing model to evaluate the GOS performances of cells, and that the statistics used are easily measurable. If the pushout scheme is introduced, the cell loss probability standard for loss-sensitive traffic as well as that for total traffic can be satisfied by this dimensioning method. The concept of this method is also applicable to call admission control. [Saito 89c]

1. Introduction

Recently, there has been considerable interest in integrated networks [Turner 83,86b], [Saito 91a], [Hui 90]. Compared with dedicated networks, service and network integration has major advantages in development, implementation, operation, and maintenance. While dedicated networks require several distinct costly subscriber access lines, broadband ISDN (B-ISDN) access can be based on a single optical fiber for each customer. Large scale production of highly integrated system components of a B-ISDN will lead to cost-effective solutions. To meet the requirements of possible future broadband services, an integration technique is necessary that is highly flexible and can support high-speed data transmission as well as moving picture services such as video telephony or TV program distribution.

Asynchronous Transfer Mode (ATM) is expected to be a target technology for B-ISDN [Roberts 87], [Eklundh 88a], [Hui 88], [Händel 89], [Kawashima 89], and CCITT is working on standardizing ATM [CCITT 89]. In the ATM-based transport network (ATM network), all the information including voice, data and video is conveyed using a fixed sized "cell." An ATM network is a high-speed-multimedia network that handles various classes of traffic with different bit-rates and with different service quality requirements. For example, voice traffic is transmitted at several kilobits per second and is delay-sensitive. High-speed data traffic used for file transfer or LAN interconnection is transmitted at hundreds of mega-bits per second and is loss-sensitive. To achieve a high-speed-multimedia network, ATM networks adopt a simplified transport protocol based on hardware cell switching without any flow control or retransmission inside a network. As the necessary number of cells is allocated based on traffic demand, bursty information is statistically multiplexed.

Although ATM is very promising, some design and performance problems still need to be solved [Eckberg 89], [Kawashima 89]. Dimensioning of ATM networks is one problem presently under discussion [Gallassi 90]. In particular, a practical dimensioning method needs to be developed, considering the complicated Grade-of-Service (GOS) requirements and cell level statistics, as well as the call level statistics in ATM networks.

In the conventional dimensioning method for integrated networks, the arrival process is first modeled for an individual source, usually by fitting a parametric distribution to that of the interarrival time. Then the necessary statistics for the model are measured, and a model for the superposed arrival process is determined from the model for the individual sources and/or the measured statistics. Finally, the GOS items are analyzed and evaluated from the model and parameter values estimated from measurements, so that the necessary amount of resources can be determined.

If this conventional method is applied to ATM networks dimensioning, there will be several problems for the modeling and measurement, as well as for flexibility with respect to the introduction of new services. Consider the problems that occur when modeling cell arrival. It has been shown in the related literature on ATM networks that the correlation and the higher moments of the cell interarrival times affect performance. The model for cell interarrival time thus needs to be more complex. This implies that a more sophisticated knowledge of the process is necessary, and that the correlation and higher moments of the cell interarrival times need to be measured. The assumptions of the model and its accuracy, however, become difficult to verify. In addition, because of the limits of the measurement systems, the correlation and the higher moments of the cell interarrival times

are impossible to measure. Furthermore, the use of elaborate models reduces the flexibility in traffic dimensioning required for the introduction of new services, with respect to which ATM networks are particularly flexible.

The objective of the dimensioing method proposed in this paper is to overcome this problem, making it easy for traffic engineers to change the dimensioning procedure and to change the dimensioning parameters as traffic conditions change. The proposed ATM network dimensioning method can be executed by using only the distribution of the number of cells arriving during a fixed length interval and does not use elaborate modeling of a cell arrival process and complicated analysis of a queneing model. Thus, it can achieve direct feedback of traffic measurement to dimensioning and increase the flexibility of dimensioning for introducing new services.

The organization of this paper is as follows. In Section 2, a notation is introduced and the buffer size is dimensioned. In Section 3, a dimensioning method is proposed for a single traffic class and an upper bound of cell loss probability is introduced. This upper bound can be obtained from the distribution of the number of cells arriving during a fixed interval. In Section 4, the dimensioning method is extended to multiple traffic classes. If a cell loss probability control is introduced, the cell loss probability requirement for loss-sensitive traffic as well as for total traffic is also satisfied. In Section 5, numerical examples are shown to demonstrate our dimensioning method. Concluding remarks are in Section 6.

2. Preliminaries

In ATM networks, as the number of accepted calls increases, quality measures such as cell loss probability and cell delay, deteriorate. Thus, for fixed transmission link capacity, it is impossible to accept an unlimited number of calls under service quality requirements, and the number of accepted calls must be limited. In this paper, for fixed transmission link capacity, the maximum number of simultaneously accepted calls under service quality requirements is called the number of virtual circuits [Mase 88]. Since it is impossible to accept more calls than the number of virtual circuits, call loss probability can be considered as a quality measure, as in STM networks.

As stated above, we can consider service quality at different levels [Filipiak 89], [Hui 88]. We assume GOS is defined at a call level and at a cell level, and is defined by call loss probability and cell loss probability and delay (Figure 1).

Cells are delayed in ATM switches and in output queues to transmission links (output links) in ATM nodes. Cell assembly delay and propagation delay over transmission links are fixed and are independent of traffic characteristics. Cell loss due to transmission error is also independent of the traffic volume. Thus, assuming that cell loss and delay in a switch are negligible, the key factors determining cell level quality are cell loss and delay in the output queue.

Let us assume that the maximum admissible delay at each output queue is T ms, the link capacity is C Mb/s and the cell length is L bits. The buffer size K in cells is determined such that the maximum delay of a cell is less than the maximum admissible

delay T ms under an FCFS discipline; that is,

$$K = 1000TC/L. \tag{2.1}$$

3. Dimensioning for a single class of traffic

Dimensioning of ATM networks is divided into two levels: the call level and the cell level. First, the number of virtual circuits is determined in call level dimensioning, and then the link capacity that can accept the number of virtual circuits dimensioned in call level dimensioning, is provided in cell level dimensioning.

Call level dimensioning provides the number of virtual circuits S, considering the traffic characteristics and the GOS of calls. In other words, in call level dimensioning, the number of trunk circuits is determined which satisfies a call loss probability requirement $\bar{B}(call)$, for a given offered load. Such dimensioning methods have been developed for STM networks, and they can be used for ATM networks as well. For example, if a call arrives in a Poisson process at rate λ and the mean holding time is μ^{-1}, the following Erlang-B formula provides the number of virtual circuits, S.

$$\min_S \ \frac{\rho^S}{S!} \left\{ \sum_{j=0}^{S} \frac{\rho^j}{j!} \right\}^{-1} \leq \bar{B}(call), \tag{3.1}$$

where $\rho = \lambda/\mu$.

In STM networks, the necessary bandwidth is proportional to the number of trunk circuits. However, in ATM networks, it is a complicated function of the number of virtual circuits, the traffic characteristics and the GOS of the cell level. Therefore, cell level dimensioning is necessary.

Cell level dimensioning provides the necessary bandwidth C for the number of virtual circuits S obtained in a call level dimensioning such as Eq. (3.1), considering the traffic characteristics and the GOS of the cell level. Since the buffer size K is determined by Eq. (2.1), cell delay satisfies the maximum admissible delay T. Thus, only cell loss probability is considered as the GOS for dimensioning the bandwidth.

For a fixed number of virtual circuits, the GOS performances of cells can be evaluated by queueing analysis or simulation [Kawashima 89], [Saito 91a]. For example, [Tanaka 82], [Stern 83, 84], [Jenq 84], [Daigle 85, 86], [Sriram 86], [Heffes 86], [Ide 88], and [Li 88d] are available for voice traffic, and [Maglaris 88], [Ogino 88], [Sen 89], [Yamada 89a] and [Saito 91a] are available for video traffic. However, the elaborate modeling and complicated analysis required for each traffic class require too much work for dimensioning and high-performance measurement systems, and are not suitable for the development of dimensioning tools. In particular, for heterogeneous traffic, as described in the next section, queueing analysis itself is difficult [Saito 91a]. In addition, the use of elaborate models reduces the flexibility in traffic dimensioning required for the introduction of new services, with respect to which ATM networks are particularly flexible. Instead of the detailed analysis of a queueing model, a simple dimensioning method using the upper bound of the cell loss probability is proposed here.

Consider a transmission link consisting of S multiplexed virtual circuits. Let $N(t)$ be the number of cells arriving in t ms, and $F_t(\cdot)$ be the stationary distribution of $N(t)$. By Appendix 1, for any constant $\tau > 0$, the cell loss probability $B(cell)$ satisfies

$$B(cell) \leq \frac{1}{E[N(\tau)]} \int_K^\infty (z - K) dF_{T+\tau}(z). \tag{3.2}$$

Here, the term $T+\tau$, which we can arbitrarily set, represents the length of the measurement interval of the cell flow. In other words, we characterize the cell arrival process using the number of cells arriving during $T + \tau$.

Thus, for the cell loss probability requirement $\tilde{B}(cell)$, the following equation (3.3) can yield the buffer size K such that the cell loss probability requirement can be satisfied. As a result, the transmission link capacity C which satisfies the call level GOS and cell level GOS, is dimensioned using Eq. (2.1).

$$\tilde{B}(cell) = \frac{1}{E[N(\tau)]} \int_K^\infty (x - K) dF_{T+\tau}(x). \tag{3.3}$$

This completes the dimensioning for single class traffic. The advantages of this dimensioning method using Eqs. (2.1), (3.1) and (3.3) are:

- Cell level dimensioning does not require analysis of queueing models. The measured frequency distribution can be taken to be the distribution function F. Hence, parametric modeling of the arrival process and the verification of the model assumption are unnecessary. In this sense, this dimensioning method can be referred as nonparameteric.

- The cell statistic used in this method is the number of cells arriving during a fixed interval, which is easily measurable.

- From the following equation, all the statistics employed are derived from the distribution of the number of cells arriving during $T + \tau$ ms:

$$E[N(\tau)] = \frac{\tau}{T + \tau} E[N(T + \tau)]. \tag{3.4}$$

- When the number of virtual circuits S is large, the number of arriving cells $N(T + \tau)$ follows a normal distribution based on the central limit theorem. Then, the distribu-

tion is specified by the average and variance, which are obtained from the average and variance of the number of cells arriving from an individual source. The average and variance are also sufficient statistics, if the number of cells arriving from an individual source can be approximated by a normal distribution.

- If $N(T + \tau)$ follows a normal distribution with an average a and a variance σ^2, the right-hand sides of Eqs. (3.2) and (3.3) are:

$$\sigma \frac{T + \tau}{a\tau} \Theta(\frac{K - a}{\sigma}). \tag{3.5}$$

Here,

$$\Theta(u) = \int_u^\infty (y - u)\phi(y) dy, \tag{3.6}$$

and $\phi(\cdot)$ is the probability density function of the standard normal distribution. Thus, if the one-dimensional table on $\Theta(u)$ is provided, Eq. (3.3) can be evaluated by very simple calculation.

- Independence between the number of cells arriving during $[0, T + \tau]$ and that during $[T + \tau, 2(T + \tau)]$ is not necessary. (More precisely, $F_{T+\tau}$ is the distribution of the number of cells arriving during an arbitrary interval with length $T + \tau$ ms.)

When the number of virtual circuits S is small and the number of cells arriving from an individual source cannot be approximated by a normal distribution, the probability generating function of $N(t)$ can be derived from that of the number of cells arriving from an individual source [Rao 73]. Hui's result [Hui 88] is also useful for obtaining F_t when the normal approximation is poor.

The right-hand sides of Eqs. (3.2) and (3.3) depend on τ. Although we can arbitrarily set τ, that is define the interval characterizing the cell arrival process, the result

of characterization depends on the interval length $T + \tau$. Hence, if τ is appropriately set, the upper bound of the cell loss probability $B(cell)$ becomes tight. As shown in Appendix 1, the upper bound is achieved for the case where cells arrive simultaneously during τ ms. Thus, too large a value of τ gives a loose bound. Numerical examples in Section 5 show that $\tau = 0.05T$ is appropriate, while another value, for example $\tau = 0.1T$ is also acceptable. The optimal τ depends on the traffic characteristics. However, for any traffic characteristics, $\tau = 0.05T$ gives a sufficiently tight bound.

