PAPER

Connection Admission Control in ATM Networks

Hiroshi.Esaki[†] **Kazuaki. Iwamura**[††] **Toshikazu Kodama** [†] *and* **Takeo Fukuda**[†††] *Members*

SUMMARY The connection admission control is one of preventive traffic control in ATM networks. The one objective of connection admission control is to keep the network load moderate so as to achieve a performance objective associated with quality of service (QOS). Because the cell loss rate is more sensitive to offered load than the average queuing delay in ATM networks, QOS requirement associated with cell loss rate is considered. The connection admission control acts as one of the major roles in traffic control. The job of connection admission control is to make an acceptance decision for connection set-up request to control the network load. This paper proposed and evaluated a connection admission control method. The proposed method is suitable for real time operation even in large diversity of connection types, because the amount of calculation for connection admission control is reduced remarkably compared to conventional algorithms. Moreover, the amount of calculation for the algorithm does not increase even when the number of connection types increases. The proposed method uses probability function for the number of cells transferred from multiplexed connections and uses recursive equations in estimating cell loss rate.

key words: ATM, CAC, connection admission control, priority control, cell loss rate, CLR

1. Introduction

An ATM network is one method for providing a flexible and cost-effective network for handling a wide variety of communications. Traffic control is one of the major problems in ATM networks [6[8]. The one objective of traffic control is to make the network achieve performance objective associated with user requirements for communication quality, as usually specified by cell loss rate and transmission delay. Transmission delay is mainly caused by propagation delay which is determined by the distance between nodes and by queuing delay. Since the upper bound of queuing delay is determined by buffer length, it is possible to estimate worst case transmission delay. On the contrary, it is generally difficult to estimate the cell loss rate.

With ATM, cell loss rate is more sensitive to load than average queuing delay, as shown in Sect.2. Usually, the cell loss rate grows large at low offered load, even though the average queuing delay is small [13]. Since the ATM traffic is more sensitive to cell loss rate than queuing delay, the traffic control in ATM networks should be cell loss rate sensitive [13]. And, therefore, the cell loss rate requirement associated with QOS requirement is considered in this paper.

The job of bandwidth management which includes connection admission control is to maintain the network load at adequate level. One objective of bandwidth management is to achieve QOS requirements which is usually specified by transmission delay and cell loss rate. The job of connection admission control is to determine whether to accept a connection set-up request or not. In connection admission control, the cell loss rate for each connection is evaluated. The connection acceptance decision must be made within the required allowable connection set-up latency, in accordance with the received connection set-up request. This means that connection admission control must operate in real-time, even when various kinds of traffic are multiplexed. many methods are proposed by many researchers [5]-[11]. But, still, these methods has certain problems. Many methods can not be applied to the case where priority control is fulfilled. And, especially when various kinds of traffic are multiplexed, many methods can not operate in real time, because they require the huge calculations in the algorithms [10], [11]. The authors proposed in [12] a connection admission control method that does not requires the huge amount of calculations. This algorithm is discussed in Sect.3.2.

ATM network does not always provide only one QOS requirement class. Each connection requests one of QOS classes which are provided by the network in a connection set-up procedure. This means that ATM networks try to achieve multiple QOS requirement simultaneously connection by connection. Priority control is one method to provide multiple QOS classes [13]. Here, priority control means that the cell from priority connections are served first at queuing point. The cell from non-priority connections are served only when there is no priority cell at queuing point. The authors proposed a connection admission control method that can be applied to the case where the priority control is fulfilled [12]. However, the proposed method can be

Manuscript received February 18, 1993.
Manuscript revised May 24, 1993.

[†]The authors are with R&D Center, TOSHIBA CORPORATION, Kawasaki-shi, 210 Japan.
[††]The author is with the Kansai Research Laboratory, TOSHIBA CORPORATION, Kobe-shi, 685 Japan.
[†††]The author is with the Communication Systems & Technology Laborarty, TOSHIBA CORPORATION, Hino-shi, 191 Japan.

applied to the case where high speed connection has a higher priority than the low speed connection. Therefore, the proposed method can not apply to the case where multiple QOS classes associated with cell loss rate requirement must be provided. Here, high speed connection is defined as the connection whose peak cell transmission speed is large and the low speed connection is defined as the connection whose peak transmission speed is small [12],[13]. As shown in Sect.3, the proposed connection admission control method in this paper can be applied to the case where the priority control is performed to provide multiple QOS classes associated with cell loss rate requirement.

The previous works, e.g., [10], discussed about CAC algorithm, but they did not discuss about the applying method for the actual switching node. Every proposed algorithm are basically for an ATM multiplexer. The proposed method in this paper can be applied not only to the ATM multiplexer, but also to the output buffer type switching node. Also, the performance of the proposed algorithm for the output buffer type switching system is evaluated in this paper.

The structure of the rest of this paper is as follows. In Sect.2, the traffic characteristics in ATM networks is evaluated by computer simulation and the cell loss rate sensitivity of ATM traffic is shown. In Sect.3 and 4, a connection admission control method is proposed and evaluated.

2. Characteristics of ATM Traffic

Cell loss rate and average queuing delay are the two of main important performance objectives provided by the network to each connection. Through the following evaluations for the characteristics of ATM traffic (characteristics of multiplexed traffic in ATM fashion), it is shown that ATM traffic is more sensitive to Cell Loss Rate than to average queuing delay.

The evaluation of ATM traffic associated with average queuing delay and cell loss rate is performed by computer simulation, when two kinds of bursty connections (VBR connections) are multiplexed in ATM fashion. Here, bursty connection is defined as the connection that generates the cells continuously at the speed of its peak cell transmission rate [12].

The evaluation is performed through the output buffer type multiplexer. V is defined as the cell transmission speed of the multiplexed line. V_p is defined as the peak cell transmission speed of each connection, and V_a is defined as the average cell transmission speed of each connection. The simulation model is as follows. Bursty connection which generates cells with a so called two-state Markov chain model are multiplexed. Each connection has active state (active period) and idle state (idle period). During bursty periods, each connection

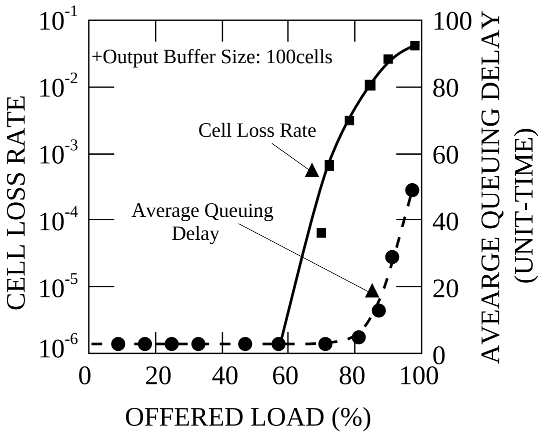

Fig.1 Cell loss rate and average queuing delay vs. offered load

transfers cells at the speed of its own peak cell transmission speed V_p. During idle period, each connection transfers no cell. The distribution of bursty periods (burst length) and idle periods (idle length) are geometrical.

Figure 1 shows the cell loss rate and the average queuing delay versus offered load. Here, the average queuing delay is shown by the normalized value. One unit-time corresponds to the transmission time of one cell, e.g., with STS-3 link and 53 bytes/cell, one unit-time corresponds to about $2.8\mu sec$. The type A connection is $V_p = V/10$ and $V_a = V_p/10$. The type B connection is $V_p = V/100$ and $V_a = V_p/2$. The average burst length of each connection types is 100 cells and the output buffer size is 100 cells. Here, the percentage of type A connection within the multiplexed connections is 10%.

