PROCEEDINGS
EUROPTO
SERIES

Broadband Networks: Strategies and Technologies

Robert A. Cryan
P. Nalinaj Fernando
Pierpaolo Ghiggino
John M. Senior
Chairs/Editors

20–23 March 1995
Amsterdam, The Netherlands

Sponsored by
EOS—The European Optical Society
SPIE—The International Society for Optical Engineering
The Commission of the European Communities, Directorate General for Science, Research, and Development

SPIE P PROCEEDINGS SERIES

Volume 2450

PROCEEDINGS
EUROPTO
SERIES

Broadband Networks: Strategies and Technologies

Robert A. Cryan
P. Nalinaj Fernando
Pierpaolo Ghiggino
John M. Senior
Chairs/Editors

20–23 March 1995
Amsterdam, The Netherlands

Sponsored by
EOS—The European Optical Society
SPIE—The International Society for Optical Engineering
The Commission of the European Communities, Directorate General for Science, Research, and Development

Published by
SPIE—The International Society for Optical Engineering

SPIE PROCEEDINGS SERIES

Volume 2450

Characterizing bursty source models applied to admission control in ATM networks

R. Bolla, F. Davoli, M. Marchese

Department of Communications, Computer and Systems Science (DIST)
University of Genoa
Via Opera Pia, 13 - 16145 Genova, Italy

ABSTRACT.

This paper proposes different models to describe a bursty source; models are based on Markov chains whose complexity is increasing with the accuracy of the description. The effectiveness of the proposed models is evaluated by varying statistical description parameters of sources like peak bandwidth and burstiness. The effect of the models presented on an ATM call admission strategy is evaluated; the latter is suited to guarantee quality of service requirements like lost and delayed cell rates in ATM networks with multi-rate bursty traffic. The efficiency of the admission control is verified by a simulation tool.

1. INTRODUCTION

In Asynchronous Transfer Mode (ATM) networks, the statistical multiplexing of cells associated to nonhomogeneous traffic flows (narrowband, broadband, time- or loss-sensitive, etc.) originates the need for a number of congestion control mechanisms that should be applied to meet the generally different performance requirements of the various users and ensure the necessary levels of Quality of Service (QoS). Moreover, the statistical description of many different sources involves a great effort in mathematical modeling, in order to capture the essential characteristics of the traffic.

In these respects, several source models have been proposed and analyzed, as well as bandwidth allocation and Call Admission Control (CAC) mechanisms (see, for instance, [8] and [10,11,12], respectively). With regard to the latter, the acceptance of a new connection should be conditioned on the possibility of guaranteeing the required performance, without impairing that of the existing connections. An architecture for CAC and bandwidth allocation among traffic classes with different statistical parameters and performance requirements has been proposed by the authors for

0-8194-1801-3/95/$6.00

an access node in [9], and later extended to include network-wide CAC and routing [13,14]. The source models used in these approaches, where the focus was on the control structure, have been limited to very simple descriptions of on-off (bursty) sources.In the present paper, we discuss a number of different models for bursty sources of the same kind, with different accuracy, but essentially capable of capturing the same characteristics. Then, we briefly review the previously mentioned CAC and bandwidth allocation scheme, and present some further simulation results.
The paper is organized as follows. In Section 2 a detailed description of a popular model for a bursty source is shown; two more complex models are described in Section 3; Section 4 shows an admission control strategy based on the bursty source model described in Section 2; simulation results to verify the effectiveness of the admission control strategy are presented in Section 5. Section 6 contains conclusions.

2. A BURSTY SOURCE MODEL

As shown, for instance, in [1], an accurate description for several types of bursty sources can be given by a Talkspurt-Silence model, a two-state Markov chain alternating periods of activity (burst periods) and periods of silence. Within a burst period cells are generated with known and constant bit rate. If the model is considered in continuous-time domain, the sojourn times in both states are exponentially distributed with known mean value; the probability density function of cell interarrival time for one source is depicted in [2]. If the model is used in discrete-time domain, sojourn times have geometric distribution and the model is equivalent to a two-state Markov Modulated Deterministic Process (MMDP) with one silence state [1]. The model described in the following, already used in [6,7,9], is an Interrupted Bernoulli Process (IBP), discrete-time version of the Interrupted Poisson Process (IPP) [1,3].
Considering a channel with capacity C_T, a bursty connection of a particular traffic class (h) can be described by a two state model: 'active' and 'idle' as in the Talkspurt-Silence discrete-time model.