When the number of multiplexed virtual circuits S reaches infinity, utilization of a transmission link provided by Eq. (3.3) approaches $T/(T+\tau)$ for any traffic characteristics (Appendix 2). For $\tau = 0.05T$, $T/(T + \tau) = 0.952$. Thus, for a high-speed link, the dimensioning method employing Eq. (3.3) can achieve high utilization for any traffic characteristics. We can estimate the tightness of Eq. (3.2) from the lower bound of the cell loss probability in Appendix 3.

4. Dimensioning for multiple traffic classes

In this section, a method for dimensioning a transmission link shared by many classes of traffic is described. In principle, the method proposed in Section 3 can be extended. However, each class of traffic has different traffic characteristics and may require different GOS. Therefore, the dimensioning method becomes complicated.

Assume that n classes of traffic share output buffers and an output link. Traffic is divided into two types: loss-sensitive traffic such as video and data, and loss-tolerant traffic such as voice. The pushout scheme [Sumita 88], [Gravey 89], [Kröner 90], [Saito 91b], one of the space priority schemes, is used as the cell loss probability control in this paper. Here,

the pushout scheme is denoted such that when an arriving cell of loss-sensitive traffic finds no empty buffers, it pushes out the last cell of loss-tolerant traffic in waiting buffers. The pushed out cell is lost. If an arriving cell of loss-sensitive traffic finds no empty buffer and cannot find any cell loss-tolerant traffic in waiting buffers, then it is lost. Loss-tolerant cells arriving when there is no empty buffer, are discarded. The order of service is FCFS (Figure 2).

Define $N_{sensitive}(t)$ as the number of loss-sensitive cells arriving during t ms, and $F_{t,sensitive}(\cdot)$ as its stationary distribution. Define $N(t)$ as the total number of cells of both types arriving during t ms, and $F_t(\cdot)$ is its stationary distribution. Since $N_{sensitive}(t)$, $F_{t,sensitive}(\cdot)$, $N(t)$ and $F_t(\cdot)$ depend on the number of class i connected calls, or equivalently, the number of class i virtual circuits $(i = 1, \ldots, n)$, the notations $N_{sensitive}(t; s)$, $F_{t,sensitive}(\cdot; s)$, $N(t; s)$ and $F_t(\cdot; s)$ are used to clarify this fact. Here, $s = (s_1, \ldots, s_n)$, and s_i denotes the number of class i virtual circuits.

For each s and C, $\hat{B}(cell; s, C)$ defined in the following equation is shown to give the upper bound of the cell loss probability for total traffic, by the same argument employed in the derivation of Eq. (3.2).

$$\hat{B}(cell; s, C) = \frac{1}{E[N(\tau; s)]} \int_K^\infty (z - K) dF_{T+\tau}(z; s), \tag{4.1}$$

For each s and C, $\hat{B}_{sensitive}(cell; s, C)$ defined in the following equation is shown to give the upper bound of the cell loss probability for loss-sensitive traffic. Proof is in Appendix 4.

$$\hat{B}_{sensitive}(cell; s, C) = \frac{1}{E[N_{sensitive}(\tau; s)]} \int_K^\infty (z - K) dF_{T+\tau, sensitive}(z; s). \tag{4.2}$$

Thus, if $\hat{B}_{sensitive}(cell; \mathbf{s}, C) \leq \tilde{B}_{sensitive}(cell)$ and $\hat{B}(cell; \mathbf{s}, C) \leq \tilde{B}(cell)$, then $(s_1, \ldots, s_n)$ is acceptable with transmission link capacity C, where $\tilde{B}_{sensitive}(cell)$ and $\tilde{B}(cell)$ are the cell probability requirement of loss-sensitive traffic and that of total traffic. Hence, for each fixed C, we can define $\Omega(C)$, the set of the number of virtual circuits which satisfy the cell loss probability requirements, as follows.

$$\Omega(C) \stackrel{\triangle}{=} \{\mathbf{s}|\hat{B}_{sensitive}(cell; \mathbf{s}, C) \leq \tilde{B}_{sensitive}(cell), \hat{B}(cell; \mathbf{s}, C) \leq \tilde{B}(cell)\}. \qquad (4.3)$$

Then, the dimensioning algorithm can be obtained as follows.

4.1 Algorithm

Step 0. *(Initialization)*

For class i traffic, the transmission capacity $C_0(i)$ is dimensioned according to the method in Section 3, assuming that the transmission link is dedicated to class i traffic, considering the traffic characteristics and the GOS of call and cell levels. Set $C_0 = max_i \, C_0(i)$, $C = C_0$, $C_{min} = C_0$ and $C_{max} = \sum_{i=1}^{n} C_0(i)$.

Step 1. *(Cell level dimensioning)*

Use Eqs. (4.1) and (4.2), and obtain $\Omega(C)$ as defined in Eq.(4.3).

If $C \geq \sum_{i \in \{loss-sensitive\}} C_0(i)$, $\Omega(C)$ reduces to

$$\Omega(C) = \{\mathbf{s}|\hat{B}(cell; \mathbf{s}, C) \leq \tilde{B}(cell)\}. \qquad (4.4)$$

Step 2. *(Call level dimensioning)*

Evaluate $B_i(call)$, the loss probability of a class i call, for a transmission link capacity C. If class i calls arrive in a Poisson process at a rate λ_i and their mean holding time is

μ_i^{-1}, then

$$B_i(call) = \sum_{\mathbf{s} \in \Gamma_i(C)} \frac{\rho_1^{s_1} \cdots \rho_n^{s_n}}{s_1! \cdots s_n!} \Big\{ \sum_{\mathbf{s} \in \Omega(C)} \frac{\rho_1^{s_1} \cdots \rho_n^{s_n}}{s_1! \cdots s_n!} \Big\}^{-1}. \qquad (4.5)$$

Here, $\Gamma_i = \{\mathbf{s}|\mathbf{s} \in \Omega(C), \mathbf{s} + \mathbf{e}_i \notin \Omega(C)\}$, and $\mathbf{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ is the i-th unit vector and all its elements are 0 except for the i-th element. That is, Γ_i is a boundary set of $\Omega(C)$.

If $B_i(call)$ satisfies the call loss probability requirement $\tilde{B}_i(call)$ for all i, set $C_{max} = C$, $C_{new} = C - \epsilon$ and go to Step 3. Here, ϵ is a small constant incremental value for searching a link capacity.

Otherwise, set $C_{min} = C, C_{new} = min\{(1 + max_i[B_i(call) - \tilde{B}_i(call)]^+)C, C_{max}\}$ and go to Step 3. Here, $max_i[B_i(call) - \tilde{B}_i(call)]^+ C$ is an approximation of the necessary additional bandwidth.

$[x]^+$ denotes that $[x]^+ = x$ for $x \geq 0$, $[x]^+ = 0$ for $x < 0$.

Step 3. *(Judgement)*

If $C_{max} - C_{min} \leq \epsilon$, stop this algorithm and adopt $C = C_{max}$ as the obtained transmission link capacity. Otherwise, set $C = C_{new}$ and go to Step 1. ∎

This algorithm has two advantages: it guarantees convergence, and that all the GOS requirements are satisfied. That is, the cell loss probability of loss-sensitive traffic as well as that of total traffic, cell delay and call loss probability for each class are less than the required levels.

This algorithm assumes that the pushout scheme is used to discriminate between two cell loss probability requirement classes. (Precisely speaking, Eq. (4.2) is derived under the

assumption that the pushout scheme is used.) For a single cell loss probability requirement class, $\Omega(C)$ is defined by Eq. (4.4), and the above algorithm is applicable.

It is also possible to extend the pushout scheme to more than two types. A class i call has the cell loss priority ξ_i, and can push out the cells with cell loss priority ξ_j, $\xi_i < \xi_j$. That is, the higher value of the cell loss priority means low priority. In this case, $\Omega(C)$ is defined as,

$$\Omega(C) \triangleq \{s | \hat{B}_1(cell; s, C) \le \tilde{B}_1(cell), \ldots, \hat{B}_m(cell; s, C) \le \tilde{B}_m(cell)\}, \qquad (4.6)$$

where $m = max_i \xi_i$, and

$$\hat{B}_i(cell; s, C) \triangleq \frac{1}{E[N_i(\tau; s)]} \int_K^\infty (z - K) dF_{T+\tau,i}(z; s). \qquad (4.7)$$

$N_i(t)$ is the number of arriving cells with priority number less than $i + 1$ during t ms, and $F_{t,i}(\cdot)$ is its stationary distribution.

Then, the proposed algorithm guarantees that the loss probability requirement of cells with priority number less than $i + 1$, $1 \le i \le m$, as well as call loss probabilities of individual classes, and maximum admission delay, are satisfied.

5. Numerical examples

In this section, two numerical examples are given, to clarify the proposed dimensioning method.

In the first example, comparisons are made between the cell loss probability evaluated by Eq. (3.2) and that evaluated by a queueing analysis with a fluid flow approximation [Maglaris 88], to examine the tightness of the upper bound of the cell loss probability provided by Eq. (3.2). In the second example, an ATM transmission link capacity shared by voice and video is dimensioned.

In the following, the cell length is assumed to be 53 octets including a 5-octet header, as agreed in CCITT standard.

5.1 Example 1.

In Example 1, cell loss probability of video traffic is evaluated by Eq. (3.2), and it is compared with that evaluated by a fluid flow approximation [Maglaris 88]. For video, there are two important measures of delay: absolute delay and delay variation (relative delay). Since queueing delay in an output queue is the main factor in delay variation, the delay variation requirement is satisfied by limiting the maximum admissible delay to below a certain specified value. In addition, the absolute delay requirement can also satisfied by limiting the maximum admissible delay.

It is assumed that the distribution of the bit rate in an individual video source is normal with an average of 3.9 Mb/s and a standard deviation of 1.725 Mb/s, following [Maglaris 88]. In [Maglaris 88], the intergeneration time of cells within a frame is not identified. In this paper, it is assumed as in [Ogino 88] that the intergeneration time of

cells within a frame is constant. Then, we can evaluate $E[N(T + \tau)]$ and $V[N(T + \tau)]$ for any $T + \tau$, using the fact that the correlation of bit rate between frames with a lag of t seconds is given by $e^{-3.9t}$ [Maglaris 88].

In Figure 3, to compare with the result obtained from [Maglaris 88], we assume that the maximum admissible delay $T = 30$ ms and that the link capacity is 52 Mb/s (SONET STS-1 rate). For 7 video sources and a cell loss probability requirement of 10^{-6} to 10^{-7}, utilization of a link in [Maglaris 88] is about 0.6, which is about 10% higher than the result obtained by our method. Also, using 10 video sources, link utilization shown in [Maglaris 88] is about 10% higher than our result. In other words, link capacity increases about 10% for the same offered load, if our dimensioning method is used instead of their result. In Figure 4, it is assumed that $T = 60$ ms and that link utilization is 0.8. In comparison with [Maglaris 88], it is shown that Eq. (3.3) gives a larger cell loss probability by a factor of 10 or 100. However, our simple method does not require the eigenvalue analysis employed in their fluid flow analysis. Furthermore, although Maglaris et al. elaborately modeled the packetized video souce, such an effort is unnecessary here.

In this example, the number of multiplexed virtual circuits is small. As described in Section 3, when the number of multiplexed virtual sources is larger, utilization approaches $T/(T + \tau) \approx 1$. Thus, for more multiplexed video sources, the result is improved.

In Example 2, $\tau = 0.05T$ is used, but we also conclude from the result that $\tau = 0.01T$ and $\tau = 0.1T$ are also applicable. In particular, for heavy loads, the result becomes insensitive to τ.

