A Dynamic Bandwidth Allocation Scheme for ATM Networks

Zsehong Tsai, Wen-der Wang, Jin-Fu Chang[t]

Dept. of Electrical Engineering
National Taiwan University
Taipei, Taiwan, R.O.C.

Dept. of Electrical Engineering[t]
National Central University
Chung-li, Taiwan, R.O.C.

Abstract

A multiple link and multiple traffic type ATM network environment is considered. In order to efficiently utilize network resource at low network management cost, we propose a dynamic bandwidth allocation scheme using call admissible regions as a control basis. The allocated bandwidth for a traffic stream belonging to a Virtual Path is set to be an integer multiple of the basic required bandwidth of a single call, which is carried by a Virtual Channel. In addition, the system state vector, a tuple of the number of calls of each traffic type, is required to be within the call admissible region such that the GoS for all traffic types is satisfied. We propose a generic mathematical model for the analysis of our dynamic bandwidth allocation scheme. Using the obtained analytical formulas, the bandwidth utilization and management cost (measured in normalized bandwidth update frequency) of two simple experimental networks are obtained. We conclude that by carefully selecting the bandwidth allocation unit, one can significantly reduce the network management cost in bandwidth adjustment and still maintain high utilization, even under a heterogeneous traffic environement.

1 Introduction

The most outstanding feature of ATM network is its flexibility in supporting a wide range of services. It is allowed in ATM networks to directly multiplex cell streams with various characteristics into the transmission links. However, it is certainly impossible to accept all connection requests without limitations. Therefore the network needs to determine whether to accept a connection request on a call-by-call basis. Such a decision process is often referred to as Connection Acceptance Control[Gri90] or

Admission Control. Under a well-designed connection acceptance control policy, the cell-level Grade-of-Service can be guaranteed and the resource(bandwidth) usage could be highly efficient. However, call request processing can still be a significant load of network management if the bandwidth allocation procedure has to be performed at each call connection. Fortunately, the ATM networks employ the concept of Virtual Path(VP) and Virtual Channel(VC) and allow bandwidth management at both VP and VC levels. With bandwidth management at the VP level, it is possible to reduce the bandwidth management load significantly. For example, in [OhS92], [OST88], and [SOT90], a dynamic scheme for VP bandwidth control has been proposed for single-link, single-traffic-type systems. In their work, both the management load and link utilization are found to be highly related to the VP bandwidth control policy.

In this paper, we extend the dynamic bandwidth allocation scheme for a multiple-link, multiple-traffic-type system. Both the bandwidth allocation scheme and the mathematical model are generic and can be employed along with a variety of connection acceptance control policies. In addition, we also investigate the effect of this bandwidth allocation scheme on various system performance measures and provide two optimization rules for different network requirements.

2 Dynamic Bandwidth Allocation Scheme

In traditional circuit-switched or STM-based networks where the required bandwidth for each call is deterministic, the connection acceptance control is simple. In order to accept an incoming call, the residual capacity of each transmission link on the designated route needs to be larger or equal to the requested bandwidth. One can thus control the bandwidth of a traffic stream as follows. The allocated bandwidth for the traffic stream can be set to be an integer multiple of the basic bandwidth required by a single call, where the integer can be said to be the

bandwidth allocation unit. When the bandwidth is completely used by the succeeding calls and a new connection request is issued, the network reallocates bandwidth to the stream. If the bandwidth is not available, the reallocation fails and the subsequent calls will be blocked.

Unfortunately, the situation is much more complicated in ATM-based networks, in which a call can not be characterized simply by a deterministic bandwidth value, but rather by a set of traffic descriptors, e.g., average bit rate, peak bit rate, burstiness and burst length. Bandwidth assignment is hence not an easy task. A call can be accepted only if the resultant cell loss probabilities on the links of its route are below a preset constraint. Most of the previous works are devoted to achieving high efficiency of direct multiplexing, yet still guaranteeing cell-level Grade-of-Service (GoS). In these articles, for example, [Hui88], [KHB91], and [MSS91], call admissible sets(regions) concerning the mixture of heterogeneous traffics are often regarded as significant results. A call admissible set is a set of multi-dimensional state vectors, of which each element indicates the number of calls of a stream. State vectors in the call admissible set are said to be "feasible". Consequently, the call acceptance control mechanism can be simplified to a procedure of checking the feasibility of the resultant state vector to determine whether a call shall be accepted.

