REVENUE MAXIMISATION IN ATM NETWORKS USING THE CLP CAPABILITY AND BUFFER PRIORITY MANAGEMENT

R. Sridhar*, C. Rosenberg†, A. Kumar*

Abstract—We are interested in the gain in network revenue that could be obtained by using the Cell Loss Priority (CLP) capability "optimally" (i.e., choosing an appropriate operating point and performing the right dimensioning) within the network, as compared with the case in which the user is not offered the CLP capability.

We study the performance of an ATM Multiplexer with two traffic classes with different QoS requirements, each class arriving as a Batch Markovian Arrival Process. The buffer priority schemes adopted are PBS (Partial Buffer Sharing) and PBS+Push-Out. We first obtain the engineering trade-off curves, between CLP=0 and CLP=1 traffic. Then we formulate a revenue optimisation problem, in which the constraints are the engineering trade-off function, and a simple model of the variation of CLP=1 demand with its price.

I. INTRODUCTION

We investigate the use of the CLP bit to allow the user to send through an ATM network two traffic classes that have different CLR requirements (the network is not allowed to tag cells). We are interested in the gain in network revenue that could be obtained by using the CLP capability "optimally" (i.e., choosing an appropriate operating point and performing the right dimensioning) as compared with a case where the user is not offered the CLP capability.

We tackle the problem in two stages. Firstly, we address the problem of joint traffic engineering of the network for CLP=0 (precious) and CLP=1 (less precious) traffic; i.e., for each level of CLP=0 load, we find the maximum CLP=1 load that can be handled so that the QoS requirements of each traffic type are met. Secondly, we propose a linear revenue function, and then, under the constraint of a simple demand versus price function for CLP=1 traffic, we obtain the point on the engineering trade-off curve at which the network should operate in order to maximise its revenue.

The CLP capability requires the implementation within the network of selective discarding schemes for giving priority to the CLP=0 traffic in case of congestion. Note that if the two classes of traffic are being offered by an application on the same Virtual Circuit (VC), then cell sequentiality should be preserved implying the use of nonspatial priority schemes. We have chosen to work with two selective discarding schemes. The first one is the well-known Partial Buffer Sharing (PBS) scheme (i.e., for a buffer of size K, there is a threshold K_1 beyond which CLP=1 cells are not accepted), and the second one is a combination of PBS with another well-known scheme called Push-out

*Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore-560 012, India.

†Département de Génie Electrique et Génie Informatique, Ecole Polytechnique de Montréal, Montréal, Canada. This work was done while Prof. Rosenberg was on sabbatical at the Indian Institute of Science.

(PBS+PO) ([6]) (i.e., an arriving CLP=1 is admitted only if the queue length is less than the threshold K_1; in addition to this, the last CLP=1 cell, if any, gets pushed out if a CLP=0 cell arrives and sees the buffer full).

We study the performance of an ATM buffer with the above selective discarding schemes and a discrete-time traffic model comprising the superposition of N independent and identical 2-state Markov modulated Bernoulli Processes (MMBPs). In the most general form of the model, both CLP=0 and CLP=1 cells can arrive in either phase of the modulating Markov process. The model can be taken to represent an ATM multiplexer with N input links, or an output queue of an N-port output queueing switch.

There is a large amount of literature on the performance analysis of ATM multiplexers. We list some representative references. The steady state analysis of the MMPP/G/1/K queue is dealt with in [1]. In [2], the buffer loss in the case of a finite capacity N/G/1 has been analysed. The performance of a statistical multiplexer for multi-class fluid sources is studied in [4]. In [6], the priority schemes such as PBS and PO have been proposed and the analysis has been carried out for Poisson arrivals and general service-time distributions. In [7], the analysis for the PBS scheme with a superpositon of N MMBPs has been carried out.

Our work differs from the above primarily in the use of the CLP bit for carrying differential QoS traffic, joint traffic engineering for CLP=0 and CLP=1 traffic, and an objective of revenue maximisation. Further, we have introduced the PBS+PO scheme. Throughout our study, the CLP=0 traffic is precious (i.e., $CLR_0 \leq \epsilon_0$), and is assumed to have been subjected to admission control procedures so that its traffic parameters are known to the network, whereas the CLP=1 (less precious) traffic can be one of the following :

(S1) uncontrolled (implying that nothing is known about these cells, all the offered traffic of this class is accepted by the network, no effort is made to police these cells) and with no QoS requirement (NQoS),

(S2a) controlled and NQoS (thus we control this traffic only for the sake of CLP=0 traffic or from a revenue point of view); in this case, the CLP=1 traffic too has known parameters.