5.2 Example 2.

In this example, the link capacity shared by voice and video traffic is dimensioned. It is assumed that both the voice and video calls arrive in a Poisson process, and that the offered load of video traffic $\rho_{video} = 5$ erl, and that of voice traffic $\rho_{voice} = 1000$ erl.

The video traffic is assumed to be loss-sensitive and the voice traffic is assumed to be loss-tolerant. The GOS requirements assumed in video and voice traffic are: the maximum admissible delay of a cell $T=1$ ms, the cell loss probability requirement of video $\tilde{B}_{sensitive}(cell) = 10^{-5}$, the cell loss probability requirement of total traffic $\tilde{B}(cell) = 10^{-3}$, the call loss probability requirement of video $\tilde{B}_{video}(call) = 10^{-1}$, and the call loss probability requirement of voice $\tilde{B}_{voice}(call) = 10^{-2}$.

The assumption on statistics for video in this example is the same as that in Example 1. For the voice traffic, speech activity detection is used: the mean talkspurt length is 650 ms and the mean of silence periods is 352 ms [Heffes 86]. (Exponential distributions are assumed for talkspurt lengths and silence periods in many works to simplify queueing analysis, but the field data do not always support these assumptions [Brady 69], [Gruber 82], [Yatsuzuka 82b], [Lee 86]. Such assumptions on talkspurt length distribution and the silence period length distribution are unnecessary in this paper. Thus, this dimensioning method is robust for these distributions, which is not the case for the queueing analysis [O'Reilly 87a].) The 32Kb/s ADPCM encoding scheme is employed, and 53-octet cells including a 5-octet header are generated every 12 ms during a talkspurt. Set $\tau = 0.05T = 0.05$ ms, and the value of ϵ used in the Algorithm is 1.0 Mb/s.

At Step 0, the necessary number of virtual circuits for voice traffic is 1030, and that

for video traffic is 8. To satisfy the GOS requirements at the cell level, the transmission link capacity for voice is 22 Mb/s and that for video is 54.3 Mb/s. Thus $C_0 = 54.3$ Mb/s.

At Step 1, for $C = 54.3$ Mb/s, $\Omega(C)$ is obtained (Figure 5). The call loss probabilities $B_{voice}(call)$ and $B_{video}(call)$ are 0.39×10^{-5} and 0.19, as evaluated in Step 2. $C_{new} = 59.3$ Mb/s, and again, at Step 1, for $C = 59.3$ Mb/s, $\Omega(C)$ is derived (Figure 5). The cell loss probabilities $B_{voice}(call) = 0.22 \times 10^{-5}$ and $B_{video}(call) = 0.12$. After that, the algorithm runs at $C = 60.5, 61.7, 62.9, 61.9$ Mb/s (Figure 5). Thus, solution is $C = 62.9$ Mb/s. As indicated in [Hui 88], the boundary of $\Omega(C)$ is almost linear.

Figure 6 shows the case where $T = 3$ ms. The solution is $C = 61.7$ Mb/s and the maximum admissible delay T is fairly insensitive to the solution.

6. Conclusions

This paper proposed a dimensioning method using the distribution of the number of cells arriving in $T + \tau$ ms, considering call loss probability, maximum admissible cell delay and cell loss probability. This method does not require complicated queueing analysis or elaborate modeling of the arrival processes. Thus, this method can achieve direct feedback of traffic measurement to dimensioning and increase the flexibility of dimensioning for introducing new services.

The concept used in this dimensioning method is also suitable for a call admission control. If the right-hand side of Eq. (3.2) is less than the cell loss probability requirement $\tilde{B}(cell)$ after accepting a required call attempt, then accept the call. Otherwise, reject the call. It is an advantage that only the one-dimensional table on $\Theta(u)$ is sufficient to achieve call admission control under the assumption of a normal distribution.

Appendix 1.

Let $n(t)$ denote the number of cells in the system including the one being transmitted at t, $l(s,t)$ denote the number of cells lost during $[s,t)$, and $a(s,t)$ denote the number of arriving cells during $[s,t)$. Then,

$$n(t) \leq [n(t - T) - K]^+ + a(t - T, t)$$

$$\leq a(t - T, t) + 1, \tag{A.1}$$

$$l(i\tau, (i+1)\tau) \leq [n(i\tau) + a(i\tau, (i+1)\tau) - K - 1]^+$$

$$\leq [a(i\tau - T, (i+1)\tau) - K]^+. \tag{A.2}$$

For $0 < s \leq \tau$,

$$l(i\tau, i\tau + s) \leq l(i\tau, (i+1)\tau). \tag{A.3}$$

Thus, for $k\tau < t \leq (k+1)\tau$,

$$l(0, t) \leq l(0, (k+1)\tau)$$

$$= \sum_{i=0}^{k} l(i\tau, (i+1)\tau)$$

$$\leq \sum_{i=0}^{k} [a(i\tau - T, (i+1)\tau) - K]^+, \tag{A.4}$$

$$\frac{l(0,t)}{a(0,t)} \leq \frac{1}{k+1} \left(\frac{a(0,t)}{k+1} \right)^{-1} \sum_{i=0}^{k} [a(i\tau - T, (i+1)\tau) - K]^+$$

$$\leq \left(\frac{a(0, k\tau)}{k} \frac{k}{k+1} \right)^{-1} \frac{1}{k+1} \sum_{i=0}^{k} [a(i\tau - T, (i+1)\tau) - K]^+. \tag{A.5}$$

Therefore, under the assumption of the ergodicity, for $t \to \infty$,

$$B(cell) \leq \frac{1}{E[N(\tau)]} \int_K^\infty (z - K) dF_{T+\tau}(z). \tag{A.6}$$

Appendix 2.

For large S, $N(T + \tau)$ follows a normal distribution, because of the central limit theorem. Assume that the average of an individual source is a and that its variance is σ^2. That is, $N(T + \tau)$ is distributed as a normal distribution with average Sa and variance $S\sigma^2$. Thus, Eq. (3.3) reduces to

$$\Theta\left(\frac{\sqrt{S}\delta_s}{\sigma}\right) = \tilde{B}(cell)\frac{\sqrt{S}a\tau}{\sigma(T + \tau)}. \tag{B.1}$$

Here, $\Theta(\cdot)$ is defined in Eq. (3.6), and

$$\delta_s \stackrel{\triangle}{=} \frac{K}{S} - a. \tag{B.2}$$

Note that $\frac{\partial\Theta(u)}{\partial u} < 0$ for $u \geq 0$. The right-hand side of Eq. (B.1) is an increasing function of S. Therefore, $\frac{\sqrt{S}\delta_s}{\sigma}$ is a decreasing function of S. Consequently, δ_s is a decreasing function of S, and $\delta_s \to 0$ when $S \to \infty$.

Thus, for sufficiently large S,

$$\frac{1000TC}{L} = aS. \tag{B.3}$$

Hence,

$$C = \frac{SaL}{1000T} = \frac{SLE[N(T)]}{1000T}\frac{T + \tau}{T}. \tag{B.4}$$

Here, $\frac{SLE[N(T)]}{1000T}$ is offered load, and as a result, $\frac{SLE[N(T)]}{1000TC}$ is the link utilization.

Appendix 3.

Let $i(0, t)$ be the total length of the idle period of the transmission link during $[0, t)$, and $u(t)$ be the unfinished work of cells in the system at t. Then,

$$u(0) - u(t) + (a(0, t) - l(0, t))L/(1000C) = t - i(0, t). \tag{C.1}$$

For $kT \leq t < (k + 1)T$,

$$l(0, t) \geq l(0, kT)$$

$$= [\{u(0) - u(kT) - kT + i(0, kT)\}1000C/L + a(0, kT)]^+. \tag{C.2}$$

Thus, using Eq. (2.1),

$$\frac{l(0, t)}{a(0, t)} \geq \left[\frac{(u(0) - u(kT) + i(0, kT))K/T}{a(0, t)} + \frac{a(0, kT)/k - K}{a(0, t)/k}\right]^+. \tag{C.3}$$

For sufficiently large t,

$$B(cell) \geq \frac{1}{E[N(T)]}\int_K^\infty (z - K)dF_T(z). \tag{C.4}$$

In particular, for heavy load, that is, $\frac{i(0, kT)K}{a(0, t)T} = 0, \quad (t \to \infty)$,

$$B(cell) = \frac{1}{E[N(T)]}\int_K^\infty (z - K)dF_T(z). \tag{C.5}$$

Appendix 4.

Let $n_{sensitive}(t)$ denote the number of cells of loss-sensitive traffic in the system at t, $l_{sensitive}(s,t)$ denote the number of lost cells of loss-sensitive traffic during $[s,t)$, and $a_{sensitive}(s,t)$ denote the number of arriving cells of loss-sensitive traffic during $[s,t)$. Define indicator variables, $s_{sensitive}$ and $s_{tolerant}$, as follows. If a cell of loss-sensitive traffic is served at t, $s_{sensitive}(t) = 1$; otherwise $s_{sensitive}(t) = 0$. If a cell of loss-tolerant traffic is served at t, $s_{tolerant}(t) = 1$; otherwise $s_{tolerant}(t) = 0$.

$$n_{sensitive}(t) \leq [n(t-T) - K]^+ + a_{sensitive}(t-T,t)$$

$$\leq a_{sensitive}(t-T,t) + s_{sensitive}(t), \qquad (D.1)$$

$$l_{sensitive}(i\tau,(i+1)\tau) \leq [n_{sensitive}(i\tau) + a_{sensitive}(i\tau,(i+1)\tau) + s_{tolerant}(i\tau) - K - 1]^+$$

$$\leq [a_{sensitive}(i\tau - T,(i+1)\tau) - K]^+. \qquad (D.2)$$

For $0 < s \leq \tau$,

$$l_{sensitive}(i\tau, i\tau + s) \leq l_{sensitive}(i\tau,(i+1)\tau). \qquad (D.3)$$

Thus, for $k\tau < t \leq (k+1)\tau$,

$$l_{sensitive}(0,t) \leq l_{sensitive}(0,(k+1)\tau)$$

$$= \sum_{i=0}^{k} l_{sensitive}(i\tau,(i+1)\tau)$$

$$\leq \sum_{i=0}^{k} [a_{sensitive}(i\tau - T,(i+1)\tau) - K]^+, \qquad (D.4)$$

$$\frac{l_{sensitive}(0,t)}{a_{sensitive}(0,t)} \leq \frac{1}{k+1} \left(\frac{a_{sensitive}(0,t)}{k+1} \right)^{-1} \sum_{i=0}^{k} [a_{sensitive}(i\tau - T,(i+1)\tau) - K]^+$$

$$\leq \left(\frac{a(0,k\tau)}{k} \frac{k}{k+1} \right)^{-1} \frac{1}{k+1} \sum_{i=0}^{k} [a_{sensitive}(i\tau - T,(i+1)\tau) - K]^+ (D.5)$$

Therefore, under assumption of the ergodicity, for $t \to \infty$, the cell loss probability of loss-sensitive traffic $B_{sensitive}(cell)$ satisfies,

$$B_{sensitive}(cell) \leq \frac{1}{E[n_{sensitive}(\tau)]} \int_K^\infty (x - K) dF_{T+\tau,sensitive}(x). \qquad (D.6)$$
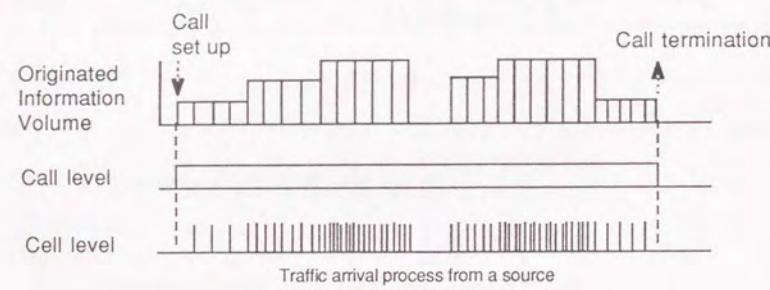
Figure 1. Call and cell levels

loss-insensitive cells ○

loss-sensitive cells ●

Figure 2. Pushout scheme

52Mb/s transmission link
T=30ms

cell loss probability

number of video sources (utilization)