As shown in this figure, cell loss rate is more sensitive to load than average queuing delay is. This means that even though the average queuing delay is small at a certain load, cell loss rate is already large. For example, at 70% offered load, even though the average queuing delay is less than five unit-times, the cell loss rate is already about 10^{-4}.

When we focus on a switching system at a certain node, the maximum allowable offered load is determined by cell loss rate rather than by queuing delay. This is because, when the actual or estimated cell loss rate reaches the upper limit of the objective cell loss rate associated with the required QOS, the average queuing delay is still small. This indicates that, in ATM networks, the cell loss rate is the dominant parameter to determine the maximum allowable offered load to the switching node.

Figure 2 shows the average queuing delay versus the percentage of type A connections, when cell loss rate requirement is either 10^{-3} or 10^{-5}. The simulation

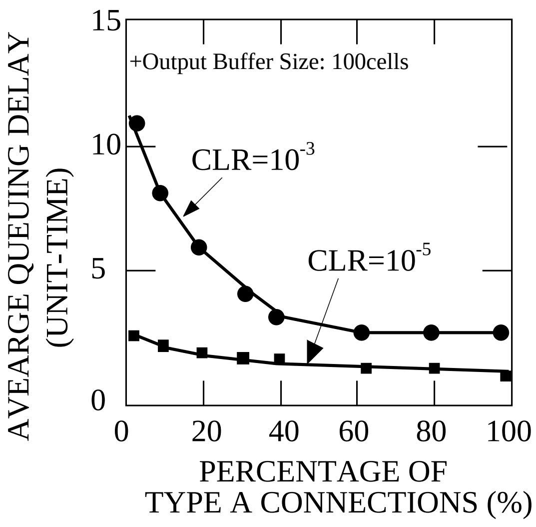

Fig.2 Average queuing delay vs. percentage of type A connections

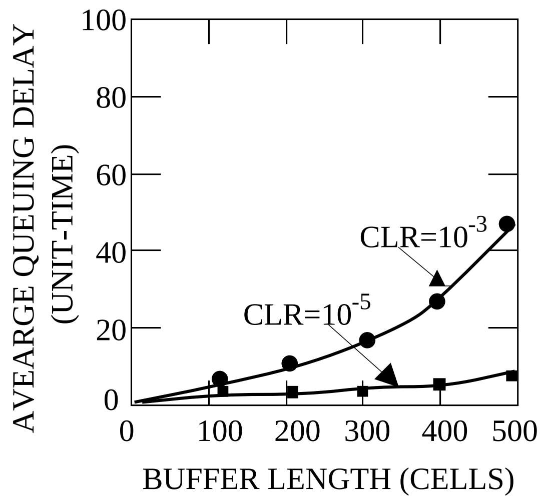

Fig.3 Average queuing delay vs. output buffer size

model is the same as Fig.1. This figure shows, again, that cell loss rate is more sensitive to load than average queuing delay is. This is because the average queuing delay is small, even when the required cell loss rate is 10^{-3}. For example, at 20% of the percentage of type A connections, both when CLR is 10^{-3} and CLR is 10^{-5}, the average queuing delay is less than five unit-times.

When the buffer size is small, say about one hundred cells, and when the burst length is large, say about one hundred cells of average burst length, the ATM traffic is more sensitive to cell loss rate than queuing delay. This may not be true for larger buffer size or for smaller burst length of multiplexed connection. Therefore, the traffic characteristics parameterizing the buffer length and the burst length of multiplexed connection is evaluated below.

First, the traffic characteristics versus the output buffer size is evaluated. From the view point of current VLSI technologies, three or four hundred cells per port is the upper bound to build in the actual switching system [14]. Figure 3 shows the average queuing delay verus the output buffer size, when the required cell loss rate are either 10^{-3} or 10^{-5}. The simulation model is the same as the one in Fig.1, except the output buffer size. As shown average queuing delay is small, even when the output buffer size is 500 cells. When the required cell loss rate is 10^{-5} with 500 cells output buffer, the average queuing delay is less than ten unit-times. Here, ten unit-time corresponds to about $28\mu sec$, with STS-3 link and 53 bytes; $28\mu sec$ corresponds to the propagation delay of transmission link for approximately 7.4km.

Second, the characteristics of average queuing delay parameterizing the burst length of the multiplexed connections is discussed. The amount of the total information field (payload) for one hundred cell burst, which is used in the above evaluation models, corresponds to 4.8Kbytes. As for the voice communication, the average length of talk-spurt is said about 200msec - 1.6 kbytes with 64kbps. For the data communication, such as Ethernet, the maximum data length is about 1.5 Kbytes for frame format. As for the video communication, one frame is 1/30 (=0.333) second - 1.875 Kbytes with 45Mbps (next target of MPEG2) or 6.25Kbytes with 1.5Mbps (MPEG1). Thus, the average burst length of one hundred cell might be too large for the voice communications or conventional data communications.

Figure 4 shows the maximum offered load and the average queuing delay versus the average burst length (M_a) of type A connection, when the required cell loss rate is smaller than 10^{-5}. The simulation model is the same as the one in Fig.1, except the average length of burst and idle periods for type A connections. The type A connection is $V_p = V/10, V_a = V_p/10$, and $M_i = M_a \times 9$. The type B connection is $V_p = V/100, V_a = V_p/2$, and $M_i = M_a = 100$ cells. Here, M_i is the average length of idle periods. As shown, the average queuing delay is quite small value at the load where the cell loss rate is 10^{-5}. For example, even when the average burst length of type A connection is ten cells, the average queuing delay is less than five unit-time. Five unit-time corresponds to about only 14 μsen with STS-3 link.

As well as the requirement of cell loss rate associated with QOS, of course, the requirement of transmission delay is also important parameter. However, since queuing delay is so small compared to propagation delay, the transmission delay is mainly composed

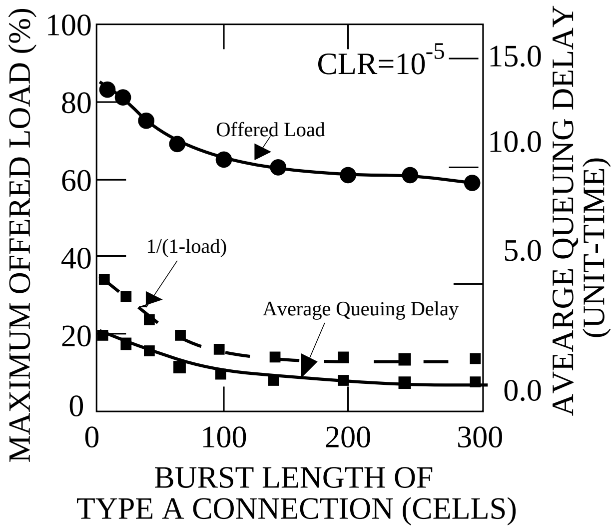

Fig.4 Maximum offered load and average queuing delay vs. burst length of type A connections

by propagation delay. Propagation delay is static value for a given path corresponding source destination node pair. This indicated that the connection acceptance decision associated with transmission delay quality is enough to be performed based on propagation delay for the source destination node pair.

