Transient characteristics of an M/M/∞ system applied to statistical multiplexing on an ATM link

Fabrice Guillemin, Gerardo Rubino, Bruno Sericola, Alain Simonian

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Transient characteristics of an $M/M/\infty$ system applied to statistical multiplexing on an ATM link

Fabrice Guillemin, Gerardo Rubino,

Bruno Sericola and Alain Simonian

N° 2386

Novembre 1994

PROGRAMME 1

apport de recherche

ISSN 0249-6399

Transient characteristics of an $M/M/\infty$ system applied to statistical multiplexing on an ATM link

Fabrice Guillemin*, Gerardo Rubino**,
Bruno Sericola** and Alain Simonian***

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués
Projet MODEL

Rapport de recherche n° 2386 — Novembre 1994 — 23 pages

Abstract: The Quality of Service (QoS) offered to telecommunication applications is related to the transient behavior of network elements in the case of congestion. In this paper, we analyze the variations of the bit rate process $\{\Lambda_t\}$ created on an ATM link transmitting data bursts in an open-loop statistical multiplexing scheme. This process is modeled by means of an $M/M/\infty$ queuing system. The performance of statistical multiplexing is then characterized by transient QoS parameters, namely the duration θ of a congestion where the process $\{\Lambda_t\}$ remains above the link transmission capacity C, the area V swept by that process and the number N of bursts arriving over this congestion period. The respective distributions of random variables θ, V and N are computed by means of a uniformization technique. The corresponding numerical algorithms are described in full detail. This approach allows us, in particular, to provide a simple proof for the so-called *local approximation* of process $\{\Lambda_t\}$, which has been previously introduced in the literature as a heuristic.

Key-words: ATM, statistical multiplexing, $M/M/\infty$ queuing system, transient characteristics.

(Résumé : tsvp)

*France Telecom, CNET Lannion-A, Route de Trégastel, 22300 Lannion
**{Gerardo.Rubino}{Bruno.Sericola}@irisa.fr
***France Telecom, CNET Paris A, 38-40 Rue du Général Leclerc, 92131 Issy-les-Mlx

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex (France)
Téléphone : (33) 99 84 71 00 – Télécopie : (33) 99 84 71 71

Caractéristiques transitoires d'un système $M/M/\infty$ appliquées au multiplexage statistique sur un lien ATM

Résumé : La qualité de service (QS) offerte à certaines applications dans le cadre des réseaux de télécommunication peuvent dépendre du comportement transitoire des éléments du réseau en cas de congestion. Dans cet article, on analyse l'évolution du processus $\{\Lambda_t\}$ décrivant le débit sur un lien de transmission ATM supportant des rafales de données dans un schéma de multiplexage statistique en boucle ouverte. Ce processus est modélisé par l'intermédiaire d'un système $M/M/\infty$. Les performances de ce multiplexage statistique sont caractérisées par des paramètres de QS transitoires, à savoir la durée θ d'une congestion pendant laquelle le processus $\{\Lambda_t\}$ est au dessus de la capacité C du lien de transmission, l'aire balayée V et le nombre N de rafales arrivant pendant une période de congestion. Les distributions respectives des variables θ, V et N sont calculées par une technique d'uniformisation. Les algorithmes numériques correspondants sont décrits en détail. Cette approche permet en particulier de donner une preuve simple de la-dite *approximation locale* pour le processus $\{\Lambda_t\}$, introduite précédemment dans la littérature en tant qu'heuristique.

Mots-clé : ATM, multiplexage statistique, file $M/M/\infty$, caractéristiques transitoires.

1 Introduction

In the emerging Broadband Integrated Services Digital Network based upon the Asynchronous Transfer Mode (ATM), a central issue is to characterize the Quality of Service (QoS) offered to users. While this task was relatively simple for classical Synchronous Transfer Mode networks such as telephone networks, the high flexibility of ATM to cope with a wide range of bit rates as well as bit rate variations within a given connection greatly complicates the characterization of relevant QoS parameters. This is all the more true as it seems very likely that statistical multiplexing will be performed by ATM networks in view of the potential benefit for a network operator to overbook the network links in order to optimize their utilization.

Several statistical multiplexing schemes have been proposed in the technical literature, notably open-loop statistical multiplexing where data bursts are directly transmitted onto the network. In such a case, overflow may occur on network links thus giving rise to network congestion. Open-loop multiplexing appears at first glance simplistic and rough when compared to other more sophisticated schemes such as the so-called "fast resource management" and "closed-loop rate based control" via explicit congestion notification. Open-loop multiplexing can nevertheless provide quite a fair and effective scheme as long as the "Connection Acceptance Control" function, which decides for the acceptance or rejection of a possible connection, is such that the probability of exceeding the transmission capacity of a link in the network is kept sufficiently small. The ratio *offered bit rate / network link transmission capacity* plays consequently a crucial role when designing a statistical multiplexing scheme.