□ τ=0.1T
■ τ=0.05T
○ τ=0.01T
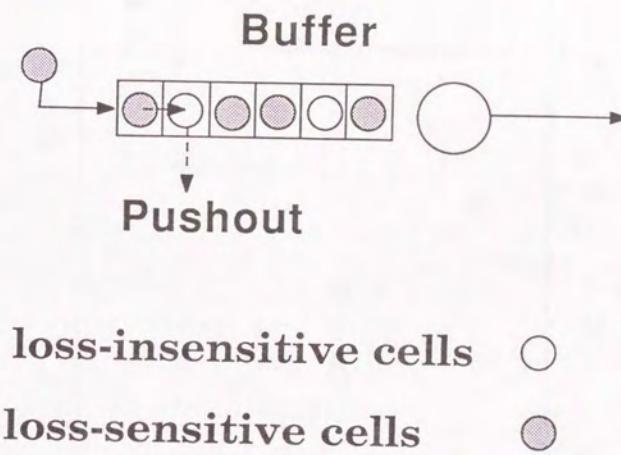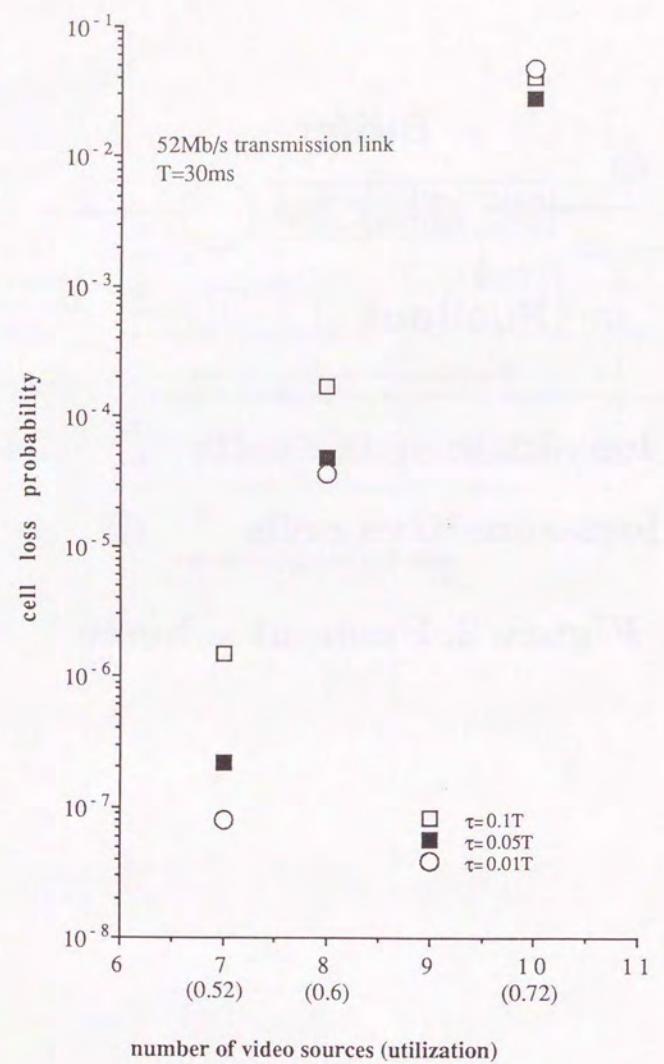
Figure 3. Cell loss probability

cell loss probability

video lines

■ τ =0.1T
□ τ =0.05T
▲ τ =0.01T
△ exact
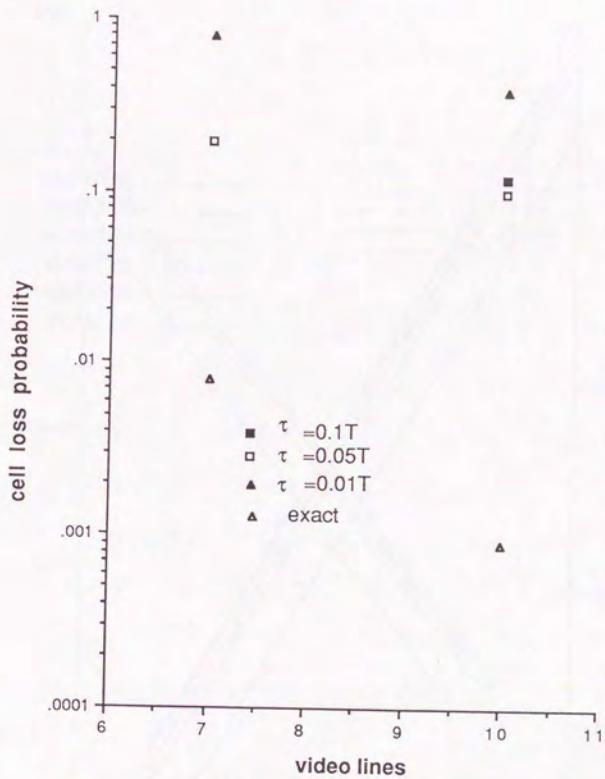
Figure 4. Cell loss probability (T=60 ms)
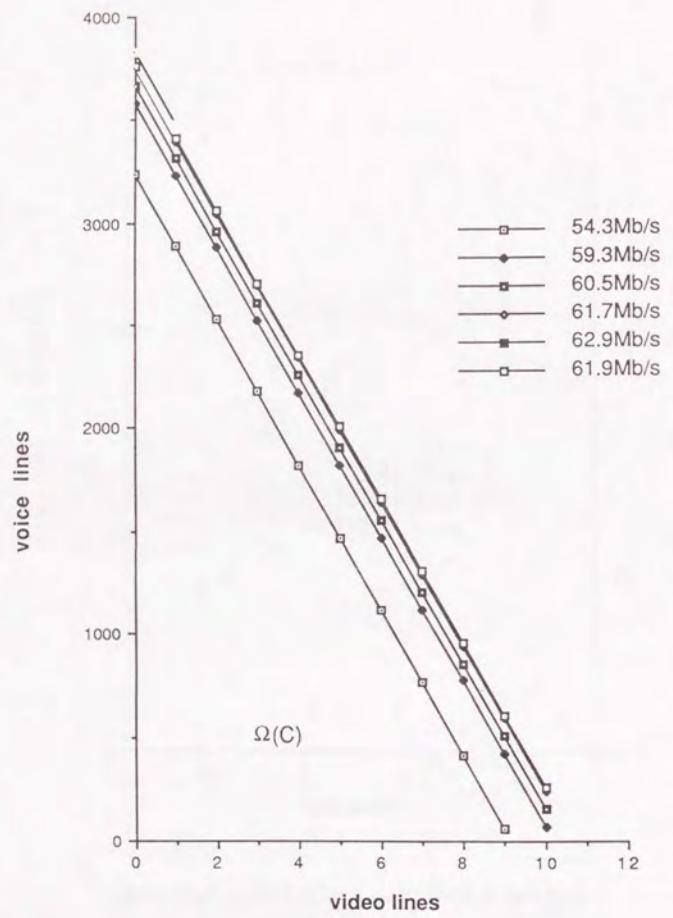
Figure 5. Number of voice and video lines (T=1 ms)
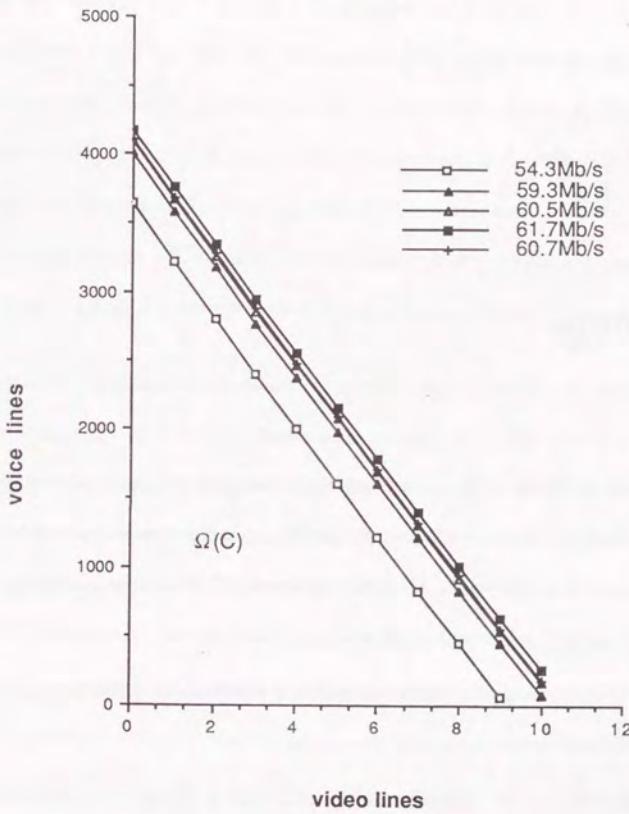
Figure 6. Number of voice and video lines (T=3 ms)

Chapter IX

Conclusions

ATM based B-ISDN will offer several benefits: adaptability to new services with different bandwidth requirements and to changing traffic, an integrated internal architecture, and a high transmission efficiency. However, although ATM is very promising, traffic design and traffic control do present problems.

This thesis focused on control and dimensioning problems in ATM networks with a synthetic approach and a non-parametric approach.

The optimal delay quality control, the optimal control for selective cell discarding, and a call admission control were synthesized. The optimal delay quality control which minimizes the number of cells delayed beyond the maximum allowable time specified for an individual call class was derived, without any assumptions on the cell arrival process or buffer management schemes. Implementation of the optimal discipline was also discussed.

The optimal control for selective cell discarding was obtained. Under embedded coding

of voice signals, cells of sampled voice signals were found to have different significances. That is, one cell may have contained bits of higher significance than that of another cell. When a network is congested, the bits are dropped by discarding cells at vocoders or multiplexers located at the network entry point and at the ATM switching nodes within the network. Thus, there is a trade-off between deterioration in voice quality and reduction of congestion. The optimal control was shown to maximize the long-run average coding rate under an average cell queue length constraint. The synthesized optimal control was shown to have a simple structure, and to be the random selection between two feedbacks.

A call admission control based only on parameters specified by users was also proposed. The key technology here is a non-parametric evaluation of the upper bound of cell loss probability. This non-parametric evaluation is done using only the specified parameters and other assumptions, as other parameters and modeling are not needed. The proposed call admission control rejects connection requests when the evaluated upper bound of cell loss probability exceeds the cell loss probability standard. Thus, the cell loss probability standard is guaranteed to be satisfied under this control. Implementation of this control to quickly evaluate cell loss probablity after acceptance of a new call was discussed. When there is no information on the cell arrival processes except for the specified parameters, the proposed call admission control is available for use in the dimensioning of ATM networks.

A dimensioning method for ATM networks was also developed with a way of thinking similar to that used to derive the call admission control. This proposed dimensioning method is 'non-parametric', and is further divided into call-level dimensioning and cell-level dimensioning. Call-level dimensioning provides a number of virtual circuits, taking into

consideration the call loss probability standard. Cell-level dimensioning yields both output buffer size and output link capacity, taking into consideration the cell loss probability standard and the maximum admissible delay. This dimensioning method is applicable to multiple traffic classes. It employs the probability density function of the number of cells arriving at a fixed interval. However, the p.d.f. need not be parameterized. The measured frequency distribution can be used as the p.d.f. This is not the case in the ordinary dimensioning method based on the queueing theory.

The proposed controls and dimensioning method were disscussed for an ATM node. However, if burstiness is reduced by concatenating several nodes in an ATM network, the discussions described for an entry node may be inapplicable to that for a node within a network. To examine this phenomenon, the cell arrival process at the input and output sides of a node have to be compared. This thesis investigates the departure process of an N/G/1 queue. The N-process, a versatile point process, can model video/voice cell arrival processes. Thus, the smoothing effect of passing through a node can be quantatively evaluated with the theory developed in this thesis. In [Kawarazaki 90], the results presented in this thesis were directly applied and the following conclusion was obtained: When transmission efficiency is low, which is the case when video traffic GOS standards are fulfilled, burstiness is not reduced by passing through nodes. Consequently, the argument for one node can be applied to all nodes in a network, and it is sufficient to consider a single node.

