

Bandwidth Allocation of Multiple QOS Classes in ATM Environment¹

Jimmy H. S. Chan and Danny H. K. Tsang

Department of Electrical and Electronic Engineering
Hong Kong University of Science & Technology
Clear Water Bay, Hong Kong

Abstract

For future Broadband-ISDN, Asynchronous Transfer Mode (ATM) is designed not only to support a wide range of traffic classes with diverse flow characteristics (e.g., burstiness, bit rate and burst length), but to guarantee the different quality of service (QOS) requirements as well. The QOS may be measured in terms of cell loss probability and maximum cell delay. In this paper, we consider the ATM networks in which the virtual path (VP) concept is implemented. By applying the Markov Modulated Deterministic Process method, we develop an efficient algorithm to compute the minimum capacity required to satisfy all the QOS requirements when multiple classes of on-off sources are multiplexed onto a single VP. Using the result, we then propose a simple algorithm to determine the VP combination to achieve the near optimum of total capacity required for satisfying the individual QOS requirements. Numerical results are also presented to demonstrate the performance of the algorithm when compared to the optimal total capacity required.

I. Introduction

ATM-based Broadband ISDN (B-ISDN) aims to support a wide variety of services (such as voice, video, and data) that exhibit diverse traffic flow characteristics, while guaranteeing the quality of service (QOS) requirements of individual users. In ATM networks, different traffic classes with different flow characteristics and QOS requirements are statistically multiplexed to increase the utilization of network bandwidth. Due to the high-speed nature of the ATM environment, the bandwidth control algorithm employed must be both simple and efficient. Simplicity would permit practical implementation of the algorithm while efficiency would minimize the bandwidth required to guarantee the different QOS requirements. QOS is normally expressed in terms of cell loss probability and maximum cell delay, in which the maximum cell delay can easily be imposed by having a finite buffer to limit the maximum queue length and consequently the maximum delay. Henceforth, we shall simply use QOS to refer to cell loss probability.

The concept of virtual path (VP) [5],[6],[9],[11],[12] has been proposed to reduce the processing and complexity of network control and has been adopted by CCITT as part of the ATM standard. A virtual path is basically a logical direct connection between two nodes. It provides a multiplexing structure in which virtual circuits are statistically multiplexed onto a virtual path, and virtual paths are then multiplexed onto a physical transmission link. The capacity allocated to a VP should be large enough to guarantee the various QOS requirements of virtual circuits supported by the VP. On the other hand, the total VP capacity allocated should not exceed the capacity of the physical link.

For homogeneous traffic flow, two possible and simple approaches to virtual path traffic management can be taken. One is to provide a single VP on which all the traffic classes are integrated and the QOS of this VP should be set for the most demanding virtual circuits (i.e., the most stringent QOS). The other is to provide multiple and different QOS classes of VP, and all traffic classes with the same QOS are integrated for statistically multiplexing on a single VP. Therefore, the different QOS classes of VP are segregated and managed individually. The graphical illustration of these two approaches is shown in Fig. 1 as in [3], [4], [7], [8]. In the first approach, although the different traffic classes can easily be managed by using a single VP, the network bandwidth may not be efficiently utilized due to the most stringent QOS selected as the VP's QOS while other traffic classes with less stringent QOS requirements are also multiplexed on the VP. In the second approach, besides the need of a more complicated control to manage the different VPs, bandwidth may also be wasted since dedicated capacity is allocated to each VP to guarantee its QOS and no sharing of bandwidth can take place across the VPs. Therefore, it is necessary to design a simple control strategy to optimize the required bandwidth for satisfying the QOS requirements of various traffic classes.

For the case of heterogeneous traffic flow, the problem is further complicated by the fact that when different traffic classes are statistically multiplexed on a VP, the QOS experienced by each traffic class differs due to the difference in flow characteristics. The VP's QOS (or the average QOS supported by the VP) cannot simply be the most stringent QOS required. The bandwidth allocated to the VP must be sufficient to guarantee all the individual QOS and not just the VP's QOS.

The contribution of this paper is two-fold. First, by applying the Markov Modulated Deterministic Process (MMDP)

1. This research was supported in part by HKUST UPGC Research Infrastructure Grant: RI92/93.EG02 and in part by Hong Kong Telecom Institute of Information Technology grant: 93/94.EG01

recently developed in [1], we develop a simple and accurate algorithm to compute the minimum and sufficient bandwidth to satisfy the QOS requirements of various traffic classes multiplexed together on a single VP. The traffic classes are considered to be on-off sources characterized by the statistical parameters such as peak rate, mean rate and mean burst length. Second, a simple algorithm to determine the near optimal VP combination is presented and the performance of the proposed algorithm is compared to the optimal capacity required.

The paper is organized as follows. Section II presents an iterative algorithm to determine the minimum and sufficient capacity of a VP for satisfying the individual QOS requirements. In Section III we apply the algorithm to investigate the performance of two simple VP combinations (integration and segregation approaches). Section IV presents a simple algorithm to determine the near optimal VP combination for maximum bandwidth efficiency. Section V presents the numerical results and compare the performance of the proposed algorithm to that of the optimal VP combination. Section VI summarizes the entire paper.