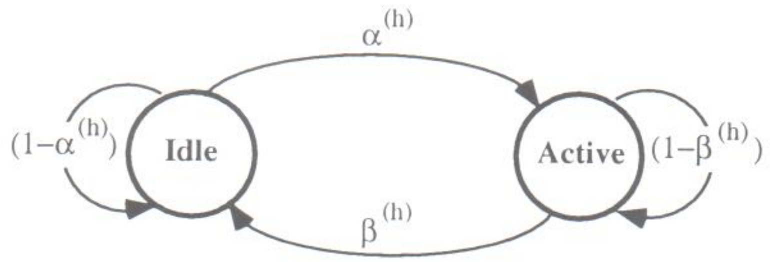

Fig.1 IPP bursty source model

The probabilities to be in the 'idle' and 'active' state are:

$$\omega_i^{(h)} = \frac{\beta^{(h)}}{\alpha^{(h)} + \beta^{(h)}} \tag{1}$$

$$\omega_a^{(h)} = \frac{\alpha^{(h)}}{\alpha^{(h)} + \beta^{(h)}} \tag{2}$$

where $\alpha^{(h)}=\frac{1}{B^{(h)}(b^{(h)}-1)}$ and $\beta^{(h)}=\frac{1}{B^{(h)}}$, $B^{(h)}$ being the mean value of the burst length in cells, and $b^{(h)}=P^{(h)}/A^{(h)}$ the value of the burstiness for traffic class (h), having peak bandwidth $P^{(h)}$ and average bandwidth $A^{(h)}$. Then, $\omega_i{}^{(h)}=\frac{b^{(h)}-1}{b^{(h)}}$ and $\omega_a{}^{(h)}=\frac{1}{b^{(h)}}$ It is important to note that these values are independent of the burst length.

Within the active state, the arrival process of cells is modelled by a Bernoulli process: at each time instant, there is a cell arrival with probability $\frac{1}{M^{(h)}}$ with $M^{(h)}=C_T/P^{(h)}$. As will be shown below, the average number of empty cells between two successive cell generations for traffic class (h) is $(M^{(h)}-1)$; in the Talkspurt-Silence model, where the distance between cells within a talkspurt is deterministic, $\lceil M^{(h)}-1 \rceil$ is actually the number of empty cells between two successive cell generations ($\lceil x \rceil$ represents the smallest integer greater than or equal to x). So, the interarrival time between two successive cells in this description has a geometric distribution with a minimum of one cell; the interarrival times (in cells) are considered to be independent and identically distributed, i.e. belonging to the class of renewal processes.

Since in this case the number of empty cells between two consecutive cell generations is a random variable with known distribution, it is possible to compute and depict the density, the mean value and the variance of this variable and, then, to show the difference with the accurate Talkspurt-Silence model.

So, defining by **k** the random variable "number of empty cells between two successive cell arrivals of traffic class (h)", it is true [4] that :

$$\Pr\{\mathbf{k}=k\}=\frac{1}{M^{(h)}}\left(1-\frac{1}{M^{(h)}}\right)^k \tag{3}$$

The distribution is, so, a staircase function:

$$F(x) = \Pr\{\mathbf{k} \le x\} = \sum_{r=0}^{k} \frac{1}{M^{(h)}}\left(1-\frac{1}{M^{(h)}}\right)^r =$$

$$= \frac{1}{M^{(h)}} \sum_{r=0}^{k} \left(1-\frac{1}{M^{(h)}}\right)^r = \frac{1}{M^{(h)}} \frac{\left(1-\frac{1}{M^{(h)}}\right)^{k+1} - 1}{\left(1-\frac{1}{M^{(h)}}\right) - 1} = 1 - \left(1-\frac{1}{M^{(h)}}\right)^{k+1} \quad , k \le x < k+1 \tag{4}$$

As a consequence, the density function of the defined random variable **k** is a sum of impulses:

$$f(x) = \sum_{r=0}^{\infty} \frac{1}{M^{(h)}}\left(1-\frac{1}{M^{(h)}}\right)^{r} \delta(x-r) \tag{5}$$

The random variable **k** has mean value

$$\eta_{\mathbf{k}}=M^{(h)}-1 \tag{6}$$

and variance

$$\sigma_{\mathbf{k}}^{2}=\frac{2\left(1-\frac{1}{M^{(h)}}\right)}{\left(\frac{1}{M^{(h)}}\right)^{2}} = 2M^{(h)}(M^{(h)}-1) \tag{7}$$

In this way, the random variable can be considered described. It can be observed by evaluating the density function and the variance that the obtained values are not quite close to the deterministic values we can obtain from the Talkspurt-Silence model [2]; however, it is shown in the following (Section 4) that by using this model with overlapped sources and with a well suited admission control, good results can be obtained without using more accurate models, as those described in the next Section.