This thesis focuses on ATM networks. However, non-parametric and synthetic approaches are also available for other applications.

In future integrated networks, it is expected that many new services will be introduced and that assumptions on traffic will be difficult to define. Thus, non-parametric approaches will become more important. A problem with the non-parametric approach is that the obtained performance measure is wide-ranging. Hence, critical measurement items for non-parametric dimensioning should be investigated, and results obtained from the investigation should be reflected in the development of the measurement systems.

Looking ahead, a problem with the synthetic approach is that it sometimes yields a control which is difficult to be realized by a simple hardware configuration. Therefore, a synthetic approach which yields a class of control schemes should be developed. Only when the control which can be easily realized is found, it is adopted in the class.

Acknowledgements

I owe a particular debt of gratitude to Professor Toshiyuki Kitamori of the University of Tokyo, who led me towards the completion of this thesis.

This research was conducted at the NTT Laboratories. I am greatly indebted to Professor On Hashida of Tsukuba University (former Research Fellow in the NTT Labs), Mr. Konosuke Kawashima and Dr. Tohru Ueda, members of the NTT Labs, who introduced me to this field and gave me continuous guidance and encouragement. I would also like to express my appreciation to Dr. Hiroshi Ishikawa and Dr. Shuji Tomita, who gave me the opportunity to develope this thesis. Special thanks are due to Professor Kunio Kodaira of the Kanagawa Institute of Technology (former Head of Teletraffic Section), Dr. Fumiaki Machihara, Dr. Shuichi Sumita, Mr. Masatoshi Kawarazaki, Dr. Yoshitaka Takahashi, Mr. Ichiro Ide for their helphul discussions on teletraffic design and control. I also thank to Dr. Jun Matsuda and Mr. Takeo Abe for their guidance to teletraffic engineering. Finally, I would like to thank Dr. Shin-ya Nogami, Dr. Frank Brochin, Mr. Hiroshi Yamada, Mr. Hideaki Yoshino, Mrs. Youko Hoshiai, Mr. Shinsuke Shimogawa and Mr. Toshihisa Ozawa, and other members of the NTT Labs for their daily discussions.

References

[Aboul-Magd 88] Aboul-Magd, O. S. and Leon-Garcia, A., "Performance analysis of a finite buffer burst switched node," INFOCOM88, pp. 669-677(1988).

[Ahmadi 86] Ahmadi, H. and Stern, T. E., "A combined fixed and demand assignment satellite multiple access protocol for integrated circuit and packet switching," IEEE International Conference on Communications 1986, pp. 17.5.1-7(1986).

[Akiyama 86] Akiyama, M., Tanaka,Y. and Hamada, T., "Performance analysis DSI voice and data mixed traffic handling system," Trans. IECE Japan, J69B, 5, pp. 442-448(1968) (in Japanese).

[Amstutz 83] Amstutz, S. R., "Burst switching - an introduction," IEEE Communications Magazine, 21, 8, pp. 36-42(1983).

[Anick 82] Anick, D., Mitrani, D. and Sondhi, M., "Stochastic theory of a data handling system with multiple sources," Bell Syst. Tech. J., 61, 8, pp. 1871-1894(1982).

[Arai 89] Arai, H., Kawarazaki, M. and Nogami, S., "An analysis of call admission control in the ATM network," Trans. IEICE, J72-B-I, 11, pp. 1000-1007 (1989)(in Japanese).

[Arthurs 79] Arthurs, E. and Stuck, B. W., "A theoretical traffic performance analysis of an integrated voice-data virtual circuit packet switch," IEEE Trans. on Communications, 27, 7, pp. 1104- 1111(1979).

[Arthurs 83] Arthurs, E. and Stuck, B. W., "Traffic analysis tools for integrated digital time-division link level multiplexing synchronous and asynchronous message streams," IEEE J. Selected Areas in Communications, 1, 6, pp.1112-1123(1983).

[**Avellaneda 82**] Avellaneda, O. A., Hayes, J. F. and Nassehi, M. M., "A capacity allocation problem in voice-data networks," IEEE Trans. on Communications, **30**, 7, pp. 1767-1772(1982).

[**Bagchi 85**] Bagchi, U. and Sullivan, R. S., "Dynamic non-preemptive priority queues with general, linear increasing priority function," Operations Research, **33**, 6, pp. 1278-1298 (1985).

[**Barberis 80**] Barberis, G. and Pazzaglia, D., "Analysis and optimal design of a packet-voice receiver," IEEE Trans. on Communications, **28**, 2, pp. 217-227(1980).

[**Barberis 81**] Barberis, G., "Buffer sizing of a packet-voice receiver," IEEE Trans. on Communications, **29**, 2, pp. 152-156(1981).

[**Bially 80a**] Bially, T., Gold, B. and Seneff, S., "A technique for adaptive voice flow control in integrated packet networks," IEEE Trans. on Communications, **28**, 3, pp.325-333(1980).

[**Bially 80b**] Bially, T., McLaughlin, A. J. and Weinstein, C. J., "Voice communicaion in integrated digital voice and data networks," IEEE Trans. on Communicaions, **28**, 9, pp. 1478-1490(1980).

[**Boyer 87**] Boyer, J., Louvion, J. -R. and Romoeuf, L., "Modelling the ATD transfer technique," 5th ITC Seminar: Traffic Engineering for ISDN Design and Planning(1987).

[**Brady 65**] Brady, P. T., "A technique for investigating on-off patterns of speech," Bell Syst. Tech. J., **44**, pp. 1-22(1965).

[**Brady 68**] Brady, P. T., "A statistical analysis of on-off patterns in 16 conversations," Bell Syst. Tech. J., **47**, pp. 73- 91(1968).

[**Brady 69**] Brady, P. T., "A model generating on-off speech patterns in two-way conversation," Bell Syst. Tech. J., **48**, pp. 2445- 2472(1969).

[**Bris 86**] Bris, H. L. and Servel, M., "Integrated wideband networks using asynchronous time division techniques," IEEE International Conference on Communications 1986, pp. 1720-1724(1968).

[**Bullington 59**] Bullington, K. and Fraser, J. M., "Engineering aspects of TASI," Bell Syst. Tech. J., **38**, pp. 353-364(1959).

[**Burgin 89a**] Burgin, J., "Dynamic capacity management in the B-ISDN," Fourth Australian Teletraffic Research Seminar (1989).

[**Burgin 89b**] Burgin, J., "Broadband ISDN resource management," ITC Specialist Seminar, Addelaide, pp. 12.2.1-11(1989).

[**Burke 56**] Burke, P. J., "The Output of a Queueing System," Operations Research, **4**, pp. 699-704(1956).

[**Buyukkoc 85**] C. Buyukkoc, P. Varaiya and J. Warlrand, "The $c\mu$ rule revisited," Adv. Appl. Prob., **17**, pp. 237-238(1985).

[**Campanella 76**] Campanella, S. J., "Digital speech interpolation," COMSAT Tech. Rev., **6**, 1, pp. 127-158(1976).

[**CCITT 84**] CCITT Recommendation G.722(1984).

[**CCITT 87**] CCITT, Study Group XVIII - Report R 46(B) (1987).

[**CCITT 89**] CCITT, Study Group XVIII, June 1989 Geneva, WP-8 Report (1989).

[**Chen 88**] Chen, T. M. and Messerschmitt, D. G., "Integrated voice/data switching," IEEE Communications Magazine, **26**, 6, pp. 16- 26(1988).

[**Conway 67**] Conway, R. M., Maxwell, W. L. and Miller, L. W., "Theory of scheduling," Addison-Wesley Publishing Company, Reading (1967).

[**Coviello 75**] Coviello, G. J. and Vena, P. A., "Integration of circuit/packet switching by a SENET(Slotted Envelope Network) concept," Proc. National Telecommunications Conference, **42**, pp. 12-17(1975).

[**Coviello 79**] Coviello, G. J., "Comparative discussion of circuit versus packet switched voice," IEEE Trans. on Communications, **27**, pp. 1153-1159(1979).

[**Cox 66**] Cox, D. R. and Lewis, P. A. W., "The Statistical Analysis of Series of Events," Methuen, London(1966).

[**Cox 80**] Cox, R. V. and Crochiere, R. E., "Multiple user variable rate coding for TASI and packet transmission systems," IEEE Trans. on Communications, **28**, 3, pp. 334-344(1980).

[**Daigle 85**] Daigle, J. N. and Langford, J. D., "Queueing analysis of a packet voice communication system," INFOCOM85, pp. 18-26(1985).

[**Daigle 86**] Daigle, J. and Langford, J., "Models for analysis packet voice communications system," IEEE J. Selected Areas in Communicaions, **4**, 6, pp. 847-855(1986).

[**Daley 76**] Daley, D. J., "Queueing Output Processes," Adv. Appl. Prob., **8**, pp.395-415(1976).

[**Derman 70**] Derman, C., "Finite state markovian decision process," Academic Press, New York (1970).

[**Descloux 85**] Descloux, A., "Models for switching networks with integrated voice and data traffic," 11th International Teletraffic Congress, pp. 134-139(1985).

[**Devault 88**] Devault, M., Cochennec, J. and Servel, M., "The "Prelude" ATD experiment: assessments and future prospects," IEEE J. Selected Areas in Communications, **6**, 9, pp. 1528-1537(1988).

[**Disney 85**] Disney, R. L. and König, D., "Queueing Networks: A Survey of Their Random Processes," SIAM Review, **27**, 3, pp. 335-403(1985).

[**Disney 87**] Disney, R. L. and Kiessler, P. C., "Traffic Processes in Queueing Networks," The John Hopkins University Press(1987).

[**Dravide 89**] Dravide, S. and Sriram, K., "End-to-end performance models for variable bit rate voice over tandem links in packet networks," INFOCOM89, pp. 1089-1097 (1989).

[**Dubnowski 79**] Dubnowski, J. and Crochiere, R., "Variable rate coding of speech," Bell Syst. Tech. J., **58**, 3, pp. 577-600(1979).

[**Eckberg 88**] Eckberg, A. E. and Hou, T. -C., "Effects of output buffer sharing on buffer requirements in an ATDM packet switch," INFOCOM88, pp. 459-466(1988).

[**Eckberg 89**] Eckberg, A. E., Jr., Luan, D. T., and Lucantoni, D. M., "Meeting the challenge: congestion and flow control strategies for broadband information transport," IEEE Global Telecommunications Conference 1989, pp. 49.3.1-5 (1989).

[**Eklundh 88a**] Eklundh, B. and Gard, I., "Design options for ATM network," 1988 International Zurich Seminar on Digital Communication, pp. 199-205(1988).

[**Eklundh 88b**] Eklundh, B., Sällberg, K. and Stavenow, B., "Asynchronous transfer modes - options and characteristics," 12th International Teletraffic Congress, pp. 1.3A.3.1-7 (1988).

[**Filipiak 89**] Filipiak, J., "Structured systems analysis methodology for design of an ATM network architecture," IEEE J. Selected Areas in Communications, **7**, 8, pp. 1263-1273 (1989).

[**Fischer 76**] Fischer, M. J. and Harris, T. C., "A model for evaluating the performance of an integrated circuit- and packet-switched multiplex structure," IEEE Trans. on Communications, **24**, 2, pp. 195-202(1976).

[**Fischer 79**] Fischer, M. J., "Data performance in a system where data packets are transmitted during voice silent periods - single channel case," IEEE Trans. on Communicaitons, **27**, 9, pp. 1371- 1375(1979).

[**Fischer 80**] Fischer, M. J., "Delay analysis of TASI with random fluctuations in the number of voice calls," IEEE Trans. on Communications, **28**, 11, pp. 1883-1889(1980).