II. Algorithm to determine the minimum capacity required by a VP

With the formula to compute the individual cell loss probabilities given in the Appendix, we now propose an iterative algorithm to determine the sufficient and minimum capacity of a VP so that all the individual QOS are satisfied when different traffic classes are multiplexed together on the VP. Suppose there are N classes of traffic. For each traffic class i , $i=1, \dots, N$, traffic parameters including the peak rate Δ_i , the mean rate Φ_i , the mean burst length L_i , the number of independent sources M_i , and the required cell loss probability Q_i , are known to the ATM multiplexer. Using these information, the ATM multiplexer can determine the minimum required capacity C so that all the individual Q_i are guaranteed. An iterative algorithm is given below to determine C for a given buffer size K .

Iterative Algorithm 1

1. Find the lower bound of the required capacity $C_{lowerbound}$ and the upper bound of the required capacity $C_{upperbound}$

$$C_{lowerbound} = \sum_{i=1}^N M_i \Phi_i$$

$$C_{upperbound} = \sum_{i=1}^N M_i \Delta_i$$

Let $a = C_{lowerbound}$, $d = C_{upperbound}$, and compute $\{p_x\}$ using Equations (12) and (13).

2. Initialize $j \leftarrow 0$, $\sigma_i^{(0)} \leftarrow 0$, $i=1, \dots, N$; $C = a + (d-a)/2$ and initialize $y^{(0)}$.
3. For $x \in \Lambda$, using Equation (14) and the local normal-

ization, compute

$$y_x^{(j+1)} \leftarrow \left(\sum_{i=1}^N \left(\alpha_{x_i^-}^{(i)} y_{x_i^-}^{(j+1)} + \beta_{x_i^+}^{(i)} y_{x_i^+}^{(j)} \right) \right) \cdot A_x$$

$$y_x^{(j+1)} \leftarrow \frac{p_x \cdot y_x^{(j+1)}}{y_x^{(j+1)} \cdot e^T}$$

4. Compute the individual cell loss probabilities $\sigma_i^{(j+1)}$, $i = 1, \dots, N$, using Equation (11).
5. If any of $\sigma_i^{(j+1)}$ has not converged by checking
$$\frac{|\sigma_i^{(j+1)} - \sigma_i^{(j)}|}{\sigma_i^{(j+1)}} \leq \epsilon,$$
then $j \leftarrow j + 1$ and go to step 3.
6. Let $\sigma_i = \sigma_i^{(j+1)}$ and test to see if $\sigma_i \leq Q_i \forall i$,
 - (i) if $\sigma_i > Q_i$ for some i , $i=1, \dots, N$, then $a = C$ & go to step 2
 - (ii) if $\sigma_i < Q_i \forall i$, then $d = C$ and go to step 2.
 - (iii) otherwise terminate the process and C is found.

In the above algorithm, step 3 is the usual Gauss-Seidel iteration which involves most of the computation time. The second equation in step 3 performs the local normalization to ensure the sum of all elements in vector y_x equal to its exact value p_x after each iteration. Step 6 is a bisection method to search for the minimum capacity for satisfying all the QOS requirements.

III. Effect of traffic parameters and VP combination on the capacity required

In this section we first investigate the influence of each traffic parameter on the capacity required by a VP supporting a single traffic class. We then look at the required total capacity of two simple VP combinations (here a VP combination simply refers to an assignment of traffic classes to VPs) as the QOS varies for multiple traffic classes, in which we consider two separate cases: homogeneous traffic flow and heterogeneous traffic flow. From the results obtained, we can then identify the most important traffic parameter, in addition to the QOS requirement, that affects the capacity required the most. The results also provide some insight on how VPs should be assigned to traffic classes so as to achieve minimum or near minimum of the total capacity required. Throughout the rest of the paper, we shall assume a fixed buffer size of $K = 50$ for each individual VP.

In order to investigate which parameter has the most significant influence on the capacity required, we use a base line case with parameters ($\Delta=5$, $b=10$, $L=200$) and compare this with other cases in which only one parameter differs while the other two remain the same. Here the number of sources M is set to be 80. In Fig. 2 we plot the capacity required as the value of QOS varies for the four cases. The solid line is the base line case while each of the other three lines corresponds to the case of changing only one of the parameters. Let the QOS (cell loss probability) = 10^{-m} . These lines plot the capacity required to satisfy the QOS

requirements when m varies from 2 to 10.

Fig. 2 shows that except for the top curve which has a larger gradient, all three other curves have approximately the same gradient. The top curve corresponds to the case of larger peak rate when compared to the base line case. The results indicate that as QOS becomes more stringent, the incremental capacity required to satisfy the QOS requirement is higher for the case of larger peak rate. This also implies that peak rate is perhaps the most important traffic parameter in determining the amount of capacity required. Although burstiness affects the capacity required somewhat (as seen from the bottom curve in Fig. 2), its incremental capacity does not differ from that of the base line case by much. On the other hand, burst length seems not to have much effect on the capacity required at all as seen from the two almost overlapping curves in Fig. 2.

Now we study the performance (i.e., the total capacity required) of two simple VP combinations for supporting multiple traffic classes with different QOS requirements. The two VP combinations to be examined are:

- i) Integration approach
Between a pair of origin and destination, all VCs from various traffic classes are multiplexed onto one VP. The capacity allocated to the VP should be large enough to satisfy all the individual QOS.
- ii) Segregation approach
Between a pair of origin and destination, multiple VPs are provided and each VP supports VCs having identical flow characteristics and the same QOS requirements. The capacity allocated to each VP should be sufficient to guarantee the required QOS of the traffic classes supported. The sum of all the VP capacities assigned would constitute the total capacity required.