As shown, the maximum allowable offered load in a certain link is determined by the cell loss rate rather than by the average queuing delay. In other words, the connection acceptance decision whether to accept a connection or not will be performed based on cell loss rate [12],[13]. In the following sections, the connection admission control algorithm is discussed.

3. Connection Admission Control Method

3.1 Requirement

Connection admission control determines whether to accept a call or not when a connection set-up request occurs so as to achieve a QOS performance objective. The proposed method focuses on satisfying user requirement for cell loss rate associated with QOS performance objective.

The requirements which a connection admission control must provide are as follows.

1. Real time response: determine the connection acceptance decision within a defined latency requirement for connection set-up request.
2. Safety margin: evaluate a cell loss rate which is never smaller than actual value
3. Usage Parameter Control: monitor whether each connection keeps negotiated traffic parameters
4. Connection diversity: operate even when many connection types are multiplexed
5. Priority control case: operate when priority control is performed
6. High utilization efficiency: use the multiplexed line at high efficiency by taking into account of statistical multiplexing effect.

Many methods are proposed by many researchers [5]-[11]. But, still, these methods have certain problems. Many methods can not be applied to the case where a priority control is fulfilled. Also, many methods can not operate in real time, when various types of connections are multiplexed. This is because that the required amount of calculation for the algorithms discussed in [10],[11] increases exponentially, in accordance with increase of number of connection types, so that huge amount of calculation is required when many types of connections are multiplexed, even though they can provide the requirement of safety margin.

The proposed method in this paper provides these requirements by using the following techniques.

1. Recursive Equations: reducing amount of calculation
2. Probability Density Function: applicability of connection diversity and of priority control case
3. Bufferless Model: evaluating never smaller cell loss rate than real value
4. Simple Traffic Parameters (peak and average): feasibility of usage parameter control easily.

The real-time operation requirement can be provided through reducing the amount of calculation in evaluating cell loss rate. This is because that a large part of calculation in connection admission control concerns evaluating cell loss rate. Using recursive equations and using probability functions for the number of cells which are transferred from multiplexed connections, the amount of calculation in connection admission control algorithm can be reduced dramatically. The safety requirement can be provided by assuming bufferless model in evaluating cell loss loss rate. By using the bufferless model, the evaluated cell loss rate is never smaller than the potential cell loss rate which is for the actual switching system. The usage parameter control requirement can be provided through defining peak cell transmission speed and average cell transmission speed as the negotiated traffic parameters within the network. There are some discussion that the usage parameter control for average cell transmission speed might not be easy. But, in CCITT standardization, usage parameter control for average cell transmission speed will be likely defined in the near future, as well as definition of peak cell transmission speed which is mandatory traffic parameter for usage parameter control in 1992 B-ISDN recommendation. Therefore, we could hopefully say that the usage parameter control for peak cell transmission speed and for average cell transmission speed would be feasible to fulfill. The requirement of applicability for connection diversity can be provided also through using recursive equations and probability functions in evaluating cell loss rate. Using these two techniques, the amount of calculation in estimating cell loss rate does not increase even when the number of connection types increases. Finally, the proposed method can fairly estimate the statistical multiplexing effect. Therefore, the proposed algorithm can achieve high utilization efficiency of the multiplexed line, while providing safety margin.

Connection admission control (CAC) is performed based on the evaluated cell loss rate assuming a new connection would be established or a connection would be teared down. In other words, when the estimated cell loss rate is larger than the required cell loss rate, the connection set-up request is rejected. In the following sections, the evaluation methods for cell loss rate when a connection is established or teared down are described.

3.2 CAC Without Priority Control

3.2.1 Fundamental Method

In this subsection, the fundamental evaluation method of cell loss rate is described [12].

The estimated cell loss rate in the proposed algorithm is the average cell loss rate over all of connections, but for individual connections. It is indicated that individual cell loss rate for the connections are not generally identical in ATM networks [3],[4]. However, the difference of the cell loss rate among the individual cell loss rate are not so large; the largest cell loss rate is generally at most ten times as large as the smallest cell loss rate. By the computer simulation when many connection types are multiplexed, it would said that the difference of the cell loss rate for each connection is a range of factor, but is not the range of order. Therefore, in this paper, the average cell loss rate over all of connections is used as the cell loss rate of each connection.

One unit-time used in the proposed CAC algorithm is defined as the time that the multiplexer or switch transfers one cell to the multiplexed line. Here, $V(cells/unit-time)$ is defined as the cell transmission speed of the multiplexed line. The peak cell transmission speed of the connection, whose peak cell transmission speed is the largest among the multiplexed connections is defined as $V_{p0}(cells/unit-time)$. The period T is defined as the inverse of V_{p0}, so that the multiplexed line can transfer M_T cells during period T. On the other hand, no connection can transfer more than one cell to the multiplexer during T. Here, M_T is numerically equal to T.

It is assumed that, when the number of cells transferred from connections is larger than M_T cells, the number of cells beyond M_T cells are discarded. Assuming this bufferless model, the estimated cell loss rate is never smaller than the value for real multiplexing system. The number of type $j(1 \leqq j \leqq L)$ connections is N_j. Peal cell transmission speed of a type j connections is V_{pj} and the average cell transmission speed of a type j connections is V_{aj}.

Cell loss rate, CLR, is given by Eq.(1). Here, $P(k)$ means the probability that k cells are transferred from multiplexed connections during T. By assuming that each connection is modeled as transferring cells at the speed of V_{p0} which is maximum peak cell transmission speed of the multiplexed connections, the cell loss rate given by Eq.(1) is surely larger than the actual cell loss rate.

$$CLR = \frac{1}{\rho M_T} \sum_{k=M_T+1}^{\infty} (k - M_T)P(k) \qquad (1)$$

where,

$$P(k) = \sum_{k=\sum_{i=1}^{L}} \prod_{j=1}^{L} \left(\begin{array}{c} N_j \\ nj \end{array} \right) \alpha_j^{n_j}(1-\alpha_j)^{N_j-n_j} \qquad (2)$$

$$\rho = \sum_{j=1}^{L} N_j V_{aj}/V \qquad (3)$$

$$\alpha_j = V_{aj}/V_{p0} \qquad (4)$$

Here, $\rho(0 < \rho \leqq 1)$ is the average offered load to the multiplexed line. $n_j(1 \leqq j \leqq L)$ represents the number of the connections which are in active state. $\alpha_j(1 \leqq j \leqq L)$ is the ratio average cell transmission speed of type j connection to the maximum peak cell transmission speed of multiplexed connections. α_j corresponds to the probability that cell is generated at a certain time slot.

Equation (1) can be transformed as Eq.(5). Through this transformation, the amount of summation is reduces from $[M_T + 1, \infty]$ to $[0, M_T]$.

$$CLR = \frac{1}{\rho M_T} \sum_{k=0}^{M_T} (M_T - k)P(k) - \frac{1-\rho}{\rho} \qquad (5)$$

When one connection of type j is newly established, the new cell loss rate, CLR', is obtained as follows. Here, $\{P'(k)\}$ can be updated using the recursive equations.

$$CLR' = \frac{1}{\rho' M_T}\sum_{k=0}^{M_T}(M_T - k)P'(K) - \frac{1-\rho}{\rho'} \quad (6)$$

$$P'(k) = \begin{cases} (1-\alpha_j)P(0) & \text{(k=0)} \\ (1-\alpha_j)P(k) + \alpha_j P(k-1) & \text{(otherwise)} \end{cases} \quad (7)$$

$$\rho' = \rho + V_{aj}/V \quad (8)$$

When one connection type j is newly established, a new CLR can be obtained by Eq.(6) and $P(k)(0 \leqq k \leqq M_T)$ are updated by Eq.(7). The amount of calculation for each equation is only order of T.