The call admissible regions turn out to be dependent on the traffic parameters, priority control mechanisms, and bandwidth assignment algorithms. In order to simplify the bandwidth control mechanisms, some articles have proposed the use of virtual bandwidth [Gri90, MST91]. The virtual bandwidth of a call is defined to be the link capacity divided by the maximum number of acceptable calls, provided only this type of calls are accepted. There are other different definitions of virtual bandwidth [GAN91], but the basic concepts are similar. In the following, we adopt the concept of virtual bandwidth to explain our bandwidth allocation scheme and take it as a performance measuring basis.

Since the call acceptance control assures that the number of calls in the network shall not lead to violation of the cell-level GOS requirements, the dynamic bandwidth allocation scheme needs to be employed along with an appropriate call acceptance control policy designed for ATM networks. For connection acceptance control, the number of calls of a stream, say stream-i, denoted as x_i, needs to be monitored. As the dynamic control scheme is employed, an additional parameter x_i' is required to indicate the allocated bandwidth. We define x_i' to be the smallest integer multiple of s which is greater than or equal to x_i, where s is an integer called *bandwidth allocation unit*. Thus the bandwidth allocated for the stream is then implicitly equal to x_i' times the virtual bandwidth of that type of

call. If a call arrives when $x_i = x_i'$, implying that the allocated bandwidth is completely utilized by connected calls, then the bandwidth needs to be reallocated and x_i' shall be increased by s. On the contrary, if a call terminates such that x_i turns to be equal to $x_i' - s$, then an amount of bandwidth shall be released and x_i' is decreased by s. By keeping the two records x_i and x_i', a connection establishment procedure combining the dynamic bandwidth control and the connection acceptance control can be efficiently realized.

For example, if we implement the dynamic bandwidth allocation scheme on the VP-level bandwidth control, the VC connection request could be processed as follows: When the number of calls x_i is smaller than x_i', then GoS is guaranteed and the VC is established without further processing. If $x_i = x_i'$, the network tries to add bandwidth to this VP by increasing x_i' to $x_i' + s$. In allocating bandwidth, the network simply checks whether the state vector composed of x_i' is feasible. Such a connection establishment procedure is found to be efficient if the bandwidth increment size is large.

Intuitively, one can predict that the usage of the network resources is less efficient than the call-by-call allocation strategy, because bandwidth may be allocated but is not immediately used. However, the bandwidth management cost is also reduced since the bandwidth adjustment is not evoked at each call connection or termination. We will concentrate on the decision of appropriate sizes of bandwidth allocation units to reduce the processing cost, without significantly losing the efficiency of resource usage.

3 The Mathematical Model And Analytical Results

By using the multiple-link, multiple traffic-type system model, it is possible to study the effect of network topology, and interaction between different traffic types. The term *link* used in the following paragraphs could represent a physical transmission line, or a VP with deterministic bandwidth.

3.1 System Descriptions

The considered system is a network consisting of a set of links indexed from 1 to R, which conveys traffics of various types. The capacity of link r (for $r = 1, 2, \cdots, R$) is C_r. Here we do not assume any specific topology. A fixed routing strategy is assumed. Routing in this network is defined by the path traversed by calls of a given traffic stream. A call is lost if it is blocked by any of the links in its path, and no call buffering is allowed. We also define the following traffic parameters.

- L traffic types are assumed. Calls of a type have identical traffic descriptors and have identical *cell-level characteristics*, e.g., distributions of cell arrival process, burst size and burstiness.

- The (virtual) bandwidth requirement for each call of type l is defined to be b_l.

- The network supports N traffic streams. Calls of a stream should belong to the same type, traverse identical routes, and have identical *call-level characteristics*, e.g., distributions of call holding time. The call arrival and service completion of different streams are mutually independent. All call arrival processes are Poisson and call holding time is exponentially distributed.