Using the algorithm given in the previous section, we compare the performance of the two approaches for both homogeneous and heterogeneous traffic flows. In this paper, we consider three types of on-off sources: data, image, and voice. The flow characteristics of these sources are from [10] and are shown in Table 1. We assume that the cell length is 53 bytes.

Table 1: Traffic characteristics of some on-off sources

Traffic Classes	Peak Rate, Δ_i (Mbps)	Burstiness, b_i	Burst length, L_i (cells)
Data	10	10	339
Still Image	2	23	2604
Voice	0.064	3	58

A) Homogeneous traffic flow

Here we consider two traffic classes with identical flow characteristics but different QOS requirements. The traffic flow is assumed to be of the Data type. Each class has the same number of sources (i.e., $M_1 = M_2 = 20$). Let the cell loss probabilities (QOS_1 and QOS_2) be 10^{-m} and 10^{-n} , respectively. In order to compare the performance of the two approaches for different QOS values, we plot in Fig. 3 the total capacity required as n varies for different fixed values of m . The solid lines are the integration approach while the dotted lines are the segregation approach. The horizontal solid lines result from the fact that in the integration approach the required VP capacity is solely determined by the most stringent QOS.

In Fig. 3, the results show that the integration approach requires less capacity than the segregation approach for almost all cases, except for the case of $m = 2$ and $n > 6$ (i.e., $QOS_1 = 10^{-2}$ and $QOS_2 < 10^{-6}$). For the case of $m = 3$, by interpolation we can see that the segregation approach also becomes better when $n > 10$. This suggests that the difference between m and n (i.e., the magnitude difference) may not be a good indicator to decide which approach to use. This also implies that the knowledge of the QOS alone may not be sufficient to determine the optimal VP combination even in the case of homogeneous traffic flow.

B) Heterogeneous traffic flow

Here we consider two traffic classes with different flow characteristics and different QOS requirements. Class 1 traffic flow is assumed to be of the Data type while class 2 traffic flow to be of the Still Image type. Each class has the same number of sources (i.e., $M_1 = M_2 = 20$). Again let the cell loss probabilities (QOS_1 and QOS_2) be 10^{-m} and 10^{-n} , respectively. Similar to Fig. 3, we plot in Fig. 4 the total capacity required for the two approaches as n varies for different fixed values of m . Since the flow characteristics of the two traffic classes are different, we also plot in Fig. 5 for varying m and fixed n . Note that for both approaches the capacity determined must satisfy both individual QOS requirements.

In Fig. 4, the 45-degree solid line seems to divide the dotted lines into two equal regions. The capacity required by the segregation approach is always less than that of the integration approach whenever $n > m$ (or $QOS_2 < QOS_1$) as indicated by the lower region below the 45-degree solid line. In other words, the integration approach becomes better when $QOS_2 \geq QOS_1$. Similar observations are also found in Fig. 5.

Since the peak rate of class 1 traffic (10 Mbps) is higher than that of class 2 traffic (2 Mbps), the above results imply that when high peak rate traffic requires less stringent QOS, the segregation approach gives a better performance. On the other hand, when high peak rate traffic demands the most stringent QOS, the integration approach would be a better choice in terms of less total capacity required. With this information available, we now propose a simple algorithm to determine the near optimal VP combination in the next section.

IV. Algorithm to determine the near optimal VP combination

Consider N on-off traffic classes joining an ATM multiplexer. Each source of traffic class i , $i = 1, 2, \dots, N$, generates cells at a peak rate of Δ_i when it is in the on state and demands a QOS requirement of Q_i . A number of VP combinations, each of which requires different total capacity, can be adopted to handle the traffic. Hence, it is desirable to have simple rules suitable for implementation at ATM multiplexers to determine the optimal VP combination that minimizes the total capacity required. We now propose a simple algorithm that can produce a VP combination of reasonable performance. The algorithm is of an iterative nature and is given below:

Iterative Algorithm 2

- i) Let
$$I = \arg \min \{Q_i\}$$
$$1 \leq i \leq N$$
$$J = \arg \max \{\Delta_i\}$$
- ii) If $I = J$, it implies that among all traffic classes, traffic class J has the highest peak rate and also demands the most stringent QOS. Thus all traffic classes should be integrated into one virtual path. The minimum capacity of this VP can be determined by using Algorithm 1 in the previous section so as to satisfy $\sigma_i \leq Q_i \forall i$.

or

If $I \neq J$, then some sort of the segregation is necessary. Initially, the segregation is to separate the traffic classes into two groups, each of which is supported by at least one VP. The first group consists of all traffic classes with $Q_i \leq Q_J$, while the second group contains the rest of the traffic classes.

- iii) After steps i) and ii), one needs to check each group to see if further segregation is necessary by going through steps i) and ii) again for each group, but the number of classes in each group becomes smaller. The recursion would stop when no further segregation is required. The VP capacity of each group can be determined by using Algorithm 1. The total capacity required is just the sum of all VP capacities.

The algorithm is very easy to implement and requires only simple computations. Thus it is suitable for implementation in the ATM multiplexers to determine how traffic classes should be grouped into different VPs so as to minimize the total capacity required. Next we are interested in knowing how well the proposed algorithm performs when compared to the optimal solution. The next section will present some numerical results to compare the performance of the proposed algorithm to that of the optimal VP combination.