Similarly, when one connection type j is teared down, $\{P'(k)\}$ and ρ' are given by Eq.(9) and Eq.(10).

$$P'(k) = \begin{cases} P(0)/(1-\alpha_j) & \text{(k=0)} \\ \frac{P(k)-\alpha_j P'(k-1)}{1-\alpha_j} & \text{(otherwise)} \end{cases} \quad (9)$$

$$\rho' = \rho - V_{aj}/V \quad (10)$$

As a result, both the amount of calculation for the connection establishing and tearing down are order of T.

With conventional methods, when one connection is newly established or one connection is teared down, the amount of calculation for evaluating the new cell loss rate is the order of L_L [9],[10]. Here, L represents the number of connection types. However, the amount of calculation in the proposed algorithm is the order of T and is independent of L.

Since each connection is modeled as transferring cells at the speed of V_{p0} which is the maximum peak cell transmission speed among the multiplexed connections in the algorithm, the cell loss rate evaluated through the fundamental method is too large compared to the real value. In some cases, the summation of peak cell transmission speed of each connection is smaller than the cell transmission speed of multiplexed line. In the following subsections, two techniques to improve this over estimating are described.

3.2.2 Improvement Technique I

N_j connections type j can transfer exactly N_j cells during period $T_j(=1/V_{pj})$ at the speed of their peak cell transmission speed V_{pj}. Accordingly, N_j connections type j transfer $N_jT/T_j(N_jV_{pj}/V_{p0}$ cells during T. Since the connections whose peak cell transmission speed is $V_{p0}(V_{pk} \leqq V_{p0}$ for any $k)$ transfers exactly one cell during T, N_j connections type j can be considered as equivalent to $N_jT/T_j(= N_jV_{pj}/V_{p0})$ connections each of peak cell transmission speed is V_{p0}. With this procedure, a connection type j is no longer modeled as the connection transferring cells at the speed of V_{p0}, but rather at the speed of V_{Pj}.

However, N_jV_{pj}/V_{p0} is not always an integer, so the cell loss rate is evaluated by the weighted average between the case of $\lfloor N_jV_{pj}/Vp0\rfloor$ and $\lfloor N_jV_{pj}/V_{p0}\rfloor + 1$. Here, $\lfloor x\rfloor$ means the largest integer that is not beyond x. Where, the weighting parameters are a_{j0} for $\lfloor N_jV_{pj}/Vp0\rfloor$ and a_{j1} for $\lfloor N_jV_{pj}/Vp0\rfloor + 1$.

$$\begin{cases} a_{j0} = \lfloor N_jV_{pj}/Vp0\rfloor + 1 - N_jV_{pj}/Vp0, \\ a_{j1} = 1 - a_{j0} \end{cases} \quad (11)$$

When the number of connection types is L, the amount of calculation for the weighted average is order of 2^L. However, when one connection type j is newly established or teared down, only one pair of weighted parameters, a_{j0} and a_{j1}, change. Using the technique described in Appendix, the amount of calculation for the weighted average can be reduced to order to T.

3.2.3 Improvement Technique II

As well known [12], the beneficial effect of statistical multiplexing falls off with increase of peak cell transmission speed for the multiplexed connections. In the analysis of improved method in the previous subsection, N_j of type j connections are modeled as $\lfloor N_jV_{pj}/Vp0\rfloor$ connections whose peak cell transmission speed is V_{p0}. Therefore, the statistical multiplexing effect of those connections is estimated through the statistical multiplexing effect of the connections whose peak cell transmission speed is V_{p0}. The benefit of statistical multiplexing for high speed connection is small than for low speed connections. Here, the high speed connection is the connection whose peak cell transmission speed is not small compared with the cell transmission speed of the multiplexed line, and the low speed connection is the connection whose peak cell transmission speed is small compared with the cell transmission speed of the multiplexed line. When there is great difference among V_{pj}'s, the estimated effect of statistical multiplexing for the low speed connections is underestimated in the proposed method. By the computer simulations, it could be said that the statistical multiplexing effect falls off, when the peak cell transmission speed of the multiplexed connection is larger than around 1/100 of the cell transmission speed of multiplexed line [12]. Therefore, 1/100 is defined as the threshold value whether the multiplexed connection is high speed or low speed connection in this paper. When the ratio of the peak cell transmission speed of the connection to the cell transmission speed of multiplexed line is larger than the threshold value, we can expect small statistical multiplexing gain. On the other hand, when the ratio of the peak cell transmission speed of of the connection to the cell transmission speed of multiplexed line is smaller than the the threshold value, we can expect enough large statistical multiplexing gain. Here, in general, the threshold value will be determined by the system, through the evaluation of statistical multiplexing gain parameterizing the peak cell transmission speed of each connection. The proposed algorithm can be applied to the case, even where the threshold value is changed from the defined value in this paper.

One possible solution is separating the bandwidth exclusively for high speed connections and for low speed connections; exclusive bandwidth separation. Through the exclusive bandwidth separation, we could improve the estimation of statistical multiplexing effect. But, when the bandwidth is separated exclusively between for high speed connection and for low speed connections, the statistical multiplexing effect between these bandwidth are not taken into account. Moreover, relative peal cell transmission speed of each connection, which is the ratio of peak cell transmission speed of the multiplexed connections to the cell transmission speed of the multiplexed line, increases by the bandwidth separation. This is relatively equivalent to reducing the bandwidth of the multiplexed line. As shown in [12], when the peak cell transmission speed of each connection versus the cell transmission speed of the multiplexed line is reduced, we can expect smaller statistical multiplexing gain. As a result, by bandwidth separation, the statistical multiplexing gain is reduced in the analysis. Through the comparison of the maximum allowable system throughputs between for bandwidth separation method using the above algorithm and for without bandwidth separation method using the following algorithm, the method without bandwidth separation can yield generally better throughput than bandwidth separation method. Moreover, from the view point of system implementation, the method without bandwidth separation is easier to implement than with bandwidth separation: bandwidth separation approach needs extra hardware and extra control algorithm, such as time frame arbitration, for the switching system.

The following technique is proposed in order to solve the over-estimation problem of the cell loss rate, due to evaluating the statistical multiplexing effect assuming low speed connection to transfer cells at the speed of high speed connection. While, the proposed algorithm described below can still provide safety margin: the estimated cell loss rate is surely larger than the actual cell loss rate of the system.

T_1 is defined as the inverse of V_{p1} that is the maximum cell transmission speed of low speed connections, $T_1 = 1/V_{p1}$. The low speed connection is defined as the connection whose peak cell transmission speed is smaller than the threshold value ($V_{p1} = V/100$). One contrary, the high speed connection is defined as the connection whose peak cell transmission speed is larger than $V_{p1}(= V/100)$.

The following bufferless model is assumed. When the number of cells which are transferred from high speed and low speed connections are larger than M_{T1} cells, the number of cells beyond M_{T1} cells are discarded. This is because that the multiplexed line can transfer M_{T1} cells during T_1. Here, M_{T1} is numerically equal to T_1. In order to evaluate the cell loss rate for this bufferless model, we obtain the probability function for the number of transferred cells from the multiplexed connections during T_1. The probability $\{G(n)\}$ that n cells are transferred from the multiplexed connections during T_1 is defined. Here, $F(k)$ is the probability that k cells are transferred from HIGH speed connections during T_1. $P(k)$, which is given by Eq.(2), is the probability that k cells are transferred from LOW speed connections during T_1.