- For the i-th stream, the following parameters are given:

$$\lambda_i = \text{mean arrival rate}$$
$$\mu_i = \text{service rate}$$
$$t_i = \text{traffic type to which the calls of}$$
$$\text{stream } i \text{ belong}$$
$$\rho_i = \frac{\lambda_i}{\mu_i}, \text{ traffic intensity}$$
$$s_i = \text{bandwidth allocation unit}$$
$$\mathbf{d}_i = (\delta_{i1}, \delta_{i2}, \cdots, \delta_{iR})$$
$$\text{where } \delta_{ir} = \begin{cases} 1, & \text{if link } r \text{ is in a} \\ & \text{stream } i \text{ path} \\ 0, & \text{otherwise} \end{cases}$$

For each link r $(r = 1, \cdots, R)$, a call admissible set denoted as Ω_r can be defined. Ω_r is a set of L-dimensional vectors of nonnegative integers. Here Ω_r could be convex or nonconvex. The only assumption on it is the *coordinate convexity*, which sustains as long as the call completion is never blocked (see [Aei78]). We follow the definition of coordinate convexity given in [JoV91].

Definition : Let $y \equiv (y_1, y_2, \cdots, y_L)$ and Ω is a subset of the set of L-tuples of nonnegative integers. The set Ω is coordinate convex if the following proposition holds:

If $y \in \Omega$ and $y_i \geq 1$, then $(y_1, \cdots, y_i - 1, \cdots, y_L) \in \Omega$.

3.2 System Steady-State Behavior

In the following, the system state, denoted as $\mathbf{x}$, and the link state, denoted as $\mathbf{y}_r$, $(r = 1, 2, \cdots, R)$, are defined.

- $\mathbf{x} \equiv (x_1, x_2, \cdots, x_N)$, where x_i is the number of active calls from stream i.

- $\mathbf{y}_r \equiv (y_{r1}, y_{r2}, \cdots, y_{rL})$, where y_{rl} is derived from the state vector $\mathbf{x}$ by the following equation:

$$y_{rl} = \sum_{i=1}^{N} \lceil x_i / s_i \rceil s_i \delta_{ir} I(t_i = l), \quad r = 1, 2, \cdots, R;$$
$$l = 1, 2, \cdots, L \tag{1}$$

$$\text{where } I(t_i = l) = \begin{cases} 1, & \text{if } t_i = l \\ 0, & \text{otherwise.} \end{cases}$$

Besides, the following notations are defined for convenience.

- $\mathbf{x} + e_i \equiv (x_1, x_2, \cdots, x_i + 1, \cdots, x_N)$
- $\mathbf{x} - e_i \equiv (x_1, x_2, \cdots, x_i - 1, \cdots, x_N)$

The state space of $\mathbf{x}$, denoted as Λ, can now be defined as follows.

$$\Lambda = \{\mathbf{x} \mid \mathbf{y}_r \in \Omega_r \text{ for } r = 1, 2, \cdots, R, \text{ where each}$$
$$\text{component } y_{rl} \text{ of } \mathbf{y}_r \text{ is given}$$
$$\text{by equation (1)} \} \tag{2}$$

The definition of Λ implies that the cell-level performance on each link is guaranteed, provided that the present state $\mathbf{x}$ falls in Λ. Since Λ is constructed on the basis of Ω_r, it inherits the property of coordinate convexity. The coordinate convexity of Λ makes it easier to evaluate the system performance.

Define $\mathbf{X}(t)$ to be the system state at time t. $\mathbf{X}(t)$ is a continuous-time stationary Markov process which takes on values from the state space Λ. The transition rate from state $\mathbf{x}_m$ to state $\mathbf{x}_n$, denoted as $q(\mathbf{x}_m, \mathbf{x}_n)$, are given by

$$q(\mathbf{x}_m, \mathbf{x}_n) = \begin{cases} \lambda_i, & \text{if } \mathbf{x}_n = \mathbf{x}_m + e_i \text{ and } \mathbf{x}_n \in \Lambda \\ \mu_i x_i, & \text{if } \mathbf{x}_n = \mathbf{x}_m - e_i, \mathbf{x}_m \in \Lambda \text{ and } x_i \\ & \text{is the } i\text{-th component of } \mathbf{x}_m. \\ 0 & \text{otherwise} \end{cases}$$
$$\tag{3}$$