V. Performance evaluation of the proposed algorithm

The key to the effectiveness of the proposed algorithm lies in how well the suggested VP combination achieves the optimal total capacity required. It is thus worth studying the performance of the proposed algorithm under various traffic conditions. Here we consider three traffic classes (Data, Still Image, and Voice) multiplexed into an ATM multiplexer. Basically, there are five possible VP combinations: complete integration (i.e., 1 VP to support all three classes); complete segregation (i.e., 3 VPs; one VP supports only one class); and three partial segregations (i.e., 2 VPs; one VP supports only one class and the other VP supports two classes). Of these five VP combinations, one will give the minimum total capacity required and this corresponds to the optimal VP combination. However, the optimal VP combination may depend on the QOS requirement being specified. Therefore, for each given set of QOS requirement, we would need to evaluate the total capacities required for all five VP combinations in order to locate the optimal VP combination.

Let the QOS requirements be represented by $(Q_{\text{data}}, Q_{\text{image}}, Q_{\text{voice}})$. In order to test the different QOS combinations in the evaluation, we allow each QOS component to vary independently between 10^{-2} to 10^{-9} . Thus altogether there are 512 different QOS combinations for testing.

In Fig. 6, the solid line represents the optimal capacity required by the optimal VP combination for different QOS combinations while the dotted line is the total capacity required by the VP combination obtained from the proposed algorithm. In most cases, the proposed algorithm yields the optimal VP combination, which is illustrated by the overlapping of the two curves. However, for some cases the dotted curve is above the solid one and it means that the proposed algorithm fails to produce the optimum. Nevertheless, the error is found to be at most 12% as seen in Fig. 7.

The error may be attributable to the similar peak rates of the traffic classes used in the numerical testing. From Table 1, we see that the peak rates of Data and Still Image are very close (10 Mbps and 2 Mbps). This observation leads us to test the performance of the proposed algorithm for traffic classes with peak rates which are at least one order of magnitude different. Suppose we change only the peak rate of Still Image to 0.5 Mbps and run the same tests again. The results are shown in Figs. 8 and 9. In Fig. 8, again the solid line is for the optimal VP combination while the dotted line for the VP combination from the proposed algorithm. The two curves overlap in almost all cases. Furthermore, the error incurred by the algorithm now becomes at most 2% as seen in Fig. 9. This observation suggests that the proposed algorithm should perform extremely well when the peak rates under consideration differ by at least one order of magnitude.

VI. Conclusion

In this paper, we develop an algorithm to compute the minimum capacity required to satisfy the individual QOS requirements of various traffic classes multiplexed together onto a single VP. Using the developed algorithm, we investigate the traffic parameters to see which parameter affects the capacity required the most. We then study two simple VP combinations (integration and segregation approaches) for supporting two traffic classes with different QOS requirements. From the numerical results obtained, we observe the conditions under which integration approach is preferred. Based on this observation, we develop a simple algorithm that by using both peak rate and QOS information can determine the near optimal VP combination for supporting multiple traffic classes requiring different QOS. The effectiveness of the proposed algorithm has been verified by numerical results under different QOS requirements for multiple traffic classes. In order to improve the accuracy of the algorithm, it is found that the peak rates of various traffic classes must differ by at least one order of magnitude. In summary, the proposed algorithm can provide a simple and effective mean of determining which traffic classes should be multiplexed together onto a virtual path so as to maximize bandwidth efficiency.

Appendix

In this Appendix, we present a method for computing the individual cell loss probabilities of different traffic classes when they are statistically multiplexed into an ATM multiplexer consisting of a finite buffer of size K cells and a transmission capacity of C cells per second. Here we restrict our attention to consider only on-off traffic sources. An on-off source alternates between two states (on and off). When a source is on, it generates cells at peak rate Δ cells per second, and when a source is off, it generates no cells at all. Both the on and off periods are distributed as independent exponential random variables with means $1/\mu$ and $1/\lambda$, respectively. Thus a source can be fully characterized by the parameters Δ , μ and λ . Another more descriptive and more commonly used set of parameters for on-off sources is: mean rate Φ (the average number of cells generated per second), the burst length L (the average number of cells generated during an on period), and the burstiness b (the ratio between the peak rate Δ and the mean rate Φ). These parameters are related by the formulas: $b = 1 + \mu/\lambda$, $\Phi = \Delta/b$, $L = \Delta/\mu$. Denote the total number of traffic classes by N and assume that there are M_i independent sources in class i for $i = 1, 2, \dots, N$ with parameters Δ_i , μ_i and λ_i . Sources within one class are statistically identical and are described by the same set of parameters.

The exact analytical analysis of the above traffic model is very difficult due to the complex characteristics of the superposed arrival process. In order to simplify the computation, a more tractable but statistically equivalent arrival process can be employed to model the superposed arrival process. Methods such as Markov Modulated Poisson Process (MMPP) [13] and stochastic fluid flow (SFF) [14] have commonly been used for this purpose. In this paper, we adopt instead the Markov Modulated Deterministic Process (MMDP) to model the superposed traffic stream since the MMDP method has been shown to be quite simple, computationally efficient, and reasonably accurate

for homogeneous on-off traffic [1]. Hence, the traffic model under consideration becomes an MMDP/D/1/K queuing model. The detailed analysis of this model for homogeneous on-off traffic can be found in [1].