$G(n)$ is given by Eq.(12).

$$G(n) = \sum_{k=0}^{n} P(k)F(n-k), \ (0 \leqq n \leqq M_{T1}) \tag{12}$$

And, then, the cell loss rate is given by Eq.(13).

$$CLR = \frac{1}{\rho M_{T1}} \sum_{n=M_{T1}+1}^{\infty} (n - M_{T1})G(n) \tag{13}$$

Equation (13) can be transformed as Eq.(14), reducing the summation from $[M_{T1}+1, \infty]$ to $[0, M_{T1}]$.

$$CLR = \frac{1}{\rho M_{T1}} \sum_{n=0}^{M_{T1}} (M_{T1} - n)G(n) - \frac{1-\rho}{\rho} \tag{14}$$

In order to estimate cell loss rate, the probability function $F(k)$ must be obtained. $F(k)$ is the probability that k cells are transferred from high speed connections during T_1. Since, T_1 is larger than the inverse value of the peak cell transmission speed of high speed connections, high speed connections can transfer more than one cell during T_1. In order to calculate the cell loss rate that is never smaller than the actual cell loss rate, the worst case cell transmission pattern is assumed to evaluate cell loss rate. When $F(k)$ represents the worst case cell transmission pattern, the evaluated cell loss rate is surely larger than the real cell loss rate.

It can be assumed that each connection keeps its peak cell transmission speed and average cell transmission speed when usage parameter control works. The worst case cell transmission pattern, which gives the largest cell loss rate, is the case where a high speed connection transfers cells continuously at the speed of its peak cell transmission speed V_{pj} during T_1. Since each connection keeps its average cell transmission speed V_{aj}, the probability that a connection type j transfers with the worst case pattern is V_{aj}/V_{pj}.

The worst case for $F(k)$ mentioned above is given by Eq.(15) and Eq.(16). Here, $Q_s(n_s)$ is the probability that n_s cells are transferred from type s connections during T_1. N_s is the number of high connections type $s (1 \leqq s \leqq L_h)$, and L_h is the number of high speed connection types.

$$F(k) = \sum_{k=n_1+n_2+\ldots+n_{L_h}} \prod_{s=1}^{L_h} Q_s(n_s) \tag{15}$$

$$Q_s(n_s) = \begin{cases} \begin{pmatrix} N_s \\ l \end{pmatrix} \alpha_s^l (1-\alpha_s)^{N_s-l} & (n_s = l\lceil V_{ps}T_1 \rceil) \\ 0 & \text{(otherwise)} \end{cases} \tag{16}$$

Here, $\lceil x \rceil$ means the smallest integer equal to or greater than x. Through the following recursive equations, the calculation for updating $F(k)$ is order of T_1. Equation (17) gives the new probability function $F'(k)$ where a connection j is newly established, and Eq.(18) gives $F'(k)$ where a connection j is teared down. Here, $F(k)$ is the probability function before a new connection is established or teared down.

$$F'(k) = \begin{cases} (1-\alpha_j)F(k) + \alpha_j F(k-m_j) & (m_j \leqq k) \\ (1-\alpha_j F(k) & \text{(otherwise)} \end{cases} \tag{17}$$

$$F'(k) = \begin{cases} \frac{F(k)-\alpha_j F'(k-m_j)}{1-\alpha_j} & (1 \leqq k \leqq M_{T1}) \\ F(0)/(1-\alpha_j) & (k=0) \end{cases} \tag{18}$$

where,

$$m_j = \lceil V_{pj} T_1 \rceil \tag{19}$$

As a result, the amount of calculation for obtaining a new $\{P(k)\}$, which is the set of probabilities for LOW speed connections, and the calculation for obtaining a new $\{F(k)\}$, which is for HIGH speed connections, are only order of T_1 when a new connection is established or teared down. This means that the calculation for evaluating cell loss rate is order of T_1^2, totally. Moreover, the amount of calculation is independent of the number of connection types. Therefore, the amount of calculation does not increase even when the number of connection types increases.

3.3 CAC With Priority Control

3.3.1 CAC Algorithm with Priority Control

In this subsection, the connection admission control method with priority control is described. Here, priority control means that the cells from priority connections are served first at queuing point. The cells from non-priority connections are served only when there is no priority cell at queuing point. The proposed algorithm takes into account the statistical multiplexing effect between priority and non-priority connections.

Once again, $T_1 (= 1/V_{p1}$ is defined as the inverse of peak cell transmission speed for the connection whose peak cell transmission speed is the maximum value among the low speed connections. Then, $\{G_p(n)\}$ and $\{G_{np}(n)\}$ are defined as the probability functions for the number of transferred cells during T_1 for priority and non-priority connections respectively. $G_p(n)$ and $G_{np}(n)$ are defined by Eq.(20).

$$G_i(n) = \sum_{k=0}^{n} F_i(k) P_i(n-k), (\text{i=p, np}) \tag{20}$$

The following bufferless model is assumed. Assume that n_p cells are transferred from priority connections and n_{np} cells are transferred from non-priority connections during T_1. The multiplexer or multiplexed line tries to deal with the cells from priority connections first. Therefore, when n_p is larger than M_{T1}, $(n_p - M_{T1})$ cells from priority connections are discarded during T_1. The cells from non-priority connections are dealt with after transferring priority cells. So, when n_p ($n_p \leqq M_{T1}$) cells are transferred from priority connections, the multiplexer can transfer $(M_{T1} - n_P)$ non-priority cells during T_1. When n_{np} is larger than $(M_{T1} - n_P)$, $\{n_{np} - (M_{T1} - n_p)\}$ cells are discarded. When n_p is larger than M_{T1}, all of non-priority cells are discarded.

Assuming this bufferless model, we can obtain the expected cell loss rate for priority connections, CLR_p, and the expected cell loss rate for non-priority connections, CLR_{np}. Here, $CLR(n)$ is defined as the the expected cell loss rate for non-priority connections. when n cells are transferred from priority connections during T_1.

$$CLR_p = \frac{1}{\rho_p M_{T1}} \sum_{n=0}^{M_{T1}} (M_{T1} - n) G_p(n) - \frac{1-\rho_p}{\rho_p} \tag{21}$$

$$CLR_{np} = \sum_{n=0}^{M_{T1}} G_p(n) CLR(n) + \sum_{n=M_{T1}}^{\infty} G_p(n) \tag{22}$$

$$= \sum_{n=0}^{M_{T1}} G_p(n) CLR(n) + (1 - \sum_{n=0}^{M_{T1}} G_p(n) \tag{23}$$

$$CLR(n) = \frac{1}{\rho_{np} M_{T1}} \sum_{k=M_{T1}-n+1}^{\infty} [k - (M_{T1} - n)] G_{np}(k) \tag{24}$$

$$= \frac{1}{\rho_{np} M_{T1}} \sum_{k=0}^{M_{T1}-n} [(M_{T1} - n) - k] G_{np}(k) + \frac{\rho_{np} M_{T1} - (M_{T1} - n)}{\rho_{np} M_{T1}} \tag{25}$$

where,

$$\rho_p = \sum_{priorityconnection} N_i V_{ai}/V \tag{26}$$

$$\rho_{np} = \sum_{non-priorityconnection} N_i V_{ai}/V \tag{27}$$

The calculation for obtaining $\{G_i(n)\}$ is the order of T_1^2 and the summation of Eq.(21) and Eq.(23) are the order of T_1. Therefore, the amount of calculation for this proposed algorithm is the order of T_1^2. Moreover, the amount of calculation for this proposed method is independent of the number of connection types, similarly to the algorithm proposed in the previous subsection.