(S2b) controlled with $CLR_1 \leq \epsilon_1$.

We want to compare situations (S1),(S2a) and (S2b) under the two schemes, viz. PBS and PBS+PO. For the situation (S2b) and each of PBS and PBS+PO, we obtain traffic engineering curves that bound the region of CLP=0 and CLP=1 loads that can be handled so that each meets its CLR requirements.

Further, defining ρ_0 (resp. ρ_1) = offered load of CLP=0 (resp. CLP=1) traffic, and γ_0 (resp. γ_1) = carried load

of CLP=0 (resp. CLP=1) traffic, we propose $R(\gamma_0, \gamma_1) = a\gamma_0 + b\gamma_1$, with $a > b$, as the network revenue function. Then, using a simple "power-law" form of the demand versus price function for CLP=1 traffic, we formulate the problem of choice of network operating point (carried traffic mix and pricing for CLP=1 service) as a constrained revenue maximisation problem.

The outline of this paper is as follows. In Section 2, we describe our model for the arrival process. The analysis of the PBS and PBS+PO schemes are not reported here due to lack of space. In Section 3, we present and discuss numerical results for traffic engineering with CLP=0 and CLP=1 traffic. In Section 4, we formulate a revenue maximisation problem and provide some numerical results. Section 5 contains our conclusions.

II. MODELS AND ANALYSIS

The ATM multiplexer (or the output queue of an ATM switch) receives cells from N independent ATM links and can buffer upto K cells. We observe the arrival and queue length processes at the epochs t_n, $n = 0, 1, 2, ...$, which are potential service completion epochs of a cell at the queue. Phase changes in the arrival processes occur at t_n^+ and cell arrivals (governed by the phase at t_n^+) occur over the interval $(t_n, t_{n+1}]$.

Each substream is a 2-phase MMBP. Each substream has a Phase 1 whose length is geometrically distributed with mean $L = 1/(1 - \beta)$ and a Phase 0 whose length is geometrically distributed with mean $S = 1/(1 - \alpha)$. During Phase 1, cells arrive in a Bernoulli process of rate p_1 and a fraction σ_1 of these are CLP=0 cells. During Phase 0, cells arrive in a Bernoulli process of rate p_0 and a fraction σ_0 of these are CLP=0 cells. In our discussions, the probability of arrival of CLP=1 cells is the same in both phases. The superposition of N such MMBPs is a Batch Markovian Arrival Process (BMAP). Given all the above parameters, the total offered load of the two classes is computed from:

$$\rho_0 = \frac{Lp_1\sigma_1 + Sp_0\sigma_0}{L + S}, \text{ and } \rho_1 = \frac{Lp_1 + Sp_0}{L + S} - \rho_0$$

Using standard techniques (see [7] and [6]), the Cell Loss Ratios (CLRs) of the CLP=0 and CLP=1 cells can be computed for PBS and PBS+PO schemes. We can thus engineer our network to carry a certain CLP=0 load and a certain CLP=1 load.

III. TRAFFIC ENGINEERING WITH CLP=0 AND CLP=1 TRAFFIC

We first compare PBS and PBS+PO. Observe that for small values of K_1 and large $\frac{K-K_1}{N}$, it is very likely that an accepted CLP=1 cell is transmitted before the buffer overflows; hence PBS and PBS+PO can be expected to be very close in performance. This is demonstrated in Figure 1 where $K=32$ and $N=4$, and we plot $\text{Log}(CLR_0)$ versus offered CLP=1 load with PBS and PBS+PO for $K_1=14$ and $K_1=24$. When $K_1=24$, PBS+PO is seen to be substantially superior to PBS.
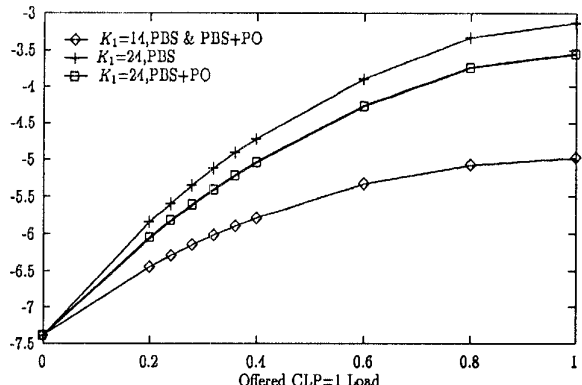