Let $\{X(t), t \geq 0\}$ be a finite, irreducible, continuous-time Markov Process with a state space $S = \{0, 1, 2, \dots, m-1\}$ and an $m \times m$ probability transition matrix P . Consider an arrival process modulated by $X(t)$ in the following way. When $X(t)$ is in state j ($j \in S$), cell arrivals occur at a fixed rate of B_j cells/s (in other words, the inter-cell arrival time is constant and takes on the value of $1/B_j$ seconds). We further assume that the cell arrival process is always renewed immediately after a transition of $X(t)$. Thus if a transition of $X(t)$ into state j ($j \in S$) occurs at time t_0 , then the first cell arrival after this transition will occur at time $t_1 = t_0 + 1/B_j$. Also let $1/\gamma_j$ ($j \in S$) be the mean sojourn time of $X(t)$ in state j . This arrival process shall be referred to as Markov Modulated Deterministic Process, which is completely defined by the vectors $B = (B_0, B_1, \dots, B_{m-1})$ and $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{m-1})$, and the transition matrix P .

For the case of heterogeneous traffic classes, let $X_i(t)$ ($i=1, \dots, N$) be the number of class i sources that are on at time t and $X_i(t)$ is a finite birth-death process. Then $X(t) = \{X_1(t), X_2(t), \dots, X_N(t)\}$ is an N -dimensional birth-death process with state space $\Lambda \equiv \{x = (x_1, x_2, \dots, x_N) \mid 0 \leq x_i \leq M_i, i = 1, 2, \dots, N\}$. Let $X(t)$ be approximated by an MMDP. If $X(t)$ is in state x ($x \in \Lambda$), then the fixed arrival rate is

$$B_x = \sum_{i=1}^N x_i \Delta_i \quad (1)$$

and the average time that $X(t)$ spends in state x becomes $1/\gamma_x$ where

$$\gamma_x = \sum_{i=1}^N [(M_i - x_i) \lambda_i + x_i \mu_i] \quad (2)$$

To simplify the notation, let us define the vectors x_i^+ and x_i^- as

$$\begin{aligned} x_i^+ &= (x_1, x_2, \dots, x_i + 1, \dots, x_N) \\ x_i^- &= (x_1, x_2, \dots, x_i - 1, \dots, x_N) \end{aligned} \quad (3)$$

Let $P_{x,z}$ be an element of the transition matrix P of $X(t)$ and denote the transition probability from state x to state z , where $x, z \in \Lambda$. Since $X(t)$ is an N -dimensional birth-death process, its transition probability $P_{x,z}$ ($x, z \in \Lambda$) can be expressed as

$$P_{x,z} = \begin{cases} \alpha_x^{(i)} & z = x_i^+, \quad i = 1, \dots, N \\ \beta_x^{(i)} & z = x_i^-, \quad i = 1, \dots, N \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where

$$\begin{aligned} \alpha_x^{(i)} &= \begin{cases} \frac{(M_i - x_i) \lambda_i}{\gamma_x} & x \in \Lambda, \quad i = 1, \dots, N \\ 0 & \text{otherwise} \end{cases} \\ \beta_x^{(i)} &= \begin{cases} \frac{x_i \mu_i}{\gamma_x} & x \in \Lambda, \quad i = 1, \dots, N \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Similar to the derivation in [1], let $\xi_0 = 0$ and ξ_η ($\eta = 1, 2, \dots$) be the η th transition epoch of $X(t)$, and $Y(t)$ be the number of cells in the buffer at time t . During the period $[\xi_\eta, \xi_{\eta+1})$, cell arrivals and departures are deterministic. However, the embedded process $[X(\xi_\eta^+), Y(\xi_\eta^+), \eta = 1, 2, \dots]$ is non-Markovian and thus the analysis of this process is very difficult. To overcome this difficulty, we introduce the service renewal assumption (i.e., when a transition occurs in the middle of a cell's transmission, the *whole* cell will be transmitted immediately after ξ_η . With this assumption, the embedded process becomes a Markov Chain. The error introduced by this assumption has been shown to be insignificant in [1]. Now for $x \in \Lambda$, define the vector

$$\pi_x = (\pi_{x,0}, \pi_{x,1}, \dots, \pi_{x,K}) \quad (5)$$

Let the steady-state probability of the embedded Markov Chain be

$$\pi_{x,k} = \lim_{\eta \rightarrow \infty} P \{X(\xi_\eta^+) = x, Y(\xi_\eta^+) = k\} \quad (6)$$

$x \in \Lambda, k = 0, 1, \dots, K$

The following system of linear equations is satisfied by π_x and

$$\pi_x = \sum_{i=1}^N \left(\alpha_{x_i}^{(i)} \pi_{x_i^-} A_{x_i^-} + \beta_{x_i}^{(i)} \pi_{x_i^+} A_{x_i^+} \right), \quad x \in \Lambda \quad (7)$$

$$\sum_{x \in \Lambda} \pi_x e^T = 1 \quad (8)$$

where $e = (1, 1, \dots, 1)$.

The stochastic matrix A_x ($x \in \Lambda$) can be defined according to the relationship between B_x and C .

For $B_x < C$, A_x is approximately a lower triangular matrix and is given by

$$A_x = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho_x & 1 - \rho_x & 0 & \dots & 0 \\ \rho_x^2 & \rho_x(1 - \rho_x) & 1 - \rho_x & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \rho_x^K & \rho_x^{K-1}(1 - \rho_x) & \rho_x^{K-2}(1 - \rho_x) & \dots & 1 - \rho_x \end{bmatrix}$$

where $\rho_x = \exp\{-\gamma_x / (C - B_x)\}$.

For $B_x = C$, A_x can be approximated by an identity matrix.