The equilibrium state distribution $p(\mathbf{x})$ should satisfy the equilibrium equations:

$$p(\mathbf{x}_n) \sum_{\mathbf{x}_m \in \Lambda} q(\mathbf{x}_n, \mathbf{x}_m) = \sum_{\mathbf{x}_m \in \Lambda} p(\mathbf{x}_m) q(\mathbf{x}_m, \mathbf{x}_n), \quad \mathbf{x}_n \in \Lambda$$
$$\tag{4}$$

Then we proceed to find the equilibrium state probability $p(\mathbf{x})$. Due to reversibility([Wan92]), the local balance equations lead to

$$p(\mathbf{x}) \lambda_i = p(\mathbf{x} + e_i) \mu_i (x_i + 1), \quad \text{for } \mathbf{x}, \mathbf{x} + e_i \in \Lambda.$$
$$\tag{5}$$

Using equation (5) repeatedly, one can build up a product form solution of the state probability as follows:

$$p(\mathbf{x}) = p(0) \prod_{i=1}^{N} \frac{\rho_i^{x_i}}{x_i!}, \quad \text{for } \mathbf{x} \in \Lambda \tag{6}$$

$$\text{where } p(0) = \left[\sum_{\mathbf{x} \in \Lambda} \prod_{i=1}^{N} \frac{\rho_i^{x_i}}{x_i!} \right]^{-1}. \tag{7}$$

3.3 System Performance

In this subsection, three measures of the system performance are used: The call blocking probability for each stream, the average link utilization, which represents the efficiency of network resources, and the normalized bandwidth update frequency, which indicates the network management cost on adjusting bandwidth for traffic streams.

We define two sets of states for stream i: $\mathbf{K}_i$ and $\mathbf{A}_i$.

$$\mathbf{K}_i \equiv \{\mathbf{x} | \mathbf{x} \in \Lambda \text{ and } \mathbf{x} + e_i \notin \Lambda\} \tag{8}$$

$$\mathbf{A}_i \equiv \{\mathbf{x} | \mathbf{x} \in \Lambda - \mathbf{K}_i \text{ and } x_i/s_i \text{ is an integer}\} \tag{9}$$

The definition of $\mathbf{K}_i$ implies that the incoming calls of stream i are lost when the state $\mathbf{x}$ is in $\mathbf{K}_i$. The definition of $\mathbf{A}_i$ implies that in these states, the incoming calls of stream i are accepted while the bandwidth should be reallocated for stream i.

For stream i, the call blocking probability, denoted as B_i, is given by

$$B_i = \sum_{\mathbf{x} \in \mathbf{K}_i} p(\mathbf{x}). \tag{10}$$

The expectation of x_i, $E(x_i)$ ($i = 1, 2, \cdots, N$), is derived as follows.

$$E(x_i) = \sum_{\mathbf{x} \in \Lambda} \rho_i \, p(0) \frac{\rho_i^{(x_i-1)}}{(x_i - 1)!} \prod_{j=1, j \neq i}^{N} \frac{\rho_j^{x_j}}{x_j!} \tag{11}$$

Owing to the coordinate convexity, equation (11) can be written as

$$E(x_i) = \rho_i \sum_{\mathbf{x} \in \Lambda - \mathbf{K}_i} p(\mathbf{x}). \tag{12}$$

Hence $E(x_i)$ can be related to B_i by the following equation.

$$E(x_i) = \rho_i \left(1 - B_i\right) \tag{13}$$

On link r, the average bandwidth occupation is given by $\sum_{i=1}^{N} \delta_{ir} \, b_{t_i} \, E(x_i)$. Summing over all links and dividing it by total link capacity, one obtains the average link utilization of the whole system, denoted as U_l.

$$U_l = \frac{\sum_{i=1}^{N} (\sum_{r=1}^{R} \delta_{ir} b_{t_i}) E(x_i)}{\sum_{r=1}^{R} C_r} \tag{14}$$

Note that the numerator in equation (14) can be regarded as the summation of $E(x_i)$ weighted by a factor $\sum_{r=1}^{R} \delta_{ir} b_{t_i}$.