3. TWO ALTERNATIVE MODELS

The first model introduced can be described by a 3-state Markov chain, as depicted in Fig.2.

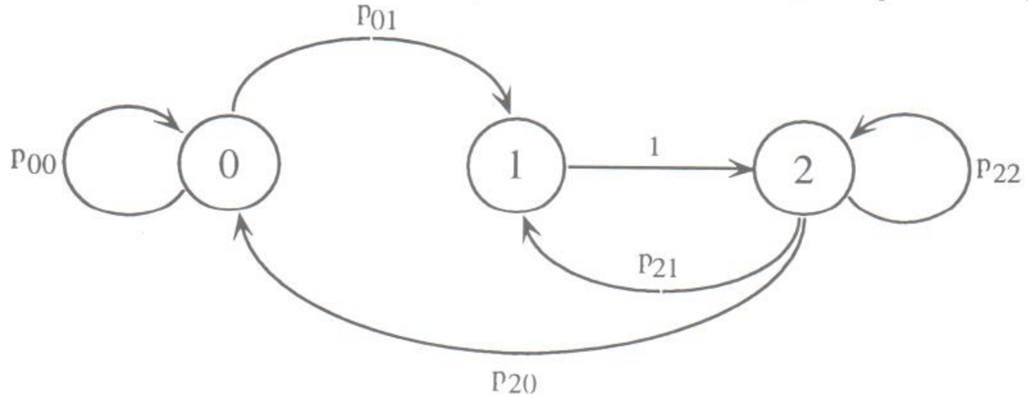

Fig. 2 Three-state source model

State 0 is the non-burst state, state 1 is the generating cell burst state, while state 2 is the non-generating cell burst state.The transition probabilities can be obtained by simple mathematical steps; it will be:

$$p_{00}=1-p_{01};\ p_{01}=\frac{1}{B^{(h)}(b^{(h)}-1)};\ p_{20}=\frac{M^{(h)}}{B^{(h)}(M^{(h)}-1)}; \tag{8}$$

$$p_{21}=\frac{B^{(h)}-M^{(h)}}{B^{(h)}(M^{(h)}-1)};\ p_{22}=\frac{M^{(h)}-1}{M^{(h)}-2}$$

The second model we describe, already introduced in [5], by using probability generating functions, is a Markov chain with $(M^{(h)}+1)$ states, matching the behaviour of the Talkspurt-Silence model. Time is assumed slotted and the cell arrival process in each slot is governed by a 'modulating Markov chain', where transitions between states of the chain take place only at slot boundaries. In Fig.3, the corresponding Markov chain is shown, supposing $(M^{(h)}-1)$ empty cells between two consecutive arrivals.

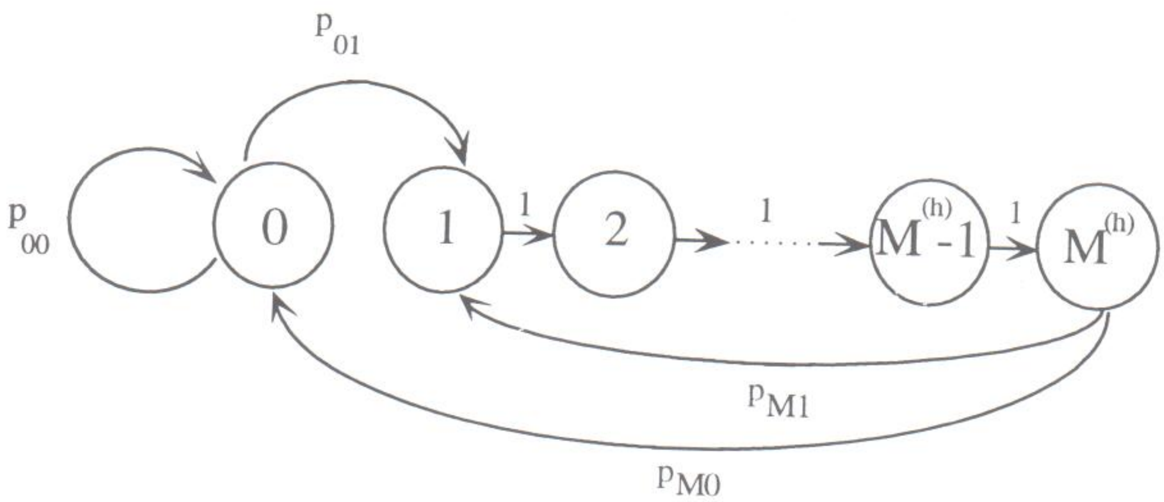

Fig. 3 M-states source model

State 0 is the Silence state, while states 1 through $M^{(h)}$ are bursty ones, but only 1 is a generating state. It is possible to compute the transition probabilities, obtaining:

$$p_{00}=1-p_{01};\ p_{01}=\frac{1}{B^{(h)}(b^{(h)}-1)};$$

$$p_{M0}=\frac{M^{(h)}}{B^{(h)}};\ p_{M1}=\frac{B^{(h)}-M^{(h)}}{B^{(h)}} \tag{9}$$