For $B_x > C$, A_x is approximately an upper triangular matrix and is given by

$$A_x = \begin{bmatrix} 1 - \rho_x & \rho_x(1 - \rho_x) & \rho_x^2(1 - \rho_x) & \dots & \rho_x^K \\ 0 & 1 - \rho_x & \rho_x(1 - \rho_x) & \dots & \rho_x^{K-1} \\ 0 & 0 & 1 - \rho_x & \dots & \rho_x^{K-2} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

where $\rho_x = \exp\{-\gamma_x / (B_x - C)\}$.

To compute the overall cell loss probability σ from π_x ($x \in \Lambda$), let $S_{x,k}$ be the total number of cells that arrive during the

period $[\xi_\eta, \xi_{\eta+1})$ given that $X(\xi_\eta^+) = x$ and $Y(\xi_\eta^+) = k$, where $x \in \Lambda$ and $k = 0, \dots, K$. Also let $R_{x,k}$ be the total number of cells rejected in the same period. Similar to the derivation given in [1], we have

$$E(S_{x,k}) = \frac{B_x}{\gamma_x}$$

$$E(R_{x,k}) = \begin{cases} \frac{\rho_x^{K-k+1}}{1 - \rho_x} & \text{if } B_x > C \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

and the overall cell loss probability σ is given by

$$\sigma = \frac{\sum_{(x,k) \in \Omega} E(R_{x,k}) \pi_{x,k}}{\sum_{(x,k) \in \Omega} E(S_{x,k}) \pi_{x,k}} = \frac{\sum_{x \in V} \frac{\rho_x y_{x,K}}{1 - \rho_x}}{\sum_{x \in \Lambda} \frac{B_x \rho_x}{\gamma_x}} \quad (10)$$

where $\Omega \equiv \{(x,k) \mid x \in \Lambda \text{ and } 0 \leq k \leq K\}$, $V \equiv \{x \in \Lambda \mid B_x > C\}$, $\rho_x \equiv \pi_x e^T$, and $y_{x,K}$ is the last element of vector $y_x \equiv \pi_x A_x$.

To compute the individual cell loss probabilities, let $S_{x,k}^{(i)}$ and $R_{x,k}^{(i)}$ be the number of class i cell arrivals and the number of class i cells lost, respectively, during the period $[\xi_\eta, \xi_{\eta+1})$ given that $X(\xi_\eta^+) = x$ and $Y(\xi_\eta^+) = k$. During the period $[\xi_\eta, \xi_{\eta+1})$, cell arrivals, cell departures, and cell losses are all deterministic. The total cell arrivals from class i is simply a fraction of the total cells arriving in this period, which is equal to $(x_i \Delta_i) / B_x$. Similarly, the total cell loss from class i is also this fraction of the total cell lost in this period. Thus the individual cell loss probability σ_i for class i is given by

$$\sigma_i = \frac{\sum_{(x,k) \in \Omega} E(R_{x,k}^{(i)}) \pi_{x,k}}{\sum_{(x,k) \in \Omega} E(S_{x,k}^{(i)}) \pi_{x,k}} = \frac{\sum_{x \in V} \frac{\rho_x y_{x,K} x_i \Delta_i}{(1 - \rho_x) B_x}}{\sum_{x \in \Lambda} \frac{x_i \Delta_i \rho_x}{\gamma_x}} \quad (11)$$

Note that p_x ($x \in \Lambda$), which is also the steady-state probability distribution of $X(t)$, can easily be determined from q_x ($x \in \Lambda$), the steady-state probability distribution of the embedded Markov chain with transition matrix P as defined above. From [2], the computation of p_x from q_x is given by

$$p_x = \frac{\gamma_x q_x}{\sum_{x' \in \Lambda} \gamma_{x'} q_{x'}}, \quad x \in \Lambda \quad (12)$$

Since $X_1(t)$, $X_2(t)$, ..., and $X_N(t)$ are all independent binomial random variables, we thus have

$$q_x = \prod_{i=1}^N \binom{M_i}{x_i} \left(\frac{\lambda_i}{\lambda_i + \mu_i} \right)^{x_i} \left(\frac{\mu_i}{\lambda_i + \mu_i} \right)^{M_i - x_i}, \quad x \in \Lambda \quad (13)$$

The probability $p_x \equiv \pi_x e^T$ can be computed without the knowledge of π_x . Similarly for computing $y_x \equiv \pi_x A_x$, since A_x is a stochastic matrix (i.e., $A_x e^T = e^T$), the following relations hold:

$$y_x e^T = \pi_x A_x e^T = \pi_x e^T = p_x$$

By multiplying both sides of Equations (7) & (8) by A_x , we obtain the following system of linear equations for y_x .

$$y_x = \left(\sum_{i=1}^N \left(\alpha_{x_i^-}^{(i)} y_{x_i^-} + \beta_{x_i^+}^{(i)} y_{x_i^+} \right) \right) \cdot A_x, \quad x \in \Lambda \quad (14)$$

$$\sum_{x \in \Lambda} y_x e^T = 1 \quad (15)$$

y_x can be solved by any iterative methods such as Gauss-Seidel method. Observe from above that $y_x e^T = p_x$, where p_x can be pre-computed from Equations (12) and (13), local normalization (i.e., $y_x e^T = p_x$) should be performed to speed up the convergence of Equation (14) without resorting to the normalization equation of (15).