3.3.2 Discussion

The proposed algorithm in this subsection assumes the priority control model that the cells for non-priority connections are only served when there is no cells from priority connections. Therefore, the average queuing delay for non-priority cells might be large because of head line of blocking by the priority cells in the buffer. However, by the following two reasons, the average queuing delay of the cells from non-priority connections should not be large.

- Even when the priority control is performed, the average queuing delay of the cells that is the average value among priority and non-priority connections is still small, say less than five unit-time, as same as evaluation result in Sect.2. As with the average delay of the cells of priority connections, it can be smaller than the above average queuing delay, but it is not too small. Say, the average queuing delay of the cells from priority connections should be few unit-times. Therefore, it can be said that unless the offered load of non-priority connections is extremely small compared to the total offered load for priority and non-priority connections, the average queuing delay of the cells from non-priority connections is also small, say less than ten unit-times.

- The proposed method assumes the pseudo-bufferless switching model. It is assumed that only when n is larger than T_1, $(n-T_1)$ cells are lost in the analysis. Here, n is the number of cells transferred from the multiplexed connections during T_1. Otherwise ($n \leq T_1$), all of transferred cells are served at queuing point during T_1. Therefore, the evaluated cell loss rate corresponds to the buffering probability of the cells in the actual switching system that has a queuing buffer. The discussed cell loss rate is fairly small value say less than 10^{-6}. This means that the probability that the cells from non-priority connections are not served is fairly small. For example, when cell loss rate of the cells for the priority connections is 10^{-6}, the probability that the cells from non-priority connections are queued in the buffer is only 10^{-6}. This indicates that the increase of the queuing delay of the cells for non-priority connections due to the head of line blocking by the priority cells in the buffer should not be large. Here, in the actual switching system, even when n is smaller than T_1, the cells are queued. However, the queuing delay at this case should be small, because n cells are surely served during T_1.

The proposed method is basically applicable for the assumed priority control policy in this paper. However, in the actual system, other priority control policies would be performed. The methods to apply to the other priority control policies are for further study.

4. Performance Evaluation

4.1 Evaluations for Multiplexer System

In this subsection, the performance of the proposed method is evaluated for the output buffered multiplexing system. In order to evaluate how the proposed methods take into account of the effect of statistical multiplexing, the performance of the proposed methods is compared with the performance when all connections are allocated their peak cell transmission speed (peak assignment strategy).

4.1.1 Evaluation Model

The performance of the proposed method is compared with computer simulation. High speed connections and low speed connections, which generate cells with two-state Markov chain model, are multiplexed. The high speed connection has peak cell transmission speed that is $V_p = V/10$. The low speed connection has peak cell transmission speed that is $V_p = V/100$.

On the contrary, the maximum allowable throughput by using the proposed CAC algorithm is evaluated by numerical analysis. The traffic parameters to be used in the algorithm in order to evaluate cell loss rate are peak and average cell transmission speed for each connections.

4.1.2 Performance Evaluation

Figure 5 shows the maximum offered load of low speed connections versus the offered load of high speed connections, when both of connections require the smaller cell loss rate than 10^{-5}. The high speed connection is $V_p = V/10$, $V_a = V_p/10$, $M_a = 100$ cells and $M_i = 900$ cells. The low speed connection is $V_p = V/100$, $V_a = V_p/2$ and $M_a = M_i = 900$ cells. Here, again, M_a is the average burst length and M_i is average idle length. The method which is described in Sect.3.2.1 is denoted as "Fundamental Method". On the contrary, the improved method that applies both technique I and technique II is denoted as "Proposed Method" in the figure.

As shown, when the offered load of high speed connections is small, say at less than 6%, the fundamental

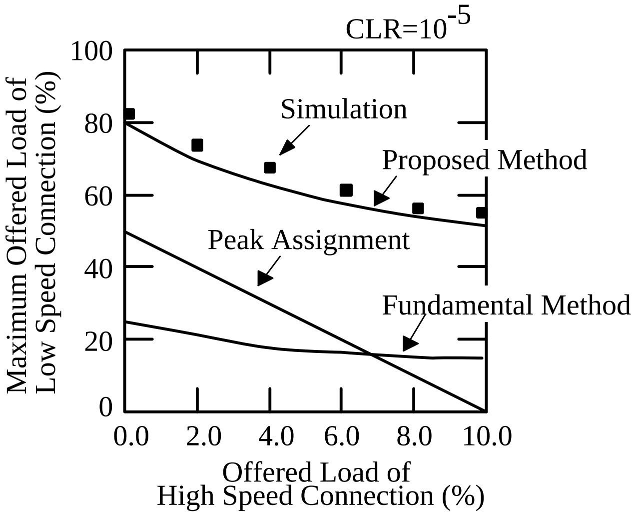

Fig. 5 Maximum offered load of low speed connections vs. offered load of high speed connections

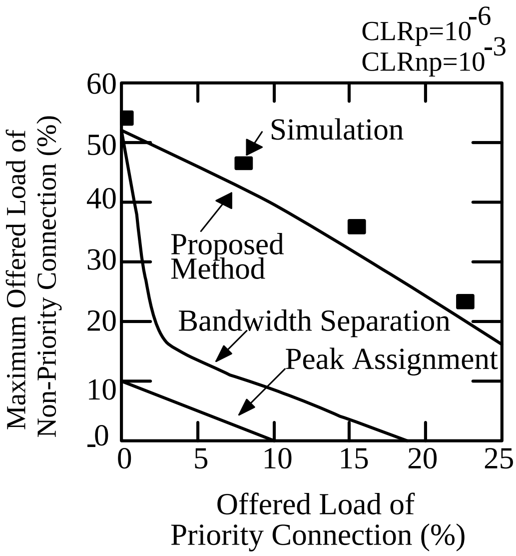

Fig.6 Maximum offered load of non-priority connections vs. offered load of priority connections

method yields smaller throughput than the peak assignment strategy does. This is derived from too large safety margin associated with the evaluated cell loss rate using the fundamental method: the estimated cell loss rate is too large. This indicates that the fundamental method is too secure algorithm. The estimated offered load is always smaller than the actual maximum offered load to the multiplexer, but it takes into account negative or so small statistical multiplexing effect.

On the contrary, the proposed method can take into account statistical multiplexing effect, fairly well. Also, it is shown that the proposed method can operate in safe side. This means that the estimated maximum offered load using the proposed method is never larger than the actual maximum offered load which is estimated by the computer simulation, while it takes into account of statistical multiplexing effect.

The performance of the proposed method, when the priority control is performed, is evaluated. The bandwidth is shared among priority connections and non-priority connections. Therefore, the effect of statistical multiplexing among these connections are taken account.