By computing the distribution and the density of the interarrival process, a high level of accuracy is obtained in both cases which have been presented; in the second one, the result obtained is the discrete-time case of the density shown in [2]. But, to fully utilize the accuracy of these descriptions, the use of queueing theory and related approximations is needed to get the solution of complex mathematical problems. For example, a possible approach is based on the usage of the characteristic function, as in [5].

The approach presented here is based on a simple observation; if the static probability p_g of generating a cell at a random instant is evaluated, it can be shown by simple mathematical steps that in all three cases we have:

$$p_g = \frac{1}{M^{(h)} b^{(h)}} \tag{10}$$

So, the accuracy of the model does not affect this result. The same result would be obtained by considering two Bernoulli processes "encapsulated" one inside the other; the first one describing the burst arrival time, the second one the cell arrival time within a burst. Considering a slot as time unit, the probability to be in "burst-state" shall be $1/b^{(h)}$ and, being within a burst, the probability of generating a cell for traffic class (h) shall be $1/M^{(h)}$. The probability that a cell is generated in a slot is p_g. In the next Section, a CAC scheme is introduced, based on the first presented model (IPP bursty source model) that is completely equivalent, as already said, to two encapsulated Bernoulli processes; the issue of batch arrival is taken into account.

4. A CONNECTION ADMISSION CONTROL

An ATM network node shared by several different traffic classes (h=1,...,H) is considered.

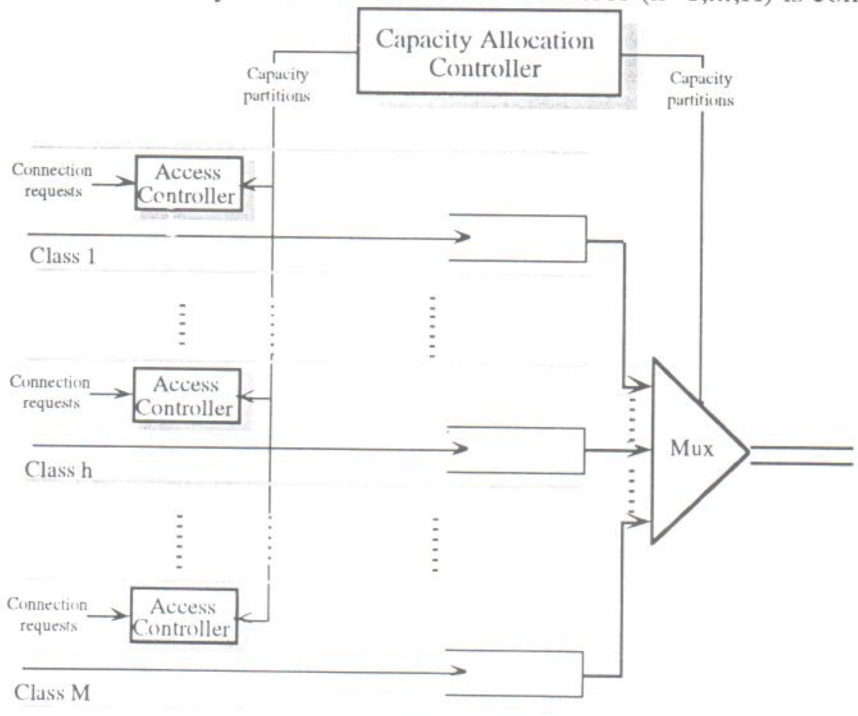

Fig. 4 Control Structure of a link

Each traffic class is characterized by statistical parameters (peak ($P^{(h)}$) and average bandwidth ($A^{(h)}$)) and by performance requirements (cell loss rate, delayed cells rate); the control structure of each link is shown in Fig. 4.

The output link (only one for semplicity) has a buffer for each traffic class, served by a "dedicated" partition of link bandwidth. The bandwidth partitions $V_m^{(h)}$, h=1,...,H, are computed by a Capacity Allocation Controller at discrete time instants m=[0,K,2K,...], by minimizing a cost function with equality and inequality constraints, explained in detail in the following.

The values of the partitions are communicated to a multiplexer, which picks up the cells from dedicated buffers, by using a strategy depending on partitions allocated to each class; the same values are transmitted to Access Controllers (one for each class) that compute the maximum number of acceptable connections for a traffic class on the output link.