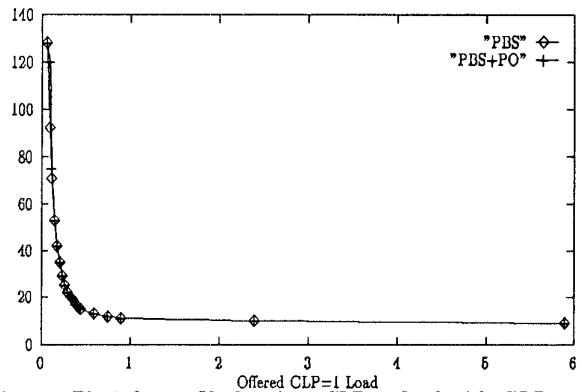Fig. 1. Plot of $\text{Log}(CLR_0)$ vs offered CLP=1 load; $K=32, N=4, L=25, S=100, \rho_0=0.24$

Fig. 2. Plot of max K_1 for given CLP=1 load with $CLR_0 < 10^{-9}$, $\rho_0=.0923$, $N=6$, $L=40$, $S=1000$, $K=128$

Uncontrolled CLP=1 traffic is represented by cells arriving on all substreams, on all slots. If we can choose $K_1 > 0$ such that $CLR_0 < 10^{-9}$ even under the worst-case conditions, we can afford to admit all the offered CLP=1 traffic into the network (although, of course, not all of it will be carried).

CLP=1 traffic will, however, need to be controlled for two possible reasons: (i) We cannot find $K_1 > 0$ for worst-case CLP=1 load [1] such that $CLR_0 \le \varepsilon_0$; (ii) We expect the real offered load of CLP=1 cells to be much below the worst case CLP=1 load. The problem with not controlling the CLP=1 traffic here is that we have to make a conservative choice for K_1 which could lead to a poor QoS for the CLP=1 cells.

The second point is clear from Figure 2 where we plot, for fixed CLP=0 load, the maximum value K_1 can take for each CLP=1 load such that $CLR_0 < 10^{-9}$. (The traffic parameters L and S and the switch parameter N, considered in Figure 2 will be used repeatedly as a running example.) Observe from Figure 2 that if we do not control CLP=1 traffic, we are forced to choose $K_1=10$. On the other hand, if the offered CLP=1 load were as small as 0.3 or so, and *knowing this*, we decide to control CLP=1 traffic to a maximum load of 0.3, then we can afford a $K_1=23$. The advantage

[1]The worst-case as per our model is a CLP=1 load of $N - \rho_0$, i.e., all substreams being flooded with either CLP=0 or CLP=1 cells.

Fig. 3. Plot of $Log(CLR_1)$ vs offered CLP=1 load; $N=6$, $K=128$, $L=40$, $S=1000$, $CLR_0 < 10^{-9}$, $\rho_0 = .0923$

Fig. 4. Plot of γ_{1max} vs γ_0 for $CLR_0 \leq 10^{-9}$ and $CLR_1 \leq 10^{-4}$, $N=6, L=40, S=1000$

of this larger K_1 together with controlled CLP=1 traffic can be seen from Figure 3 where we plot $Log(CLR_1)$ versus offered CLP=1 load for the uncontrolled and controlled cases, for the parameters of Figure 2. Figure 3 shows that, with a higher K_1, a CLR_1 of 10^{-4} can be achieved for a CLP=1 load of about 0.3 whereas, with $K_1=10$ (the value for uncontrolled CLP=1 traffic), CLR_1 is worse than 10^{-4} even for CLP=1 loads smaller than 0.3.

We now turn to the problem of obtaining engineering trade-off curves for CLP=0 and CLP=1 traffic. We assume that CLP=1 traffic is controlled and is given a QoS guarantee i.e., $CLR_1 \leq \epsilon_1$, $\epsilon_1 \geq \epsilon_0$. We compute the maximum CLP=1 throughput, γ_{1max}, that can be handled for a given CLP=0 throughput, γ_0, assuming that CLR requirements on both the classes are respected. In Figure 4, we have plotted γ_{1max} vs γ_0 for $CLR_0 \leq 10^{-9}$ and $CLR_1 \leq 10^{-4}$. The traffic parameters are the same as in Figures 2 and 3. Two sets of curves are shown, one for $K=64$ and the other for $K=128$. Observe that the performance of PBS+PO is different from that of PBS only when the carried CLP=0 load is small, i.e., when K_1 is large.