The probability that an incoming stream-i call necessitates bandwidth reallocation, denoted as F_i, is given by

$$F_i = \sum_{\mathbf{x} \in \mathbf{A}_i} p(\mathbf{x}). \tag{15}$$

The overall bandwidth reallocation frequency is given by $\sum_{i=1}^{N} F_i \lambda_i$. Normalizing it by the total call arrival rate, one obtains the normalized bandwidth update frequency, denoted as $\mathbf{F}$.

$$\mathbf{F} = \frac{\sum_{i=1}^{N} F_i \lambda_i}{\sum_{i=1}^{N} \lambda_i} \tag{16}$$

4 Numerical Results And Optimization Rules

In order to investigate the effect of the dynamic bandwidth allocation scheme, we first study a single link system in the following scenario.

The transmission link has a capacity of 150 Mbps, corresponding to the transmission capability of a SONET STS-3c line. The link accommodates two traffic streams. The first stream consists of CCITT E-1 circuit-emulated VCs which has a capacity of 2.048 Mbps. The second stream is composed of CCITT E-3 circuit-emulated VCs with capacity equal to 34.368 Mbps. Since a E-3 circuit occupies more than one fifth of the link capacity, the value of s_2 is always set to 1. Conversely, the bandwidth of E-1 is quite small, therefore the value of s_1 can be set to be in the range from 1 to 17. The maximum value 17 is nearly the ratio of E-3 bandwidth to E-1 bandwidth. Our consideration is that further increasing of s_1 has similar effect as simply using E-3.

The offered traffic load for stream i, denoted as a_i, is defined to be the traffic intensity ρ_i multiplied by basic bandwidth requirement b_i. All offered traffic loads are given in Mbps. The total offered traffic load of the 2 streams is 200 Mbps. The proportion of stream-2 load is set to be 10%, 50%, and 90%. The utilization and bandwidth update frequency are shown in Figure 1(a) and (b), respectively. As the proportion of offered traffic load of E-3 increases from 10% to 90%, the resultant utilization U_l decreases and $\mathbf{F}$ increases. As the E-3 traffic dominates, the variance of U_l to s_1 is small. It is found that the utilization does not decrease monotonically with respect to s_1. The difference in utilization under various s_1 could be as large as 0.15, and the difference in normalized bandwidth update frequency could be up to 0.8. That is, the performance could vary significantly with different values of s_i. Therefore the selection of s_i is of critical importance. In spite of the large variance of utilization, the peak values are pretty close, with the difference of at most 5%. In addition, the normalized bandwidth update frequency is nearly 1.0 for small values of s_1. However, as s_1 increases to 6, the bandwidth update frequency soon settles down to a small value. As a result, it is possible to choose an appropriate value of s_1 such that the bandwidth update frequency is dramatically reduced, while the utilization does not degrade significantly.

Optimization Rules

Based on the above observations, an intuitive way to choose s_i is proposed. The basic idea is to find the maximum U_l within the region where the value of F is acceptable.

Optimization Problem 1

$$\max_{s_i} \quad U_l \tag{17}$$

subject to $\quad F \leq F^*$

where F^* is a given constant, $0 < F^* \leq 1.0$

The solution of this optimization problem, if any, guarantees an acceptable frequency of bandwidth update. However, under some traffic conditions, there could be no feasible solutions. The search of s_i according to (17) is called optimization rule 1. Alternatively, relative constraints could be set instead of absolute constraints. One can set constraint on U_l and minimize F. As depicted before, the utilization reaches maximum at $s_i = 1$. Hence one can constrain U_l to be slightly lower than the maximum value. The second type of optimization problem can now be described as follows.

Optimization Problem 2

$$\min_{s_i} \quad F \tag{18}$$

subject to $\quad U_l \geq K^* \times U_{l(Max)}$

where $U_{l(Max)}$ is U_l obtained under $s_i = 1$, and K^* is a given constant, $0 < K^* \leq 1.0$

The search of s_i according to (18) is referred to as optimization rule 2. Since it is impossible to predict the behavior of U_l with respect to s_i, the only way to obtain the optimal solution of either problem is exhaustively search the entire set of s_i's.