[**Forst 86**] Forst, V. S., Friedman, E. M. and Minden, G. J., "Multirate voice coding for load control on CSMA/CD local computer networks," Computer Networks and ISDN Systems, **11**, pp. 99-110 (1986).

[**Fredericks 86**] Fredericks, A. A., "Traffic models for analyzing certain overload strategies for integrated voice/data networks," ITC Specialist Seminar(1986).

[**Gafni 84**] Gafni, E. M., Bertsekas, D. P., "Dynamic control of session input rates in communication networks," IEEE Transaction on Automatic Control, **29**, 11, pp. 1009-1016 (1984).

[**Gallassi 89**] Gallassi, G., Rigolio, G. and Fratta, L., "ATM: bandwidth assignment and bandwidth enforcement policies," IEEE Global Telecommunications Conference 1989, pp. 49.6.1-6(1989).

[**Gallassi 90**] Gallassi, G., Rigolio, G. and Verri, L., "Resource management and dimensioning in ATM networks", IEEE Network Magazine, pp. 8-17, May (1990).

[**Gaver 82**] Gaver, D. P. and Lehoczkey, J. P., "Channels that cooperatively service a data stream and voice messages," IEEE Trans. on Communications, **30**, 5, pp. 1153-1162(1982).

[**Gelenbe 80**] Gelenbe, E. and Mitrani, I., "Analysis and synthesis of computer systems," Academic Press, London(1980).

[**Giocelli 87**] Giocelli, S., Demichelis, C., Giandonato, G. and Melen, R., "Experimenting with fast packet switching techniques in first generation ISDN environment," International Switching Symposium 1987, pp. B5.4.1-7(1987).

[**Goldberg 77**] Goldberg, H. M., "Analysis of the earliest due date scheduling rule in queueing systems," Mathematics of Operations Research, **2**, 2, pp. 145-154 (1977).

[**Gonet 86**] Gonet, P., "Fast packet approach to integrated broadband networks," Computer Communicaions, **9**, 6, pp. 292-298(1986).

[**Gonet 86**] Gonet, P., Servel M. and Coudreuse, J. P., "Asynchronous time-division switching: the way to flexible broadband communication networks," International Zurich Seminar on Digital Communications (1986).

[**Goodman 80**] Goodman, D. J., "Embedded DPCM for variable bit rate transmission," IEEE Trans. on Communications, **28**, 7, pp. 1040- 1046(1980).

[**Gopal 84**] Gopal, P. M., Wong, J. W. and Majithia, J. C., "Analysis of playout strategies for voice transmission using packet switching techniques," Performance Evaluation, **4**, pp. 11-18(1984).

[**Gopal 86**] Gopal, P. M. and Kadaba, B., "A simulation study of network delay for packetized voice," IEEE Global Telecommunications Conference 1986, **2**, pp. 932-938(1986).

[**Gravey 89**] Gravey, A., "A space priority queueing mechanism for multiplexing ATM channels", ITC Specialist Seminar, Adelaide, pp. 7.4.1-8 (1989).

[**Green 87**] Green, P. E., Jr. and Godard, D. N., "Prospects and design choices for integrated private networks," IBM System Journal, **26**, 1, pp. 37-54(1987).

[**Gruber 81a**] Gruber, J., "Performance considerations for integrated voice and data networks," Computer Communications, **4**, 3, pp. 106-126(1981).

[**Gruber 81b**] Gruber, J. G., "Delay related issues in integrated voice and data networks," IEEE Trans. on Communications, **29**, 6, pp. 786-800(1981).

[**Gruber 82**] Gruber, J. G., "A comparison of measured and calculated speech temporal parameters relevant to speech activity detection," IEEE Trans. on Communications, **30**, 4, pp. 728-738(1982).

[**Gruber 83**] Gruber, J. G. and Le, N. H., "Performance requirements for integrated voice/data networks," IEEE J. Selected Areas in Communications, **1**, 6, pp. 981-1005(1983).

[**Gruber 85**] Gruber, J. G. and Strawczynski, L., "Subjective effects of variable delay and speech clipping in dynamically managed voice systems," IEEE Trans. on Communications, **33**, 3, pp. 801-808(1985).

[**Händel 89**] Händel, R., "Evolution of ISDN towards broadband ISDN," IEEE Network, pp. 7-13 (1989).

[**Harrington 80**] Harrington, E. A., "Voice/data integration using circuit switched networks," IEEE Trans. on Communications, **28**, 6, pp. 781-793(1980).

[**Haselton 83**] Haselton, E. F., "A PCM frame switching concept leading to burst switching network architecture," IEEE Communications Magazine, **21**, 6, pp. 13-19(1983).

[**Haskell 72**] Haskell, B. G., "Buffer and channel sharing by several interframe picturephone coders," Bell Syst. Tech. J., **51**, 1, pp. 261-289(1972).

[**Hayashida 86**] Hayashida, Y., Fujii, S. and Tezuka, Y., "Traffic analysis of voice/data integrated system with controled arrival rate of data packet," Trans. IECE Japan, **J69B**, 9, pp. 888-897(1986) (in Japanese).

[**Heffes 80**] Heffes, H., "A Class of a Data Traffic Processes - Covariance Function Characterization and Related Queueing Results," Bell Syst. Tech. J., **59**, pp. 897-929(1980).

[**Heffes 86**] Heffes, H. and Lucantoni, D.M., "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," IEEE J. Selected Areas in Communications, **4**, pp. 856-868 (1986).

[**Heggestad 82**] Heggestad, H. M., McAulay, R. J. and Tierney, J., "Practical considerations for speech digitizing system at rates from 64.0 to 0.6 kbps," IEEE Global Telecommunications Conference 1982 (1982).

[**Hirano 89**] Hirano, M. and Watanabe, N., "Characteristtics of multiplexed burst traffic in ATM switching systems," Trans. IEICE, **J72-B-I**, pp. 264-271(1989)(in Japanese).

[**Holtzman 71**] Holtzman, J. M., "Bounds for a dynamic priority queue," Operations Research, **19**, 2, pp. 461-468 (1971).

[**Holtzman 85**] Holtzman, J. M., "The interaction between queueing and voice quality in variable bit rate packet voie systems," 11th International Teletraffic Congress, pp. 151-154(1985).

[**Honig 84**] Honig, M. L., "Analysis of TDMA network with voice and data traffic," Bell Laboratories Techinical Journal, **63**, 8, pp. 1537- 1563(1984).

[**Howard 71**] Howard, R. A., "Dynamic probabilistic systems," John Wiley & Sons, New York (1971).

[**Huang 84**] Huang, A. and Knauer, S., "Starlite: A wideband digital switch," IEEE Global Telecommunocations Conference 1984, pp. 5.3.1-5(1984).

[**Hui 88**] Hui, J. Y., "Resource allocation for broadband networks," IEEE J. Selected Areas in Communications, **6**, 9, pp. 1598-1608(1988).

[**Hui 90**] Hui, J. Y., "Switching and traffic theory for integrated broadband networks", Kluwer Academic, Boston (1990).

[**Ide 88**] Ide, I., "Superposition of interrupted Poisson process and its application to packetized voice mutiplexers," 12th International Teletraffic Congress, pp. 3.1B2.1-7(1988).

[**Irland 87**] Irland, M., "Buffer management in a packet switch," IEEE Trans. Communications, **26**, 3, pp.328-337 (1987).

[**Jack 86**] Jack, C. M. and Morse, J. D., "Simulation of a burst-switched integrated voice and bulk data communications network," Annual Simulation Symposium, 19th, pp. 145-157(1986).

[**Jackson 60**] Jackson, J. R., "Some problems in queueing with dynamic priorities," Naval Res. Log. Quart., **7**, 3, pp. 235-249 (1960).

[**Jackson 61**] Jackson, J. R., "Queues with dynamic priority discipline," Management Science, **8**, 1, pp. 18-34 (1961).

[**Jackson 62**] Jackson, J. R., "Waiting time distribution for queues with dynamic priorities," Naval Res. Log. Quart., **9**, 1, pp. 31-36 (1962).

[**Jaiswal 68**] Jaiswal, N. K., "Priority queues," Academic Press, New York (1968).

[**Janakiraman 84a**] Janakiraman, N., Pagurek, B. and Neilson, J. E., "Performance analysis of an integrated switch with fixed or vaiable frame rate and movable voice/data boundary," IEEE Trans. on Communications, **32**, 1, pp. 34-39(1984).

[**Janakiraman 84b**] Janakiraman, N., Pagurek, B. and Neilson, J. E., "Delay versus TASI advantage in a packet voice multiplexer," IEEE Trans. on Communications, **32**, 3, pp. 319-320(1984).

[**Jayant 81**] Jayant, N. S. and Christensen, S. W., "Effects of packet losses in waveform coded speech and improvements due to an odd-even sample interpolation procedure," IEEE Trans. on Communications, **29**, 2, pp. 101-109(1981).

[**Jenq 84**] Jenq, Y. C., "Approximations for packetized voice traffic in statistical multiplexer," INFOCOM84, pp. 256-259(1984).

[**Kawarazaki 88**] Kawarazaki, M., Okada, T. and Sasakawa, M., "Perspective of ATM communication technique - Evolution to broadband communication network -," J. of IEICE, **71**, 8, pp. 806-814(1988) (in Japanese).

[**Kawarazaki 90**] Kawarazaki, M., Saito, H. and Yamada, H., "Traffic analysis of an ATM transport network," IEEE International Conference on Communications 1990.

[**Kawashima 89**] Kawashima, K. and Saito, H., "Teletraffic issues in ATM networks," ITC Seminar, Adelaide, pp. 17.5.1-8 (1989).

[**Kim 83**] Kim, B. G., "Characterization of arrival statistics of multiplexer voice packets," IEEE J. Selected Areas in Communications, **1**, 6, pp. 1133-1139(1983).

[**Kim 88**] Kim, B. G., "Integrated voice/data multiplexers with slotted and nongated packet services," IEEE Trans. on Communications, **29**, 6, pp. 800-807(1981).

[**Kirton 87**] Kirton, P. Ellershaw, J. and Littlewood, M., "Fast packet switching for integrated network evolution," International Switching Symposium 1987, pp. B6.2.1-7(1987).

[**Kishimoto 89**] Kishimoto, R., Ogata, Y. and Inumaru, F., "Generation interval distribution characteristics of packetized variable rate video coding data streams in an ATM networks," IEEE J. Selected Areas in Communications, **7**, 5, pp. 833-841(1989).

[**Konheim 84**] Konheim, A. G. and Pickholtz, R. L., "Analysis of integrated voice/data multiplexing," IEEE Trans. on Communications, **32**, 2, pp. 140-147(1984).

[**Körner 83**] Körner, U., "Congestion control in packet switching computer communication network," 10th International Teletraffic Congress, pp. 3.3-5 (1983).

[**Kou 85**] Kou, K. Y., O'Neal, Jr. and Nelson, A. A., "Digital speech interpolation for variable rate coders with application to subband coding," IEEE Trans. on Communications, **33**, 10, pp. 1100-1108(1985).

[**Kowalk 88**] Kowalk, W. and Lehnert, R., "The "policing function" to control user access in ATM networks - definition and implementation - ," ISSLS88, pp. 12.2.1-6(1988).

[**Krämer 76**] Krämer, W. and Lagenbach-Belz, M., "Approximate Formulas for the Delay in the Queueing System GI/G/1," 8th International Teletraffic Congress, pp. 235.1-8 (1976).

[**Kröner 90**] Kröner, H., "Comparative performance study of space priority mechanisms for ATM networks", INFOCOM90, pp. 1136-1143 (1990).

[**Kuczura 73**] Kuczura, A., "The interrupted Poisson process as an overflow process," Bell Syst. Tech. J., **52**, pp. 473-448(1973).

[**Kümmerle 74**] Kümmerle, K., "Multiplexer performance for integrated line and packet switched traffic," International Conference of Computer Communication, pp. 507-515(1974).