Figure 4 gives an example of a set of engineering trade-off curves. Next, we formulate the problem of determining a "good" operating point on the traffic engineering curve via a revenue maximisation approach.

IV. REVENUE MAXIMISATION

A natural form for the network's revenue function is $R_{a,b}(\gamma_0, \gamma_1) = a\gamma_0 + b\gamma_1$ where a/b (≥ 1) represents the proportionality factor between what the network charges for CLP=0 traffic versus CLP=1 traffic. To simplify matters, we assume $a=1$, and so we are trying to find for a given K, a given selective discarding scheme and a given b ($\in [0, 1]$), the maximum revenue. We now formulate the revenue maximisation problem. The first element of this problem is an engineering curve (for fixed K and selective discarding scheme) which we denote by $\eta(\gamma_0)$. Denote $\overline{\gamma_1} = \eta(0)$.

The next element of the formulation is the variation of CLP=1 demand with b, the CLP=1 tariff.[3] We denote this function by $x_1(b)$. Since CLR is very small (10^{-4} or less), the carried load is practically the same as offered load; hence we will think of $x_1(b)$ as an achievable bound on the carried load of CLP=1.

We consider the following form of $x_1(b)$: $x_1(b) = A_1 b^{-\alpha}$ where $\alpha \geq 0$. As may be expected, demand for CLP=1 service decreases with increasing price; the decrease is steeper for larger α. $A_1 > 0$ models the residual CLP=1 demand when CLP=1 is priced the same as CLP=0. The point here is that, even though CLP=1 service is priced the same as CLP=0 service, all the CLP=1 demand cannot shift to CLP=0, as the network cannot carry that much CLP=0 traffic.

In economic terms, α is called the elasticity of demand with price. We will consider three cases.

- Case 1: $x_1(b) = A_1$, ($\alpha = 0$) and $A_1 \geq \overline{\gamma_1}$. This case corresponds to no demand constraint on the problem; the network will get as much CLP=1 traffic as it wants.
- Case 2: $x_1(b) = A_1 b^{-1/2}$ (i.e. elasticity =1/2) and $A_1 < \overline{\gamma_1}$.
- Case 3: $x_1(b) = A_1 b^{-2}$ (i.e. elasticity =2) and $A_1 < \overline{\gamma_1}$. Cases 2 and 3 correspond to the situation where demand for CLP=1 service decreases with price (more steeply for $\alpha=2$), and for $b=a=1$, the demand is less than the maximum CLP=1 load that the network can carry.

With the above elements, the revenue maximisation problem becomes: $\max_{\gamma_0, \gamma_1} R = \gamma_0 + b\gamma_1$ subject to $\gamma_0 \leq \gamma_{0max}$, $\gamma_1 \leq min(\eta(\gamma_0), x_1(b))$.

For the parameters as in Figures 2, 3 and 4, Figure 5 shows revenue optimisation results for the CLP=1 demand curve of Case 1 ($\alpha=0$). Observe, in Figure 5, that the revenue obtained increases almost linearly with b. This is because the CLP=1 traffic we have considered is not bursty. We find that for every b, the revenue as obtained in Figure 5 is very close to the bound $(a-b)\rho_{0max} + b$. Table 1 gives the optimum operating point (in terms of offered loads) and the corresponding revenue for the PBS and PBS+PO cases, as a function of b, (again, with $a=1$) for the traffic parameters in Figure 5. Observe that the operating points for the same b can be considerably different for PBS and PBS+PO.