In the following the results obtained with the two optimization rules are presented. Let there be two streams, where stream 1 is E-1 type and stream 2 is E-3 type. The offered traffic load of stream 2 is 10 Mbps and the offered traffic load of stream 1 is varied from 50 Mbps to 140 Mbps. Shown in Figure 2 are the utilization and normalized bandwidth frequency versus E-1 load. Both optimization rules are used. The constant F^* for rule 1 is set to be 0.4, while the constant K^* for rule 2 is 0.98. The curves of U_l and F for $s_i = 1, (i = 1, 2)$ are also presented. As depicted previously, $s_i = 1$ leads to largest values of both U_l and F. The maximum degradation of U_l under the two rules, however, does not exceed 2%. Note that U_l under optimization rule 1 is higher than that under rule 2 with $F^* = 0.4$ and $K^* = 0.98$. On the other hand, F with $s_i = 1$ is nearly 1.0, while the values of F under rule 1 and rule 2 are about 0.4 and 0.1, respectively. This implies that one can reduce F significantly while at the same time retains

high utilization. Using optimization rule 1, the reduction of F is about 0.6, and the loss of U_l is around 0.5%. Using optimization rule 2, the reduction of F is about 0.9, and the loss of U_l is around 2%. One can select one of the two rules according to his need.

In order to illustrate the validity of the multiple-link system model, we study another example with two parallel links as shown in Figure 3. Stream 3 and stream 4 are E-3 traffic streams and traverse link 1 and 2, respectively. Stream 0 is composed of E-1 calls and can traverse either link. The routing ficility decides the route for each incoming stream-0 call according to a random number, with probability q to link 1. Then stream 0 is implicitly separated into two independent streams, stream 1 and 2. Therefore the average link utilization U_l and the normalized bandwidth update frequency F of this system can be obtained using the model and analytical results presented in Section 3. It is found that one can maximize the average link utilization by using a line search algorithm to find the optimal routing probability q^*.

The optimized results of this two-parallel-link system are presented in Figure 4. Suppose both links have capacity 150 Mbps. The offered traffic loads are set to be $a_0 = 200$ Mbps, $a_3 = 10$ Mbps, and $a_4 = 40$ Mbps. First, the line search algorithm is used to find q^* under all possible values of bandwidth allocation units s_i's. Then the optimization rule can be employed to find optimal bandwidth allocation units. In this figure, only the results with $s_i = 1$ $(i = 1, \cdots, 4)$, and those obtained using optimization rule 2 with constant K^* are shown. Again, the loss in utilization U_l is quite small, while the reduction in normalized bandwidth update frequency F is significant.

5 Conclusions

In this paper, a dynamic bandwidth allocation scheme is studied. Under the scheme, the bandwidth allocation size is set to be an integer multiple of the basic bandwidth requirement. Hence the bandwidth control and the connection acceptance control do not need to be done on a call-by-call basis. A control procedure combining the dynamic allocation scheme and the connection acceptance control is also discussed. Besides, the analytical and numerical results are presented. In the numerical examples, it is found that the dependence of the system performances on the values of bandwidth allocation units can not be predicted intuitively. Therefore it is necessary to search the entire set of bandwidth allocation units to find optimal values. Two optimization rules are proposed: one maximizes the utilization subject to acceptable values of the normalized bandwidth update frequency, and the other minimizes the normalized bandwidth update frequency subject to acceptable degradation in utilization. One can

employ any one of the rules and set the constraint parameters to meet his purpose. As a result, the dynamic allocation scheme could be highly efficient in resources usage, and at the same time reduce the bandwidth management cost significantly.

References

[Aei78] J. M. Aein, "A Multi-user-class, Blocked-calls-cleared, Demand Access Model," IEEE Trans. Commun., Vol. COM-26, pp.378-385, Mar. 1978.

[CCI89] CCITT Recommendation I.121, Broadband Aspects of ISDN, Blue Book, Geneva, Switzerland, 1989.

[GAN91] R. Guérin, H. Ahmadi, and M. Naghshieh, "Equivalent Capacity and Its Application to Bandwidth Allocation in High Speed Networks," IEEE J. Select. Areas Commun., Vol.9, No.7, pp.968-981, Sept. 1991.

[Gri90] T. R. Griffiths, "Analysis of a Connection Acceptance Strategy for Asynchronous Transfer Mode Networks," Proc. GLOBECOM, pp.505.4.1-505.4.7, 1990.

[Hui88] J. Y. Hui, "Resource Allocation for Broadband Networks," IEEE Trans. Commun., Vol.6, No.9, pp.1598-1608, Dec. 1988.