Figure 6 shows the maximum allowable offered load of non-priority connections versus the load of priority connections. Priority connection requires the cell loss rate which is less than 10^{-6}, and non-priority connection requires the cell loss rate which is less than 10^{-3}. Both for priority connections and for non-priority connections, there are high speed connections and low speed connections. The high speed connection is $V_p = V/10, V_a = V_p/10, M_a = 100$ cells and $M_i = 900$ cells. The low speed connection is $V_p = V/100, V_a = V_p/10, M_a = 100$ cells and $M_i = 900$ cells. Here, both for priority connections and for non-priority connections, the percentage of high speed connections is 50%. 50% of the multiplexed connections is high speed connection and the other 50% is low speed connections.

Here, in the figure, the performance of "Bandwidth Separation" method is shown. In the Bandwidth Separation method, the bandwidth of the multiplexed line is separated exclusively by the cell loss rate requirement class. Bandwidth Separation method is one method to provide multiple QOS classes simultaneously without applying priority control. In this evaluation, the bandwidth is separated exclusively for priority connections and non-priority connections. This means that the performance of Bandwidth Separation method is equal to the performance when priority control is not performed in order to provide multiple cell loss rate classes. As shown in the figure, by applying the priority control associated with cell loss rate requirement class, we can expect fairly higher throughput compared to the case where priority control is not performed. For example, when the offered load of priority connections is 10%, the multiplexer applying the priority control can yield about 250% of the system throughput compared to the throughput of the system that does not apply the priority control; throughput of the proposed method is about 50%, and that of Bandwidth Separation method is only about 20%.

Of course, compared to the simulation results, the proposed method operates with safety margin. The evaluated throughput by the proposed method is never larger than the throughput which is measured by the

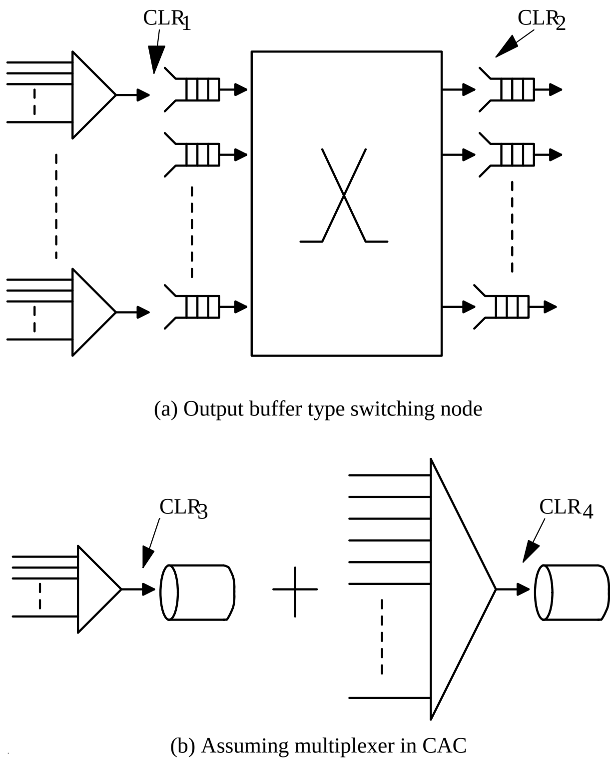

Fig.7 Switching node model

computer simulation.

4.2 Evaluations for Switching System

4.2.1 CAC on a Switching System

Basically, the proposed method in the previous section is for an output buffer type multiplexer. As shown in the following evaluation, the proposed method can apply not only to a multiplexer but also to a switch. Figure 7 (a) shows the evaluated system model which is an output buffer type switching system. With a real switching system, the cell may be multiplexed first, for instance, at a remote multiplexer, before cells enter to the switch. The cell loss rate at these multiplexer are defined as CLR_1. After going through the buffers in the multiplexer, cells are switched. Cell loss rates at the switch output buffer are defined as CLR_2 – assuming a perfect switch fabric which means non-blocking switch. In computer simulation, when both of measured CLR_1 and CLR_2 are smaller than the cell loss rate which is requested by the connection at connection set-up procedure, the connection set-up request is accepted. Otherwise, the connection set-up request is rejected.

In order to apply the proposed method to the switch shown in Fig.7 (a), the cell loss rate of each connection is evaluated assuming the model shown in Fig.7 (b). CLR_3 and CLR_4 are the estimated cell loss rate by the proposed methods described in the previous section. In the CAC using the proposed algorithm, when either CLR_3 or CLR_4 is larger than the required cell loss rate of each connection, the connection set-up request is rejected.

4.2.2 Evaluation Model

The performance evaluation is as shown in Fig.7 (a). The switch is an 8x8 output buffer type whose buffer size is 100 cells. The number of both input and output ports are eight. Each multiplexer before the switch fabric has 100 cells output buffer. Input port of the multiplexer and output port of the switch fabric are selected randomly at connection set-up – random traffic. High speed connections and low speed connections are multiplexed. The high speed connection is $V_p = V/10, V_a = V_p/10, M_a = 100$ cells and $M_i = 900$ cells. The low speed connection is $V_p = V/100, V_a = V_p/2$, and $M_a = M_i = 900$ cells. Here, again, M_a is the average burst length and M_i is the average idle length.

On the contrary, the maximum allowable throughput by using the proposed CAC algorithm is evaluated by numerical analysis, assuming the model shown in Fig.7 (b).

4.2.3 Performance Evaluation

Figure 8 shows the maximum offered load of low speed connection versus the offered load of high speed connections, when the cell loss rate for each connection is smaller than or equal to 10^{-5}.

Figure 9 shows the maximum offered load of non-priority connections versus the offered load of priority connections. Here, the required cell loss rate for priority connections is less than or equal to 10^{-5} and the required cell loss rate for non-priority connection is 10^{-3}. Both for priority connections and for non-priority connections, the percentage of high speed connection is 50%

Again, compared to the simulation results, the proposed method never ever-estimate, which means that the maximum allowable load evaluated by the proposed algorithm is never larger than the load measured by simulation.

5. Conclusion

This paper proposes a connection admission control method which can be applied to the priority control case. It is shown that, because of cell loss rate sensitivity of ATM traffic, the connection admission control should be performed based on cell loss rate.

The proposed connection admission control method has the following features. First, the

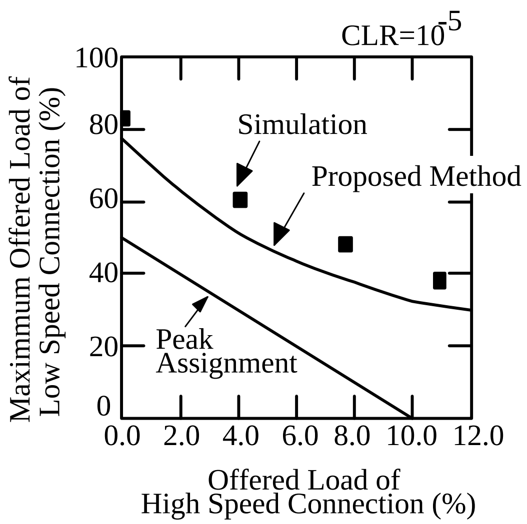

Fig.8 Maximum offered load of low speed connections vs. offered load of high speed connecitons

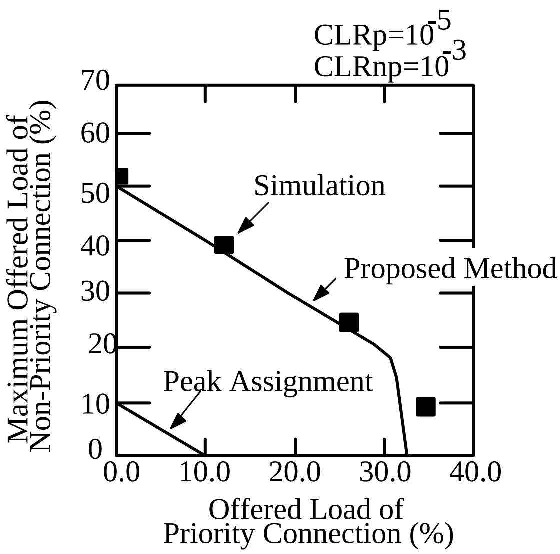

Fig.9 Maximum offered load of non-priority connections vs. offered load of priority connecitons

amount of calculation of algorithm is reduced dramatically, and does not increase even when the number of connection types increase. Second, evaluated cell loss rate is never smaller than the real cell loss rate due to bufferless model. Third, usage parameter control could be performed easily, because the traffic parameters used in the algorithm are only peak cell transmission speed and average cell transmission speed.