[3] Strictly speaking, we also ought to consider the variation of demand for CLP=0 service with a, but, in order to get a simple model, we fix a at 1 and assume that the CLP=0 demand exceeds γ_{0max}.
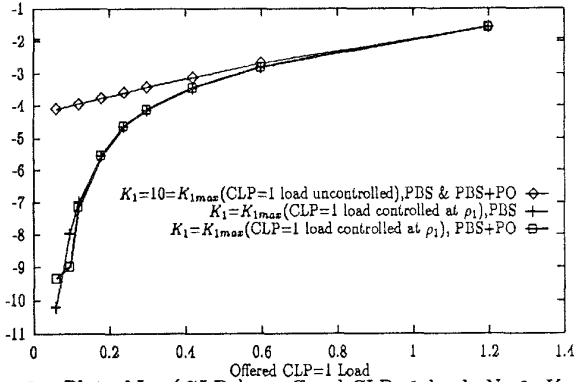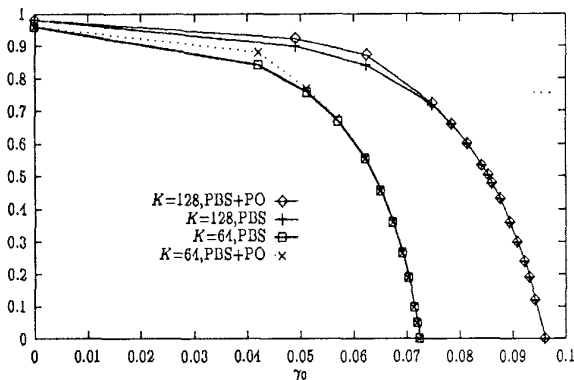
Fig. 5. Plot of maximum revenue R^* vs b ($\alpha = 0$); K=128, N=6, L=40, S=1000, a=1, $A_1 \geq \overline{\gamma_1}$

Fig. 6. Plot of optimum revenue R^* and corresponding CLP=0 load γ_0 vs b for price-dependent CLP=1 load, N=6, L=40, S=1000, K=128, a=1, A_1=0.2.

a=1.0	PBS				PBS+PO			
	R^*	ρ_0	ρ_1	K_1	R^*	ρ_0	ρ_1	K_1
b=0	.096	.096	0	0	.096	.096	0	0
b=.05	.111	.081	.6	53	.111	.081	.6	53
b=.1	.15	.075	.72	76	.15	.062	.87	125
b=.015	.19	.062	.84	99	.19	.062	.87	125
b=.2	.23	.062	.84	99	.24	.062	.87	125
b=.25	.27	.049	.9	114	.28	.062	.87	125
b=.3	.32	.049	.9	114	.33	.049	.92	128
b=.4	.41	.049	.9	114	.42	.049	.92	128
b=.5	.5	.049	.9	114	.51	.049	.92	128
b=.7	.69	0	.98	128	.69	.049	.92	128
b=.9	.88	0	.98	128	.88	0	.98	128
b=1.0	.98	0	.98	128	.98	0	.98	128

TABLE I

MAXIMUM REVENUE AND OPTIMUM POINT FOR VARIOUS b, α=0, N=6, L=40, S=1000, K=128, $A_1 \geq \overline{\gamma_1}$

Observe from Table 1 that if b is not too small, our operating point has a high ρ_1 and $\rho_0 < \rho_{0max}(= 0.096)$ and there is an appreciable improvement in the revenue when compared with single-class operation (i.e., b=0). (This gain can be even greater than that obtained by increasing the buffer size without adding CLP=1 traffic; for example, Figure 4 shows that if K is increased from 64 to 128, the increase in γ_{0max} is only from 0.072 to 0.096. A better improvement could be achieved through using the CLP capability.) On the other hand if the ratio a/b is very large, the operating point has a very low ρ_1 and $\rho_0 \approx \rho_{0max}$, and we might not be able to gain much in revenue by introducing CLP=1 cells.

Further, from Figure 5 as well as Table 1, it is clear that the total revenue obtained using the PBS scheme is nearly the same as that obtained using PBS+PO, hence PBS, being simpler to implement, should be preferred.

Recall that Figure 5 is for the demand-price function with α=0 and $A_1 \geq \overline{\gamma_1}$. Thus, the demand curve is not a constraint at all and the results in Figure 5 are simply obtained by maximising, for each fixed b, the linear revenue function over the traffic trade-off curve $\eta(\gamma_0)$. If we do the same optimisation problem with demand curves $0.2b^{-1/2}$ and $0.2b^{-2}$, we get the optimal revenue curves (denoted by
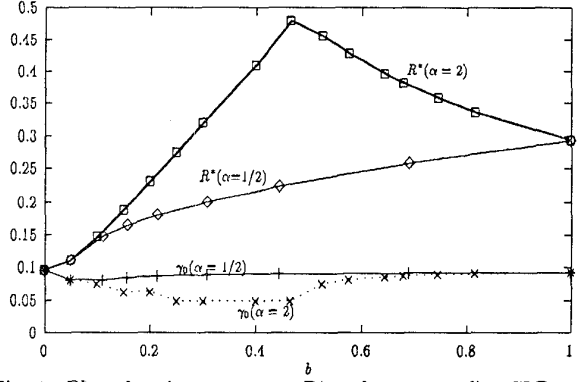
R^*) in Figure 6; also shown are the values of γ_0 at the optimal operating point for each b.