[JoV91] S. Jordan and P. Varaiya, "Throughput in Multiple Service, Multiple Resource Communication Networks," IEEE Trans. Commun., Vol.39, No.8, pp.1216-1222, Aug. 1991.

[KHB91] H. Kröner, G. Hébuterne, P. Boyer, and A. Gravey, "Priority Management in ATM Switching Nodes," IEEE J. Select. Areas Commun., Vol.9, No.3, pp.418-427, Apr. 1991.

[MSS91] T. Murase, H. Suzuki, S. Sato and T. Takeuchi, "A Call Admission Control Scheme for ATM Networks Using a Simple Quality Estimate," IEEE J. Select. Areas Commun., Vol.9, No.9, pp.1461-1470, Dec. 1991.

[MST91] T. Murase, H. Suzuki, and T. Takeuchi, "A Call Admission Control for ATM Networks Based on Individual Multiplexed Traffic Characteristics," Proc. ICC, pp.6.3.1-6.3.6, 1991.

[OhS92] S. Ohta and K.-I. Sato, "Dynamic Bandwidth Control of the Virtual Path in an Asynchronous Transfer Mode Network," IEEE Trans. Commun., Vol.40, No.7, pp.1239-1247, July 1992.

[OST88] S. Ota, K.-I. Sato, I. Tokizawa, "A Dynamically Controllable ATM Transport Network Based on the Virtual Path Concept," Proc. GLOBECOM, pp.39.2.1-39.2.5, 1988.

[SOT90] K.I. Sato, S. Ohta and I. Tokizawa, "Broad-band ATM Network Architecture Based on Virtual Paths," IEEE Trans. Commun. Vol.38, No. 8, pp. 1212-1222, Aug. 1990.

[T1S90] T1S1.5/90-001 R1, Broadband Aspects of ISDN, Baseline Document, Draft, August 1990.

[Wan92] W. D. Wang, Design and Analysis of The Labeled-Multiplexer in ATM Networks, Master Dissertation, National Taiwan Univ., June 1992.
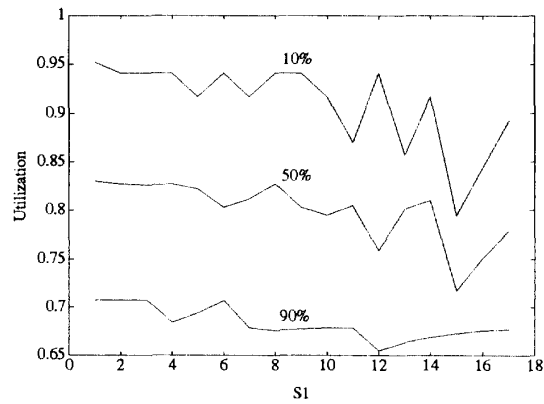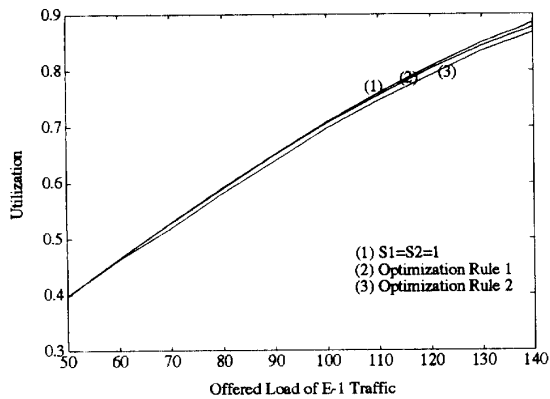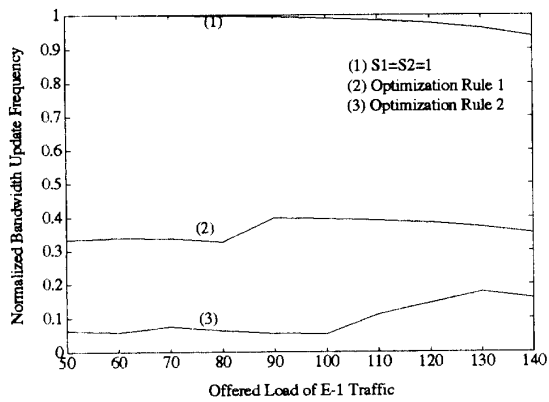
(a) Link Utilization U_l
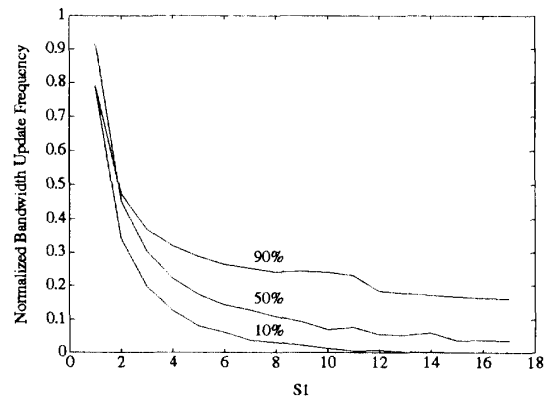
(b) Normalized Bandwidth Update Frequency F