[**Kurose 88**] Kurose, J. F., Schwartz, M. and Yemini, Y., "Controlling window protocols for time-constrained communication in multiple access networks," IEEE Trans. on Communications, **36**, 1, pp. 41-49(1988).

[**Kwong 84**] Kwong, R. H. and Leon-Garcia, A., "Performance analysis of an integrated hybrid-switched multiplexer structure," Performance Evaluation, **4**, pp. 81-91(1984).

[**Lazar 83a**] Lazar, A. A., "The throughput time delay function of an M/M/1 queue," IEEE Trans. on Information Theory, **29**, 6, pp. 914-918 (1983).

[**Lazar 83b**] Lazar, A. A., "Optimal flow control of a class of queueing networks in equilibrium," IEEE Trans. on Automatic Control, **28**, 11, pp. 1001-1007 (1983).

[**Lee 86**] Lee, H. and Un, C., "A study of an on-off characterisation of conversational speech," IEEE Trans. on Communications, **34**, 6, pp. 630-637(1986).

[**Lehoczky 81a**] Lehoczky, J. P. and Gaver, D. P., "Diffusion approximations for the cooperarive service of voice and data messages," Journal of Applied Probability, **18**, pp. 660-671(1981).

[**Lehoczky 81b**] Lehoczky, J. P. and Gaver, D. P., "Performance evaluation of voice/data queueing systems," Applied Probability - Computer Science, The Interface Volume II (R. L. Disney and T. J. Ott editted), pp. 329-346(1981).

[**Leon-Garcia 82**] Leon-Garcia, A., Kwong, R. H. and Williams, G. F., "Performance evaluation methods for an integrated voice/data link," IEEE Trans. on Communications, **30**, 8, pp.1848-1858(1982).

[**Leon-Garcia 86**] Leon-Garcia, A. and Aboul-Magd, O. S., "Performance analysis of integrated voice and data tandem link network," IEEE Global Telecommunications Conference 1986, 2, pp. 944-948(1986).

[**Li 85**] Li, S. and Mark, J. W., "Performance of voice/data integration on a TDM system," IEEE Trans. on Communications, **33**, 12, pp. 1265-1273(1985).

[**Li 87**] Li, S., "A new performance measurement for voice transmission in burst and packet switching," IEEE Trans. on Communications, **35**, 10, pp. 1083-1094(1987).

[**Li 88a**] Li, S. and Mark, J. W., "Performance trade-offs in an integrated voice/data services TDM System," Performance Evaluation, **8**, pp. 51-64(1988).

[**Li 88b**] Li, S. and Mark, J. W., "Simulation study of a network of voice/data integrated TDM's," IEEE Trans. on Communications, **36**, 1, pp. 126-132(1988).

[**Li 88c**] Li, S. and Mark, J. W., "Traffic characterization for integrated services," INFO-COM88, pp. 781-790(1988).

[**Li 88d**] Li, S., "Study of packet loss in a packet switched voice system," IEEE International Conference on Communications 1988, pp. 1519-1526(1988).

[**Liang 85**] Liang, T. Y., "A simpler approach to 'analysis of integrated voice/data multiplexing'," IEEE Trans. on Communications, **33**, 1, pp. 104-106(1985).

[**Lim 86**] Lim, Y., "Performance analysis of an integrated voice and data TDM link with digital speech interpolation," IEEE Global Telecommunications Conference 1986, **2**, pp. 960-964(1986).

[**Lim 88**] Lim, Y. and Kobza, J., "Analysis of a delay-dependent priority discipline in a multi-class traffic packet switching node," INFOCOM88, pp. 9A.4.1-10 (1988).

[**Listanti 83**] Listanti, M. and Villani, F., "Voice communication handling in X.25 packet switching networks," IEEE Global Telecommunications Conference 1983, pp. 2.4.1-5(1983).

[**Lucantoni 88**] 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," Working paper 88-008, Systems and Industrial Engineering Department, The University of Arizona (1988).

[**Luderer 87**] Luderer, G., Mansell, J., Messerli, E., Staehler, R. and Vaidya, A., "Wide-band packet technology for switching system," International Switching Symposium 1987, pp. B6.1.1-7(1987).

[**Luhanga 88**] Luhanga, M. L., "A fluid approximation model of an integrated packet voice and data multiplexer," INFOCOM88, pp. 687-692(1988).

[**Lyghounis 74**] Lyghounis, E., Poretti, I. and Monti, G., "Speech interpolation in digital transmission systems," IEEE Trans. on Communications, **22**, 9, pp. 1179-1189(1974).

[**Ma 86**] Ma, D. -J., Makowski, A. M. and Shwartz, A., "Estimation and optimal control for constrained Markov chains," 25th Conference on Decision and Control, pp. 994-999 (1986).

[**Ma 87**] Ma, B. N. W. and Mark, J. W., "Performance analysis of burst- switching for integrated voice/data sevices: a voice delay scheme," INFOCOM87, pp. 864-871(1987).

[**Ma 88**] Ma, B. N. W. and Mark, J. W., "Performance analysis of burst switching for integrated voice/data services," IEEE Trans. on Communications, **36**, 3, pp. 282-297(1988).

[**Machihara 88**] Machihara, F., "Completion time of service unit interrupted by PH-Markov renewal custormers and their applications," 12th International Teletraffic Congress, pp. 5.4B5.1-7 (1988).

[**Maglaris 82**] Maglaris, B. S. and Schwartz, M., "Optimal fixed frame mutiplexing in integrated line- and packet-switched communication networks," IEEE Trans. on Information Theory, **28**, 2, pp. 263-273(1982).

[**Maglaris 88**] Maglaris, B., Anastassiou, D., Sen, P., Karlsson, G. and Robbins, J., "Performance models of statistical multiplexing in packet video communications," IEEE Trans. on Communications, **36**, 7, pp. 834-844(1988).

[**Mase 88**] Mase, K. and Kawano, H., "A study on dimensioning and control for ATM networks," IEICE Technical Report, IN88-118 (1988) (in Japanese).

[**McDysan 88**] McDysan, D. E. and Pickholtz, R. L., "Performance analysis of a moving boundary/fixed frame integrated voice/data multiplexing technique with correlatd sources," INFOCOM88, pp. 661-668(1988).

[**Mehmet 86**] Mehmet-Ali, M. K. and Woodside, C. M., "Analysis of re- assembly buffer requirements in a packet voice network," INFOCOM86, pp. 233-238(1986).

[**Minoli 79**] Minoli, D., "Optimal packet length for packet voice communication," IEEE. Trans. on Communications, **27**, 3, pp. 607-611(1979).

[**Mitrani 86**] Mitrani, I., Adams, J. and Falconer, R., "A modelling study of the Orwell ring protocol," Teletraffic Analysis and Comuter Performance Evaluation, North Holland(1986).

[**Miyahara 78**] Miyahara, H. and Hasegawa, T., "Integrated switching with variable frame and packet," IEEE International Conference on Communications 1978, pp. 20.3.1-5(1978).

[**Miyake 89**] Miyake, K. and Kawashima, K., "Traffic control and design issues in ATM network," 1989 Natl. Conf. Rec. Spring, IEICE, SB-5-1 (1989) (in Japanese).

[**Montogomery 83**] Montogomery, W. A., "Techniques for packet voice synchronization," IEEE J. Selected Areas in Communiacations, **1**, 6, pp. 1022-1028(1983).

[**Morris 81**] Morris, R. J. T., "An algorithmic technique for a class of queueing models with packet swtching applications," IEEE International Conference on Communications 1981, pp. 41.2.1-8 (1981).

[**Morse 85**] Morse, J. D., "Performane evaluation of burst-switched integrated voice-data networks," 11th International Teletraffic Congress, pp. 155-160(1985).

[**Muise 86**] Muise, R. W., Schonfeld, T. J. and Zimmerman, G. H., III., "Experiments in wideband packet technology," International Zurich Seminor on Digital Communications 1986, pp. 135-139(1986).

[**Murase 89**] Murase, T., Suzuki, H., Miyao, Y., Sato, S. and Takeuchi, T., "A traffic control for ATM networks," IEICE Technical Report, SSE89-69(1989) (in Japanese).

[**Murata 89**] Murata, M., Oie, Y., Suda, T. and Miyahara, H., "Analysis of a discrete time single-server queve with bursty inputs for traffic control in ATM networks," IEEE Global Telecommunications Conference 1989, pp. 49.5.1-7 (1989).

[**Murata 90**] Murata, M., Oie, Y., Suda, T. and Miyahara, H., "Analysis of a discretetime single-server queue with bursty inputs for traffic control in ATM networks," IEEE J. Selected Areas in Communications, to be published.

[**Nain 86a**] Nain, P. and Ross, K. W., "Optimal multiplexing of heterogeneous traffic with hard constraint," Performance 86 and ACM Sigmetrics 1986, Raleigh (1986).

[**Nain 86b**] Nain, P. and Ross, K. W., "Optimal priority assignment with hard constraint," IEEE Trans. on Automatic Control, **31**, 10, pp. 883-888 (1986).

[**Netterman 79**] Netterman, A. and Adiri, I., "A dynamic priority queue with general concave priority functions," Operations Research , **27**, 6, pp. 1088-1100 (1979).

[**Neuts 79**] Neuts, M. F., "A versatile Markovian Point Process," J. Appl. Prob., **16**, pp. 764-779(1979).

[**Neuts 81**] Neuts, M. F., "Matrix geometric solutions in stochastic models," Baltimore, The John Hopkins University Press (1981).

[**Noguchi 90**] Noguchi, K. and Okada, T., "Admission control in an ATM network," 1990 Natl. Conf. Rec. Spring, IEICE (1990) (in Japanese).

[**Nomura 87**] Nomura, M., Fujii, T. and Ohta, N., "A study on bursty signal modeling in a low bit rate video coding," IEICE Technical Report, IN87-39 (1987) (in Japanese).

[**Nomura 89a**] Nomura, M., Suzuki, J. and Ohta N., "A stydy on burstiness in variable rate coding of broadcast video," The 1st workshop on Information Signal Transmission and Signal Processing, pp. 135-140(1989).

[**Nomura 89b**] Nomura, M., Fujii, T. and Ohta, N., "Basic characteristics of variable rate video coding in ATM environment," IEEE Journal on Selected Areas in Communications, **7**, 5, pp. 752-760(1989).

[**Ogino 88**] Ogino, N., "Characteristics in ATM multiplexing of video signals," IEICE Technical Report, SSE 88-199(1988).

[**Ohnishi 88**] Ohnishi, H., Okada, T. and Noguchi, K., "Flow control schemes and delay / loss tradeoff in ATM networks," IEEE J. Selected Areas in Communications, **6**, 9, pp. 1609-1616 (1986).

[**Okada 86**] Okada, H., "Delay behavior of data traffic in an integrated voice/data multiplexer structure: Multi-capacity-limits(MCL)," IEEE Trans. on Communications, **34**, 3, pp. 300-303(1986).

[**O'Reilly 85**] O'Reilly, P., "A fluid-flow approach to performance analysis of integrated voice-data system with speech interpolation," Modelling Techniques and Tools for Performance Analysis '85, pp. 115-130(1985).

[**O'Reilly 86a**] O'Reilly, P., "Burst and fast-packet switching: performance comparisons," INFOCOM86, pp. 653-666(1986).

[**O'Reilly 86b**] O'Reilly, P., "Performance analysis of data in burst switching," IEEE Trans. on Communications, **34**, 12, pp. 1259-1263(1986).

[**O'Reilly 87a**] O'Reilly, P. and Ghani, S., "Data performance in burst switching when the voice silence periods have a hyperexponential distribution," IEEE Trans. on Communications, **35**, **10**, pp. 1109-1112(1987).

[O'Reilly 87b] O'Reilly, P., "Burst and fast packet switching: Performance comparisons," Computer Networks and ISDN systems, **13**, 1, pp. 21-32(1987).

[Panwar 89] Panwar, S. S., Towsley, D. and Wolf, J. K., "Optimal scheduling policies for a class of queues with customer deadlines to the beginning of service," Journal of the Association for Computing Machinery, **35**, 4, pp. 832-844(1989).