Before introducing the Admission Control Strategy in detail it is necessary to define:

$N_{max}^{(h)}(m)$ Maximum number of acceptable connections for traffic class (h) computed at the instant m (intervention instant of the Capacity Allocation Controller).

$N_a^{(h)}(k)$ Number of connections in the system for traffic class (h) computed at the generic instant k.

$N_L^{(h)}(m)$ Maximum number of acceptable connections for traffic class (h) that maintains the cell loss rate below a given upper bound (cell loss requirement).

$N_D^{(h)}(m)$ Maximum number of acceptable connections for traffic class (h) that maintains the delayed cells rate below a given upper bound (cell delay requirement);

The maximum number of acceptable connection for traffic class (h) shall be the minimum between the maximum number of acceptable connections satisfying the loss requirement and the delay requirement, i.e.:

$$N_{max}^{(h)}(m) = \min\{ N_L^{(h)}(m); N_D^{(h)}(m) \} \qquad (11)$$

The Admission Strategy is now quite simple: a new connection of traffic class (h) is accepted if the sum between the connections of class (h) already in the system and the requesting one does not exceed the maximum number of acceptable connections for traffic class (h). More briefly:

$$\begin{aligned} &\text{connection accepted if} \quad N_a^{(h)}(k)+1 \le N_{max}^{(h)}(m) \\ &\text{connection rejected if} \quad N_a^{(h)}(k)+1 > N_{max}^{(h)}(m) \end{aligned} \qquad (12)$$

The computation of $N_L^{(h)}(m)$ and $N_D^{(h)}(m)$ is performed by imposing upper limits ($\varepsilon^{(h)}$,$\delta^{(h)}$) on the average value of cell loss rate and delayed cell rate, respectively:

$$N_L^{(h)} = \max_N \left\{ N: \sum_{n=0}^{N} P_{loss}^{(h)}(n) v_{n,N}^{(h)} \le \varepsilon^{(h)} \right\} \tag{13}$$

$$N_D^{(h)} = \max_N \left\{ N: \sum_{n=0}^{N} P_{delay}^{(h)}(n) v_{n,N}^{(h)} \le \delta^{(h)} \right\} \tag{14}$$

where

$$v_{\mathbf{n,N}}^{(h)} = \binom{N}{n} \left(\omega_{\mathbf{a}}^{(\mathbf{h})}\right)^{\mathbf{n}} \left(\omega_{\mathbf{i}}^{(\mathbf{h})}\right)^{\mathbf{N}-\mathbf{n}} \tag{15}$$

is the probability that n connections are active, having N connections accepted in the network (traffic class (h)). The formula written above is immediately deduced from the IPP model with batch arrivals as in [1].

The cell loss rate for the considered output link, given n calls in the active state is

$$P_{loss}^{(h)} = \sum_{i=0}^{Q^{(h)}} \Pi_i^{(h)} \frac{\sum_{j=0}^{n} \max(i+j-Q^{(h)},0) f_j^{(h)}(n)}{n\,\Gamma^{(h)}} \tag{16}$$

where

$$\Gamma^{(h)} = \frac{P^{(h)}}{C_T} = \frac{1}{M^{(h)}} \tag{17}$$

$$n\,\Gamma^{(h)} = \sum_{j=0}^{n} j f_j^{(h)}(n) \tag{18}$$

and

$$f_i^{(h)}(n) = \begin{cases} 0 & i < 0 \\ \binom{n}{i} (\Gamma^{(h)})^i (1-\Gamma^{(h)})^{n-i} & 0 \le i \le n \end{cases} \tag{19}$$

$f_i^{(h)}(n)$ is the probability of having i connections of traffic class (h) generating a cell with n connections of the same class in the active state.

$Q^{(h)}$ is the fixed dimension of the output buffer of traffic class (h) and $\Pi_i^{(h)}$ the steady state probability of having i cells inside the buffer, as computed in [6,7,9].

Similarly, the delayed cell rate is:

$$P_{delay}^{(h)} = \sum_{i=0}^{Q^{(h)}} \Pi_i^{(h)} \frac{\sum_{j=0}^{n} \max\left[(i+j-\max(i,\hat{Q}_m^{(h)}) - \max(i+j-Q^{(h)},0)\right] f_j^{(h)}(n)}{n\,\Gamma^{(h)}} \tag{20}$$

where

$$\hat{Q}_m^{(h)} = \frac{V_m^{(h)} D^{(h)}}{C_T} \quad (21)$$

and $D^{(h)}$ is the number of cells beyond which a cell is considered delayed.