For b=0, the operating point is at ($\gamma_0 = \gamma_{0max}, \gamma_1 = 0$) and the demand constraint is not operative. As b increases, the operating point moves up along the engineering curve (and behaves just like in Figure 5) until the demand constraint becomes operative. This happens at b=0.1 for α=1/2 and at b=0.46 for α=2, in our example (Figure 6). Beyond this value of b, the revenue deviates from that in the case of α=0 and the operating point retraces its path along the engineering curve. For $\alpha < 1$, it can be shown that the revenue continues to increase, and when $\alpha > 1$, the revenue may decrease after a point; it can be argued that for $\alpha > 1$, and for very bursty CLP=0 traffic, the revenue will be optimised for $b < 1$.

We observe from Figure 6 that, unlike in Figure 5, the carried CLP=0 load at the revenue maximising operating points is a substantial fraction of γ_0. This is because in order for the revenue to be maximised for small values of γ_0 the value of b has to be large, but for large b the demand for CLP=1 also reduces. Further, there is significant improvement in network revenue if CLP=1 service is introduced provided it is priced correctly and the network is appropriately engineered. Finally, the operating point is quite sensitive to the demand versus price function for CLP=1 traffic; as per Figure 6, for $x_1(b) = 0.2b^{-1/2}$, the optimal operating point is ($\gamma_0 = 0.093, \gamma_1 = 0.2$), whereas, for $x_1(b) = 0.2b^{-2}$, the optimal operating point is ($\gamma_0 = 0.0491, \gamma_1 = 0.9$).

V. CONCLUSIONS

We have studied the advantage of using the CLP bit to carry traffic streams with differential QoS requirements, in an attempt to maximise network revenue. A single ATM multiplexer with PBS or PBS+PO is studied as a test case. The revenue is quantified by a linear revenue function of the form $a\gamma_0 + b\gamma_1$, where γ_0 and γ_1 are the carried loads of the CLP=0 and CLP=1 traffic.

If the multiplexer is engineered for uncontrolled CLP=1 traffic without QoS constraints then the PBS limit K_1 has to be set for the worst case. Then CLP=1 cell loss ratio is

very poor and it would be expected that a/b is large. In this case, there is no appreciable revenue gain in adding CLP=1 traffic. On the other hand, if CLP=1 traffic obeys a traffic contract, and demands a QoS ($CLR_1 >> CLR_0$) then it can be expected that a/b is not too large and some CLP=0 load can be traded off for carrying CLP=1 load, resulting in an overall increase in revenue. We have demonstrated this using a simple demand versus price formulation for CLP=1 traffic.

We have provided the CLP=0 vs CLP=1 trade-off curves, which, in conjunction with more sophisticated economic models, can be used to determine optimal network operating points.

ACKNOWLEDGEMENT

REFERENCES

[1] A. Baiocchi and N. Blefari-Melazzi *Steady State Analysis of the MMPP/G/1/K queue*: IEEE Transactions on Communications, Vol.41, No.4, April 1993.

[2] C.Blondia *The N/G/1 finite capacity queue*: Commn. Statist. Stochastic Models, Vol.5, No.2, 1989.

[3] L.Breiman *Probability* Addison-Wesley Series, 1968.

[4] A. Elwalid and D. Mitra *Fluid Models for the Analysis and Design of Statistical Multiplexing with Loss Priorities on Multiple Classes of Bursty Traffic*: INFOCOM'92

[5] W.K.Grassman, M.I.Taksar and D.P.Heyman *Regenerative Analysis and Steady State Distributions for Markov Chains*: Operations Reasearch , Vol.33, No.5, 1985.

[6] H.Kroner, G.Hebuterne, P.Boyer and A.Gravey *Priority Management in ATM Switching Nodes*: IEEE Journal on Selected Areas in Communications, Vol.9, No.3, April 1991.

[7] J.Y.Le Boudec *An Efficient Solution Method for Markov Models of ATM Links with Loss Priorities* IEEE Journal on Selected Areas in Communications, Vol.9, No.3, April 1991.