[Rabiner 78] Rabiner, L. R. and Scafer, R. W., "Digital processing of speech signals," Prentice-Hall, New Jersey (1978).

[Ramaswami 80] Ramaswami, V., "The N/G/1 queue and its detailed analysis," Adv. Appl. Prob., **12**, pp. 222-261(1989).

[Rao 73] Rao, C. R., "Linear statistical inference and its application," John Wiley & Sons, New York (1973).

[Roberts 87] Roberts, J. W., "Some performance issues arising in a fast packet switching enviroment," The second Australian Teletraffic Seminar, Adelaide, December (1987).

[Ross 82] Ross, M. J. and Mowafi, O. A., "Performance analysis of hybrid switching concepts for integrated voice/data communications," IEEE Trans. on Communications, **30**, 5, pp. 1073-1087(1982).

[Ross 88] Ross, K. W. and Chen, B., "Optimal scheduling of interactive and noninteractive traffic in telecommunication syetems," IEEE Trans. on Automatic Control, **33**, 3, pp. 261-267(1988).

[Rossiter 87] Rossiter, M. H., "A Switched Poisson Model for Data Traffic," A.T.R., **21**, 1, pp. 53-57(1987).

[Saito 89a] Saito, H., "Optimal control of variable rate coding in integrated voice/data packet networks," Performance Evaluation, 10, 2, pp. 115-128 (1989).

[Saito 89b] Saito, H., "The departure process of an N/G/1 queue," Performance Evaluation, 11, pp. 241-251(1990).

[Saito 89c] Saito, H., "A simplified dimensioning method of ATM networks," IEICE Technical Report, SSE89-112(1989).

[Saito 90a] Saito, H., "Optimal queueing discipline for real-time traffic at ATM switching nodes," IEEE Trans. Communications, to be published.

[Saito 90b] Saito, H., "Optimal control of variable rate coding with incomplete observation in integrated voice/date packet networks," European J. of Operational Research, to be published.

[Saito 90c] Saito, H. and Kawashima, K., "Performance modelling of integrated voice and data communications networks," Stochastic Analysis of Computer and Communication Networks, North Holland pp. 807-849 (1990).

[Saito 90d] Saito, H., "Call admission control in an ATM network using upper bound of cell loss probability," IEEE Trans. Communications, to be published.

[Saito 90e] Saito, H., "New dimensioning concept for ATM networks," 7th Intenational Teletraffic Congress, Specialist Seminar (1990).

[Saito 91a] Saito, H., Kawarasaki, M., and Yamada, H., "An analysis of statistical multiplexing in an ATM transport network," IEEE J. Selected Areas in Communications, to be published.

[Saito 91b] Saito, H., "Queueing analysis of cell loss probability control in ATM networks", 13th International Teletraffic Congress (1991).

[Sato 88] Sato, K., Nakada, H. and Sato, Y., "Variable rate speech coding and network delay analysis for universal transport network," INFOCOM88, pp. 771-780(1988).

[Sato 89] Sato, K., Ohta, S. and Tokizawa, I., "Broodband transport network architecture based on virtual path concept," Trans. IEICE, **J72-B-I**, pp. 904-916 (1989)(in Japanese).

[Schoute 88] Schoute, F. C., "Simple decision rules for acceptance of mixed traffic streams," Philips TDS Review, **46**, 2, pp. 35-48(1988).

[Segall 77] Segall, A., "Optimal control of noisy finite- state Markov processes," IEEE Trans. on Automatic Control, **22**, 2, pp. 179-186(1977).

[Seguel 82] Seguel, J., Tanaka, Y. and Akiyama, M., "Simulation analysis of the waiting time distribution of a packetized voice concentrator," Trans. IECE Japan, **65**, 2, pp. 115-122(1982).

[Sen 89] Sen, P., Maglaris, B., Rikli, N. and Anastssiou, D., "Models for packet switching of variable-bit-rate video sources," IEEE J. Selected Areas in Communications, **7**, 5, pp. 865-869(1989).

[Shioyama 85] Shioyama, T., Ohno, K. and Mine, H., "An optimal control of an integrated circuit-and packet-switching system," 11th International Teletraffic Congress, pp. 147-150(1985).

[Sriram 83] Sriram, K., Varshney, P. K. and Shanthikumar, J. G., "Discrete-time analysis of integrated voice/data multiplexers with and without speech activity detectors," IEEE J. Selected Areas in Communications, **1**, 6, pp. 1124-1132(1983).

[**Sriram 86**] Sriram, K. and Whitt, W., "Characterizating superposition arrival processes in packet multiplexers for voice and data," IEEE J. Selected Areas in Communications, 4, 6, pp. 833-846(1986).

[**Sriram 88**] Sriram, K. and Lucantoni, D. M., "Traffic smoothing effects of bit dropping in a packet voice multiplexer," INFOCOM88, pp. 759-770(1988).

[**Stanford 85**] Stanford, D. A., "Modeling priority queueing characteristics in approximate analytical tools for open queueing networks," in Modelling Techniques and Tools for Performance Analysis 85, pp. 131-146(1985).

[**Stern 83**] Stern, T. E., "A queueing analysis of packet voice," IEEE Global Telecommunications Conference 1983, pp. 2.5.1-6(1983).

[**Stern 84**] Stern, T. E., "Analytical and computational techniques for packet voice system performance evaluation," Proc. Conf. Inf. Sci. Syst. 1984, pp. 415-419(1984).

[**Suda 84**] Suda, T., Miyahara, H. and Hasegawa, T., "Performance evaluation of a packetized voice system-simulation study," IEEE Trans. on Communications, 32, 1, pp. 97-102(1984).

[**Sumita 88**] Sumita, S. and Ozawa, T., "Achivability of Performance objectives in ATM switching nodes," Proceeding of the International Seminar on Performance of Distributed and parallel Systems Kyoto, pp. 45-56 (1988).

[**Tanaka 82**] Tanaka, Y., Seguel, J. and Akiyama, M., "Performance analysis of a voice packet multiplexer," Trans. IECE Japan, J65B, 5, pp. 539-546(1982) (in Japanese).

[**Tham 83**] Tham, Y. K. and Hume, J. N. P., "Analysis of voice and low priority data traffic by means of brisk periods and slack periods," Computer Communications, 6, 1, pp. 14-22 (1983).

[**Thomas 84**] Thomas, A., Coudreuse, J. P. and Servel, M., "Asynchronous time-division techniques: an experimental packet network integrating videocommuncation," International Switching Symposium 1984, pp. 32C.2.1-7(1984).

[**Tijms 86**] Tijms, H. C., "Stochastic modelling and analysis: A computational approach," John Wiley & Sons, Chichaster (1986).

[**Tominaga 89**] Tominaga, H., "New directions in broadband networks," Trans. IEICE, J72-B-I, 11, pp. 876-885(1989).

[**Tucker 88**] Tucker, R. C. F., "Accurate method for analysis of a packet speech multiplexer with limited delay," IEEE Trans. on Communications, 36, 4, pp. 479-483(1988).

[**Turner 83**] Turner, J. S. and Wyatt, L. F., "A packet network architecture for integrated services," IEEE Global Telecommunications Conference 1983, pp. 45-50(1983).

[**Turner 85**] Turner, J. S., "Design of an integrated services packet network," Data Communications Symposium, 9th, pp. 124-133(1985).

[**Turner 86a**] Turner, J. S., "Design of an integrated services packet network," IEEE J. Selected Areas in Communications, 4, 8, pp. 1373-1380(1986).

[**Turner 86b**] Turner, J. S., "New directions in communications (or which way to the information age?)," IEEE Communications Magazine, 24, 10, pp.8-15(1986).

[**Turner 88**] Turner, J. S., "The challenge of multipoint communications," Traffic Engineering for ISDN Design and Planning (M. Bonatti and M. Decina, Editors), North-Holland, pp. 263-279(1988).

[**Ueda 83**] Ueda, H., Tokizawa, I. and Aoyama, T., "Evaluation of an experimental packetized speech and data transmission system," IEEE J. Selected Areas in Communiacations, 1, 6, pp. 1039-1045(1983).

[**Ushijima 72**] Ushijima, T., "A queueing system with Markov arrivals," J. Operations Research Society Japan, 15, pp. 167-193(1972).

[**Vakil 87**] Vakil, F. and Lazar, A. A., "Flow control protocols for integrated networks with partially observed voice traffic," IEEE Trans. on Automatic Control, 32, 1, pp. 2-14(1987).

[**Verbiest 88**] Verbiest, W., Pinnoo, L. and Voeten, B., "Statistical multiplexing of variable bit rate video sources in asynchronous transfer mode networks," IEEE Global Telecommunications Conference 1988, pp. 7.2.1-6(1988).

[**Verbiest 89**] Verbiest, W. and Pinnoo, L., "A variable bit rate bideo codec for asychronous transfer mode networks," IEEE J. Selected Areas in Communications, 7, 5, pp. 761-770(1989).

[**Webber 77**] Webber, S. A., Harris, C. J. and Flanagan, J. L., "Use of variable-quality coding and time-interval modification in packet transmission of speech," Bell Syst. Tech. J., 56, 8, pp. 1569-1573(1977).

[**Weinstein 79**] Weinstein, C. J. and Hofstetter, E. M., "The tradeoff between delay and TASI advantage in a packetized speech multiplexer," IEEE Trans. on Communications, 27, 11, pp. 1716-1720(1979).

[**Weinstein 80**] Weinstein, C. J., Malpass, M. L. and Fisher, M. J., "Data traffic performance of an integrated circuit-and packet-switched multiplexer structure," IEEE Trans. on Communications, 28, 6, pp. 873-878(1980).

[**Weinstein 83**] Weinstein, C. J. and Forgie, J. W., "Experience with speech communication in packet networks," IEEE J. Selected Areas in Communications, **1**, 6, pp. 963-980(1983).

[**Whitt 83**] Whitt, W., "The queueing network analyzer," Bell Syst. Tech. J., Part 1, **62**, 9, pp. 2779-2815(1983).

[**Williams 84**] Williams, G. F. and Leon-Garcia, A., "Performance analysis of integrated voice and data hybrid-switched links," IEEE Trans. on Communications, **32**, 6, pp. 695-706(1984).

[**Woodruff 88**] Woodruff, G. M., Rogers, R. G. H., and Richards, P. S., "A congestion control framework for high-speed integrated packetized transport," IEEE Global Telecommunications Conference 1988, pp. 7.1.1-5(1988).

[**Yamada 89a**] Yamada, H., Miyake, K., Kishino, F. and Manabe, K., "Modeling of arrival process of packetized video and related statistical multiplexer performance," 1989 Natl. Conf. Rec. Autumun, IEICE, B-6 (1989) (in Japanese).

[**Yamada 90**] Yamada, H., Komine, T., and Sumita, S., "Characteristics of statistical multiplexing with heterogeneous inputs," IEICE Technical Report, SSE90-49 (1990).

[**Yatsuzaka 82a**] Yatsuzaka, Y., "Highly sensitive speech detector and high-speed voice-band data discriminator in DSI-ADPCM systems," IEEE Trans. on Communications, **30**, 4, pp. 739-750(1982).

[**Yatsuzuka 82b**] Yatsuzuka, Y., "High-gain digital speech interpolation with adaptive differential PCM coding," IEEE Trans. on Communications, **30**, 4, pp. 750-761(1982).

[**Yin 87a**] Yin, N., Stern, T. E. and Li, S., "Performance analysis of a priority-oriented packet voice system," INFOCOM87, pp. 856-863(1987).

[**Yin 87b**] Yin, N., Li, S. and Stern, T. E., "Congestion control for packet voice by selective packet discarding," IEEE Global Telecommunications Conference 1987, pp. 1782-1786(1987).

[**Yum 87**] Yum, T. and Schwartz, M., "Packet-switched performance with different circuit-switched routing procedures in nonhierarchical integrated circuit-switched and packet-switched networks," IEEE Trans. on Communications, **35**, 3, pp. 362-366(1987).