References

- [1] T. Yang and D.H.K. Tsang, "A novel approach to estimating cell loss probability in an ATM multiplexer loaded with homogeneous bursty sources", Proc. GLOBECOM 1992, pp. 511-517.
- [2] S.M. Ross, "Stochastic Processes.", John Wiley & Sons, Inc., New York 1983.
- [3] CCITT, "Recommendations of the I.370", Study Group XVIII R17 July 1992.
- [4] K. Yasuhiro and K. Toshikazu, "QOS control in an ATM network considering virtual path concept", Proc. IEEE INFOCOM 1990, pp. 701.4.1-701.4.7.
- [5] K. Sato, S. Ohta and I. Tokizawa, "Broadband ATM network architecture based on virtual path", IEEE Trans. Commun. vol. 38, no. 8, pp. 1212-1222, Aug 1990.
- [6] I. Tokizawa and K.Sato, "Broadband transport techniques based on virtual paths", Proc. GLOBECOM 90, pp. 1269-1273.
- [7] K. Sato and I. Tokizawa, "Flexible asynchronous transfer mode networks utilizing virtual paths", Proc. ICC 1990, Atlanta, GA, pp. 318.4.1-318.4.8, April 16-19 1990.
- [8] K. Sato, H. Hadama, and I. Tokizawa, "Network reliability enhancement with virtual path strategy", Proc. GLOBECOM 1990, pp. 403.5.
- [9] Y. Sato, and K. Sato, "Virtual path and link capacity design for ATM networks", IEEE J. Select. Areas Commun., vol. 9, no. 1, pp. 104- 111 Jan 1991.
- [10] P. Castelli, E. Cavallero, and A. Tonietti, "Policing and call admission problems in ATM networks", ITC-13, pp. 847-852, 1991.
- [11] S. Ohta and K. 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.
- [12] S.Ohta, K.I. Sato, and I. Tokizawa, "A dynamically controllable ATM transport network based on virtual path concept", Proc. IEEE GLOBECOM 1988, pp.1272-1276.
- [13] H. Heffes and D.M. Lucantoni, "A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance", IEEE J. Select. Areas Commun., vol. 4, no. 6, pp. 856-868, Sept 1986.
- [14] D. Anick, D. Mitra, and M.M. Sondhi, "Stochastic theory of

a data-handling system with multiple sources", Bell Sys. Tech. J., vol. 61, no. 8, pp. 1871-1894, Oct 1982.

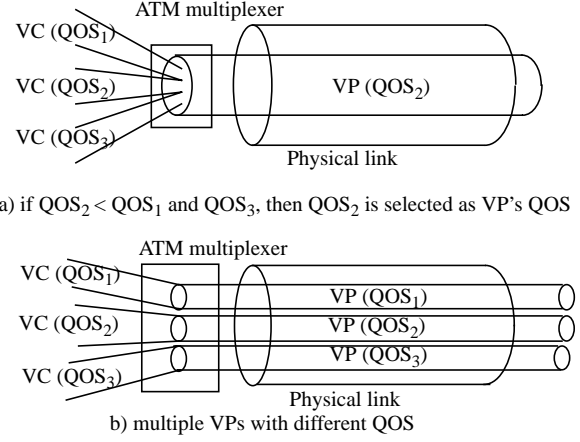


Figure 1. a) Integration approach and b) Segregation approach.

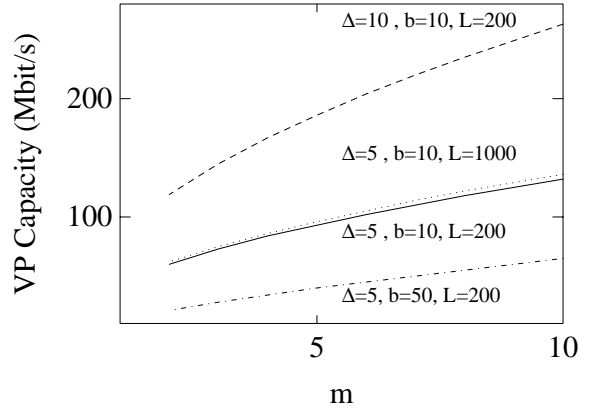


Figure 2. Sensitivity analysis of traffic flow parameters ($QOS = 10^{-m}$)

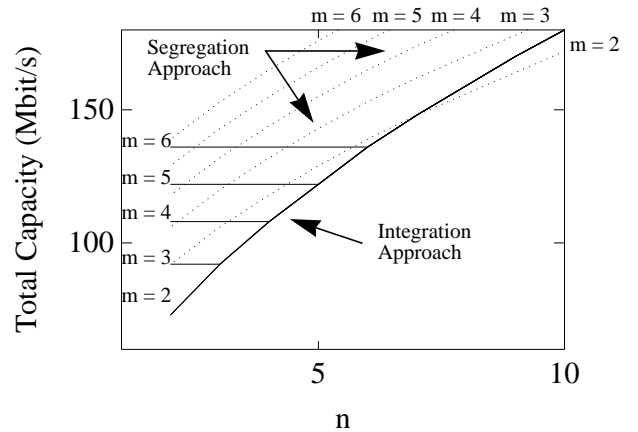


Figure 3. Integration approach vs. segregation approach for homogeneous traffic flow (classes 1 & 2 = Data, $QOS_1 = 10^{-m}$ and $QOS_2 = 10^{-n}$)