As the Bandwidth Allocation Mechanism is concerned, the aim of the allocation controller is to provide an efficient bandwidth sharing on the basis of the total traffic intensity of the various classes.

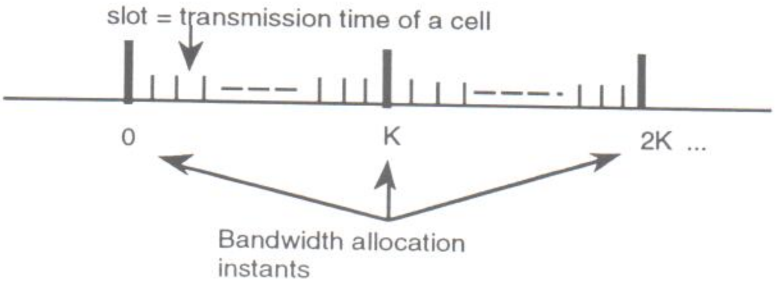

Fig. 5 Bandwidth Allocation Controller decision instants

At each decision interval (0, K, 2K,....), the minimization is performed with respect to the bandwidth, of a cost function of the form

$$J_p = \sum_{h=1}^{H} \sigma^{(h)} \sum_{n=0}^{N^{(h)}} P_{loss}^{(h)}(n)\, v_{n,N^{(h)}}^{(h)} \quad (22)$$

where $N^{(h)} = N_a^{(h)} + N_b^{(h)}$, $N_a^{(h)}$ is the current number of accepted connections at the current reallocation instant and $N_b^{(h)}$ is the number of connections refused in the previous interval (the interval between the current reallocation instant and the previous one). The quantity $\sigma^{(h)}$ is a coefficient to weight the different traffic classes.

After minimization, the values of capacity partitions are obtained for each traffic class. The value $V_m^{(h)}$ is the bandwidth allocated to traffic class (h) in the reallocation instant m, as already said.

The minimization is subject to the following constraints

$$\sum_{h=1}^{M} V_m^{(h)} = C_T$$

$$V_m^{(h)} \geq V_{min}^{(h)} \quad (23)$$

with $V_{min}^{(h)}$ the minimum capacity necessary to cope with the connections already in progress for traffic class (h).

5.SIMULATION RESULTS

In this Section, we show and briefly comment some results of simulation results, which were based on the following data.

C_T = 150 Mbit/s; M =3; T_s = slot duration = $2.83 \cdot 10^{-6}$ s (53 bytes/cell); K = $2 \cdot 10^6$ cells

$P^{(1)}$ = 128 kbit/s; $P^{(2)}$ = 2 Mbit/s; $P^{(3)}$ = 10 Mbit/s

$b^{(1)} = 2$; $b^{(2)} = 5$; $b^{(3)} = 10$

$B^{(1)} = 100$; $B^{(2)} = 500$; $B^{(3)} = 1000$ cells

$\varepsilon^{(1)} = \varepsilon^{(2)} = \varepsilon^{(3)} = 1 \cdot 10^{-4}$ $\delta^{(1)} = \delta^{(2)} = \delta^{(3)} = 1 \cdot 10^{-3}$

$D^{(1)} = 400$; $D^{(2)} = 200$; $D^{(3)} = 100$ slots

$Q^{(1)}=11$; $Q^{(2)}=12$; $Q^{(3)}=12$ cells

Fig. 6 depicts the maximum number of acceptable connections for traffic class 1 by applying the admission rule (12) and the maximum number obtained by a simulation tool, designed to test the efficiency of the access strategies.