Figure 1: Performance versus Bandwidth Allocation Unit s_1. $a_1 + a_2 = 200$ Mbps, $\frac{a_2}{a_1+a_2} = 10\%, 50\%,$ and 90% .

(a) Link Utilization U_l

(b) Normalized Bandwidth Update Frequency F

Figure 2: Performance versus Offered Load of Stream 1, Under Both Optimization Rules. $a_1 = 50 \sim 140$ Mbps, $a_2 = 10$ Mbps, $F^* = 0.4$, $K^* = 0.98$.
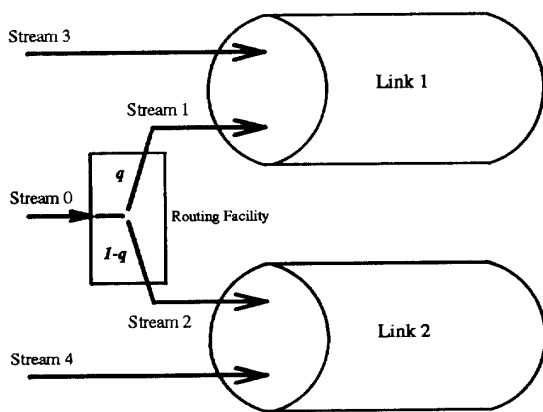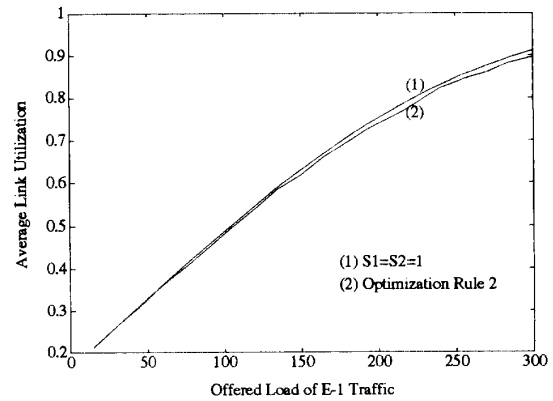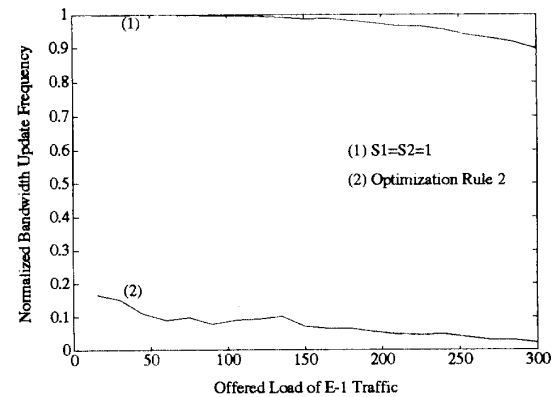

Figure 3: Two-parallel-link System with Probabilistic Routing

(a) Average Link Utilization U_l

(b) Normalized Bandwidth Update Frequency F

Figure 4: Performance versus Offered Load of E-1 Traffics, with Optimal Routing Probability, under Optimization Rule 2. $a_0 = 15 \sim 300$ Mbps, $a_3 = 10$ Mbps, $a_4 = 40$ Mbps, $K^* = 0.98$.