The proposed algorithm can apply to the case where the multiple QOS classes associated with cell loss rate requirements are provided. The proposed algorithm can apply to the case where there are multiple priority classes on cell loss rate requirements using priority control mechanisms.

References

[1] Ma, B.N. and Mark, J.W.,"Performance Analysis of Burst Switching for Integrated Voice/Data Services", *IEEE Trans. on Communications*, vol.COM-36, no.3, pp.282-297, Mar. 1988.
[2] Hirano, M. and Watanabe, N., "Characteristics of a Cell Multiplexer for Bursty ATM Traffic", *ICC89*, 13.2, vol.1, pp.399-403, Jun. 1989.
[3] Murata, M. et al, "Analysis of a Discrete-Time Single-Server Queue with Bursty Inputs for Traffic Control in ATM Networks", *IEEE Journal on Selected Area in Communications*, vol.8, Mar. 1990.
[4] Yamada, H. et al., "Characteristics of Statistical Multiplexing with Heterogeneous Inputs", *IEICE Technical Report*, SSE90-49, Jul. 1990.
[5] Kamitake, T. and Suda, T., "Evaluation of Admission Control Scheme for an ATM Network Considering Fluctuations in Cell Loss Rate", *GLOBECOMM89*, 49.4, vol.3, pp.1774-1780, Nov. 1989.
[6] Ohnishi, H., Okada, T. and Noguchi, K., "Flow Control Scheme and Delay/Loss Tradeoff in ATM Network", *IEEE Journal on Selected Area in Communications*, vol.6, no.9, pp.1609-1616, Dec. 1988.
[7] Eckburg, A.E., Kuan, D.T. and Lucantoni, D.M., "Meeting the Challenge: Congestion and Flow Control Strategies for Broadband Information Transport", *GLOBECOM89*, 49.3, vol.3, pp.1769-1773, Nov. 1989.
[8] Hui, J.Y., "Resource Allocation for Broadband Networks", *IEEE Journal on Selected Area on Communications*, vol.6, no.9, pp.1598-1608, Dec. 1988.
[9] Joos, P. and Verbiest, W., "A Statistical Bandwidth Allocation and Usage Monitoring Algorithm for ATM Network", *ICC89*. 13.5, vol.1, pp.415-422, Jun. 1989.
[10] Murase, T., Suzuki, H. and Takeuchi, T., "A Call Admission Control for ATM Networks", *ICC91*, vol.6.5, Jun. 1991.
[11] Jabbari, B. and Yegenolu, F., "An Upper Bound of Cell Loss Priority of Bursty Sources in Broadband Packet Networks", *ICC91*, vol.2, 23.3, Jun. 1991.
[12] Esaki, H., Iwamura, K. and Kodama, T., "Simple and Effective Admission Control Method for an ATM Network", *GLOBECOMM90*, 300.4, vol.1, Dec. 1990.
[13] Esaki, H., "Call Admission Control Method in ATM Network", *ICC92*, 354.4, vol.3, Jun. 1992.
[14] Endo, N., Ohuchi, T., Kozaki, T. and Kuwahara, H., "Traffic Characteristics Evaluation of a Shared Buffer ATM Switch", *GLOBECOMM90*, 905.1, vol.3, Dec. 1990.

Appendix A.

Define $\{R_k(n_1, n_2, ..., n_L) | 0 \leq k \leq M_T\}$ as the set of probabilities that k cells are transferred from all multiplexed connections, when the number of connections type j is n_j. Here, n_j corresponds to $\lfloor N_j V_{pj}/V_{p0} \rfloor$. Then $P(k)$ and $R_k(n_1, n_2, ..., n_L)$ are given by Eq.(28) and Eq.(29).

$$P(k) = a_{j0} R_k(.., n_j, ..) + a_{j1} R_k(.., n_j + 1, ..) \quad (28)$$

$$R_k(.., n_j + 1, ...) = \alpha_j R_{k-1}(..., n_j, ...) + (1 - \alpha_j) R_k(.., n_j, ..) \quad (29)$$

where, $R_{-1}(..., n_j, ...) = 0$.

When a_{j0}, a_{j1} and $P(k)$ are known, we can calculate $\{R_k(n_1, n_2, ..., n_L)\}$ using Eq.(29).

When type j connection is newly established, $\{P'(k)\}$ can be obtained by Eq.(30). Here, a'_{j0} and a'_{j1} are the updated values.

$$P'(k) = a'_{j0} R_k(..., n_j, ...) + a'_{j1} R_k(..., n_j + 1, ...) (30)$$

As a results, the amount of calculation for $\{P'(k)\}$ is order of T.

Incidentally, when $\lfloor (N_j + 1) V_{pj}/V_{p0} \rfloor$ becomes $n_j + 1$, we must obtain $\{R_k(.., n_j + 2, ..)\}$. We can obtain $\{R_k(.., n_j + 2, ..)\}$ by the order of T calculation, using Eq.(29).

Hiroshi Esaki received the B.S. and M.S. degrees from Kyushu University, Fukuoka, Japan, in 1985 and 1987, respectively. In 1987, he joined Research and Development Center, Toshiba Corporation, where he engaged in the research of ATM systems. He has been at Bellcore in New Jersey (USA) as a residential researcher from 1990 to 1991, and has engaged in the research on high speed computer communications. He is currently interested in ATM internet architecture and high speed computer networks.

Kazuaki Iwamura received the B.S. and M.S. degrees from Kyoto University, Kyoto, Japan, in 1986 and 1988, respectively. In 1988, he joined Research and Development Center, Toshiba Corporation. Since 1991, he has been with Kansai Research Laboratory, Toshiba Corp., where he has been engaged in the research of B-ISDN switching systems.

Toshikazu Kodama received the B.E., M.E. and Ph.D degrees in electrical engineering from Tokyo Institute of Technology, Tokyo, Japan, in 1971, 1973, 1987, respectively. In 1973, he jopined the Toshiba Research and Development Center, Kawasaki, Japan, where he has been engaged in the research and development of surface acoustic wave devices and ATM switching systems. He is currently a Chief Research Scientist at the Communication and Information Systems Labs., Toshiba Research and Development Center.

Takeo Fukuda received the B.S. and M.S. degrees from Kyushu University, Fukuoka, Japan, in 1963 and 1965, respectively. In 1965 , he joined Research and Development Center, Toshiba Corporation, where he engaged in the research and development of telecommunication systems and equipment. He is currently a General Manager of Communication Systems and Technology Laboiratory, Toshiba Corporation.