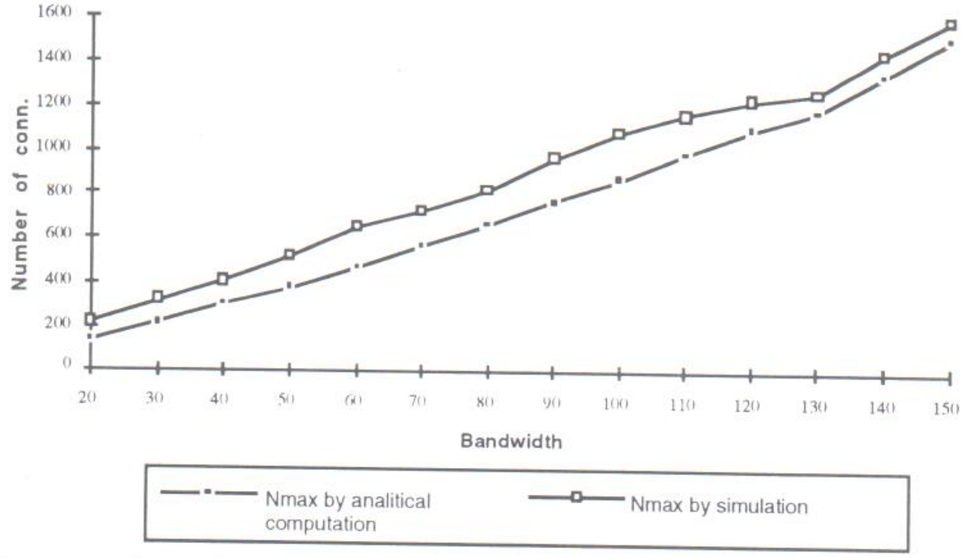

Fig. 6 Maximum number .of acceptable connections vs allocated bandwidth [Mbit/s] (class 1)

It can be seen that analytical and simulation values are quite close, and can guarantee a good performance of the access control. The same graph is shown in Fig. 7 and Fig. 8 for traffic class 2 and 3 respectively.

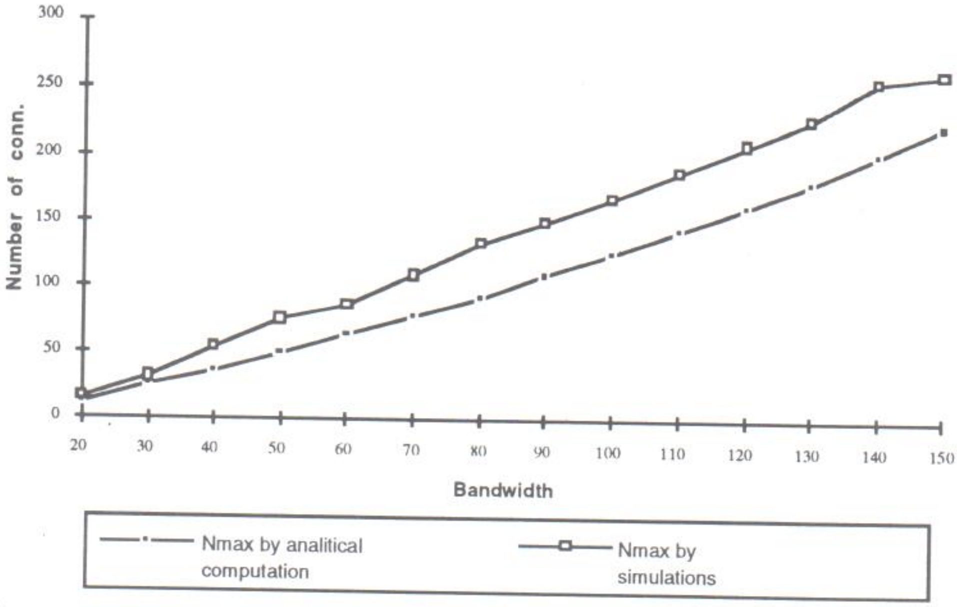

Fig. 7 Maximum number .of acceptable connections vs allocated bandwidth [Mbit/s] (class 2)

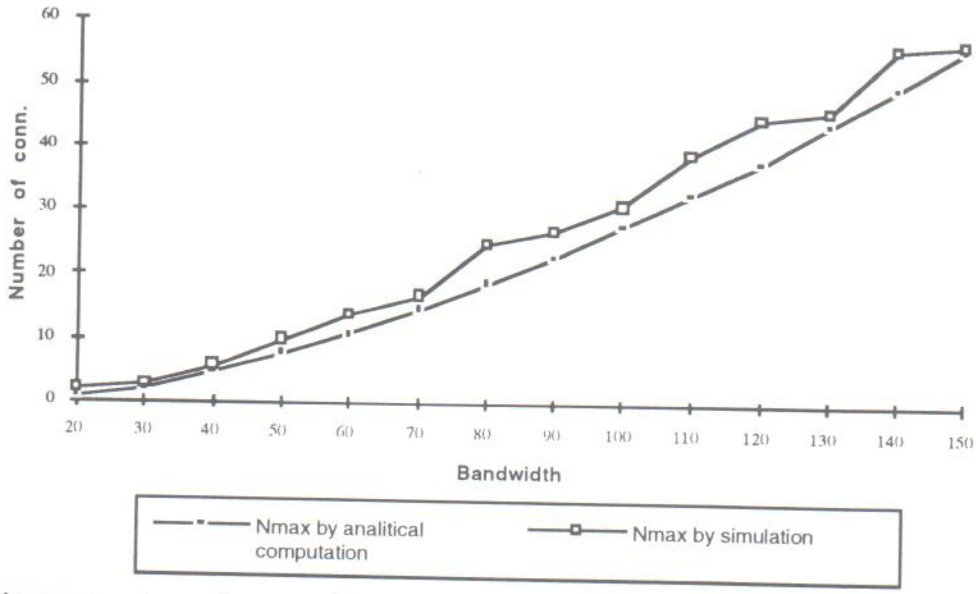

Fig. 8 Maximum number .of acceptable connections vs allocated bandwidth [Mbit/s] (class 3)

The degree of approximation for the cell loss rate can be seen in Fig.9, where cell loss rate (P_{loss}) for class 1,2,3, obtained applying the formula (16) and obtained by simulation, are depicted versus the number of connections in progress.