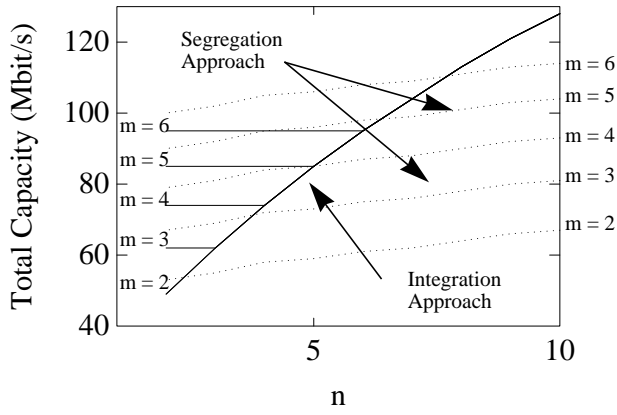


Figure 4. Integration approach vs. segregation approach for heterogeneous traffic flow (class 1 = Data, class 2 = Still Image, $QOS_1 = 10^{-m}$ and $QOS_2 = 10^{-n}$)

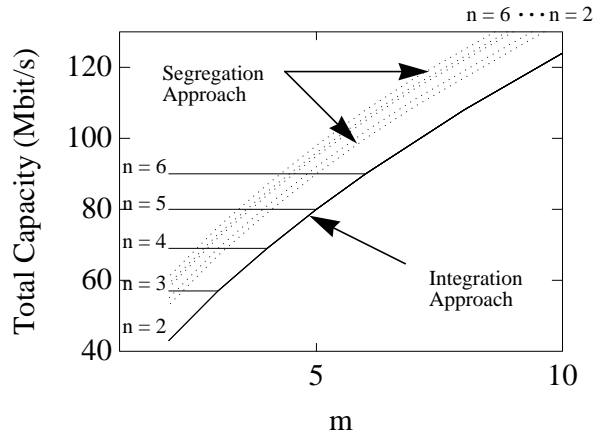


Figure 5. Integration approach vs. segregation approach for heterogeneous traffic flow (class 1 = Data, class 2 = Still Image, $QOS_1 = 10^{-m}$ and $QOS_2 = 10^{-n}$)

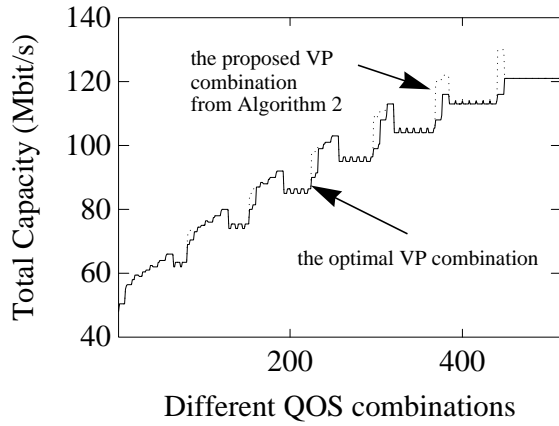


Figure 6. The optimal VP combination vs. the proposed VP combination.

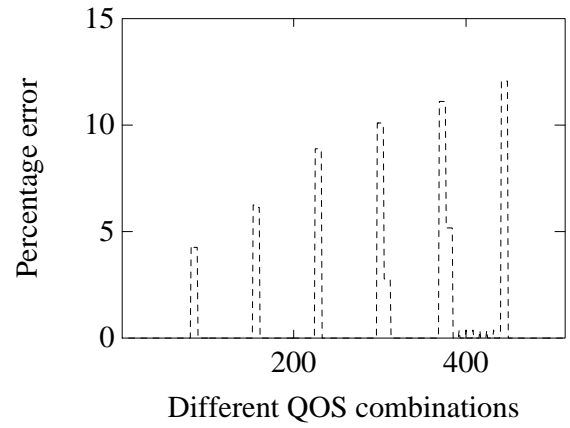


Figure 7. The percentage error of capacity required, $(C_{alg.2} - C_{opt})/C_{opt} \times 100\%$.

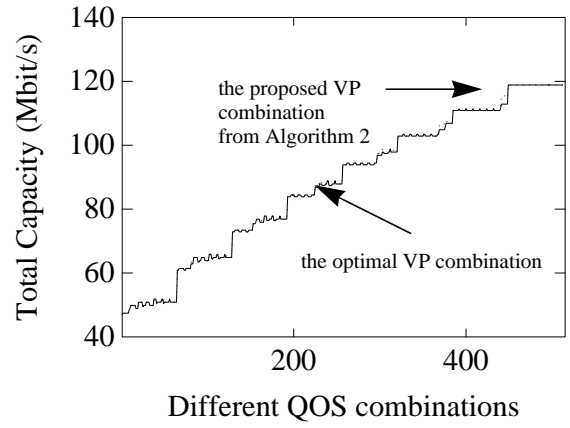


Figure 8. The optimal VP combination vs. the proposed VP combination

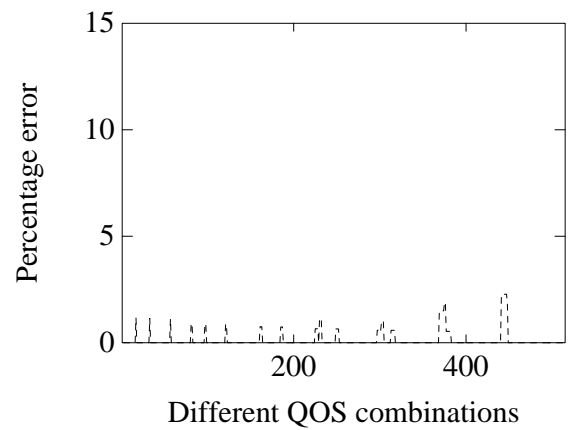


Figure 9. The percentage error of capacity required, $(C_{alg.2} - C_{opt})/C_{opt} \times 100\%$.