# of conn.	900	1050	175	200	40	50
Class 1 P_{loss} (analyt.)	$1.53 \cdot 10^{-4}$	$1.14 \cdot 10^{-3}$				
Class 1 P_{loss} (simul.)	$1.93 \cdot 10^{-5}$	$1.12 \cdot 10^{-4}$				
Class 2 P_{loss} (analyt.)			$4.94 \cdot 10^{-3}$	$1.98 \cdot 10^{-2}$		
Class 2 P_{loss} (simul.)			$1.43 \cdot 10^{-4}$	$1.87 \cdot 10^{-3}$		
Class 3 P_{loss} (analyt.)					$1.20 \cdot 10^{-3}$	$5.27 \cdot 10^{-3}$
Class 3 P_{loss} (simul.)					$5.39 \cdot 10^{-4}$	$2.77 \cdot 10^{-3}$

Fig.9 Cell loss rate for each class vs number of connections in progress

6. CONCLUSIONS

Some models to describe a bursty source have been presented and evaluated from an analytical point of view. A strategy for admission control, based only on the steady-state distribution of cell generation, has been introduced in an ATM node, where traffic is organized in service classes, characterized by performance requirements, and bandwidth partitions are dynamically allocated among classes by a Bandwidth Allocator.

Simulation results have been reported to verify the efficiency of the admission control and the accuracy of the cell loss rate approximation.

REFERENCES

1. J.P.Cosmas et al., "A Review of Voice, Data and Video Traffic Models for ATM", ETT, Vol.5, no. 2, Mar.-Apr. 1994, pp. 11-26.
2. K.Sriram, W.Whitt, "Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data", IEEE JSAC, Vol SAC-4 , no. 6, Sep. 1986, pp. 833-846.
3. H.Heffes, D.M.Lucantoni, "A Markov modulated charcterization of packetised voice and data traffic and related statistical multiplexer performance", IEEE JSAC, Vol SAC-4 ,no. 6, Sep. 1986, pp. 856-867.

4. A.Papoulis, Probability, Random Variables, and Stocastic Processes, McGraw-Hill International Editions, Singapore, 1984.

5. K.Sohraby, "On the Theory of General ON-OFF Sources With Applications in High Speed Networks", Proc. INFOCOM '93, San Francisco, CA, pp. 401-410.

6. T.Kamitake, T.Suda, "Evaluation of an admission control scheme for an ATM network considering fluctuations in cell loss rate", *Proc. Globecom '89*, Dallas, TX, Nov. 1989, pp. 1774-1780.

7. R.Bolla, F.Davoli, A.Lombardo, S.Palazzo, D.Panno, "Hierarchical dynamic control of multiple traffic classes in ATM networks", ETT, vol. 5, no. 6, Nov. 1994, pp. 747-755.

8. J.W.Roberts (Ed.), Performance evaluation and design of multiservice networks, COST 224 Final Report, 1991.

9. R.Bolla, F.Danovaro, F.Davoli, M.Marchese "An integrated dynamic resource allocation scheme for ATM networks", *Proc. Infocom '93*, San Francisco, CA, March 1993, vol. 3, pp. 1289-1297.

10. H.Saito, Teletraffic Technologies in ATM Networks, Artech House, Inc., Norwood, MA, 1994.

11. R.O.Onvural, Asynchronous Transfer Mode Networks: Performance Issues, Artech House, Inc., Norwood, MA, 1994.

12. D.Hong, T.Suda, "Congestion control and prevention in ATM networks", IEEE Network Mag., vol. 5, no. 4, July 1991, pp. 10-16.

13. R.Bolla, F.Davoli, M.Marchese, "A distributed routing and access control scheme for ATM networks", *Proc. IEEE Internat. Conf. on Commun., ICC '94*, New Orleans, LA, May 1994, vol. 1, pp. 44-50.

14. R.Bolla, F.Davoli, M.Marchese, "Performance of Hop-by-Hop distributed routing and resource allocation in an ATM network", *Proc. Internat. Conf. on Computer Communications and Networks, ICCCN'94*, San Francisco, CA, September 1994, pp. 235-241.