The above condition, however, is not sufficient to characterize the QoS offered to a connection. In fact, the stationary congestion probability says very little about the transient phenomena involved in a congestion period, whereas the QoS of some applications is precisely related to the transient behavior of the network when congestion occurs. For instance, some data transfer protocols are very sensitive to consecutive cell losses within the network. As an example, if too many consecutive cells are lost during the transfer of a frame, upper layer protocols may request the retransmission of the whole frame. A QoS parameter must then be defined via the entire probability distribution of the number of consecutive lost cells in a transmission frame or equivalently, the maximum value of this number within a confidence interval.

Characterizing statistical multiplexing on an ATM link, that is, the way congestion occurs, consists in analyzing the excess of the process representing the bit rate of the aggregate traffic above the link transmission capacity. In this paper, we study the aggregation of identical bursty sources via a fluid flow approach [7] and we model the global bit rate process as the occupation process of an $M/M/\infty$ system. The Poisson assumption for burst arrivals may be justified if sources transmit bursts sparsely enough. At first glance, the $M/M/\infty$ model may appear very specific. A key advantage is however that it is mathematically tractable and yields numerical results which can then be used in more general situations as approximations. Moreover, recent studies have shown that it can be used as a quite accurate approximation in a Large Deviations [4] as well as a Heavy Traffic [5] setting when the link capacity is assumed very large compared to the individual source peak bit rate, whatever the distribution of bursts can be.

The QoS parameters introduced in this paper for characterizing statistical multiplexing on an ATM link are the duration θ of the congestion period where the occupation process remains above the link transmission capacity C, the area V swept by that process and the number N of bursts arriving over this congestion period, respectively. The distribution of these random variables may be analyzed via different techniques. Their mean values have been, in particular, studied in [11] and their Laplace transforms have been determined in [4]. In the present case, taking advantage from the Markov property of the analyzed process, the variables of interest can be defined as sojourn times for Markov processes with rewards. The distribution function of such variables, expressed in terms of the exponential of a matrix, can then be numerically computed by using the so-called "uniformization" technique. The algorithms used for such calculations are described in detail for each variable θ, V, and N. Furthermore, assuming the mean load u of the associated $M/M/\infty$ system is $u = \rho C$ with fixed $\rho \in (0,1)$, it is shown that $C\theta$, CV, and N weakly converge as $C \to \infty$ to the duration of a busy period duration, the area swept by the occupation process over a busy period and the number of customers served over a busy period of an $M/M/1$ queue with arrival rate ρ and unit mean service time, respectively. This result consequently justifies the so-called local approximation heuristic presented in [1] which could be invoked for deriving such convergence results for the mean values of θ, V, and N.

The organization of this paper is as follows. In Section 2, the modeling of the superposition of bursts on an ATM link is presented and the QoS parameters θ, V, and N are introduced. The algorithms for computing the distributions of θ, V, and N are given in Section 3, 4 and 5, respectively. The convergence of $C\theta$, CV, and N for increasing C with fixed $\rho = u/C$ is studied in Section 6. Concluding remarks are given in Section 7.

2 Modeling the superposition of bursts on an ATM link

Consider identical traffic sources transmitting ATM cells at peak cell rate taken as unity. Cells originated by each source are gathered in bursts and successive bursts are interspaced with silence periods corresponding to time intervals when the source is idle. All sources are statistically multiplexed on an ATM link with transmission capacity C. This link may represent the output link of a multiplexer or a switching element. Statistical multiplexing consists in overbooking the link in the sense that the cumulated peak rate Λ_t of active sources at an arbitrary instant t may be larger than the link capacity C. More information than can be carried may then arrive to the link. To absorb this volume of information in excess, the link is usually equipped with a buffer. In such a case, however, the buffer must be dimensioned in the so-called "burst scale" congestion domain [6] and earlier studies [8] have shown that queue dimensioning in such conditions may be unsafe due to the high sensitivity of the queue behavior to the input process characteristics (burst durations, variability, etc.).

To avoid such burst scale congestion phenomena, we assume in the following that the buffer attached to the link is intended to absorb cell scale congestion only in a peak cell rate allocation scheme. More specifically, suppose that purely periodic cell streams are multiplexed on the link. We can then totally ignore the occurrence of link overflow due to the burst scale fluctuations of the arrival process. The buffer is then dimensioned by using a simple $M/D/1/r$ queuing model and the queue capacity r is chosen so that the rejection probability is small (typically about 10^{-10}). The queue is thus intended to absorb the asynchronism of cell arrivals (namely, the random phase between the periodic cell arrival patterns generated by each source). For more general arrival processes (e.g., "bursty" traffic), a conservative assumption for estimating the cell loss due to link overflow is to consider that all information in excess with respect to the link capacity is lost when the instantaneous arrival rate is larger than the link rate.

In the present case, we assume that typical burst durations are much larger than cell inter-arrival times within a burst. This assumption allows us to ignore the discrete nature of the cell arrival process and to handle continuous bit rates only. This is the classical fluid flow approximation (see [7] for example) which has been shown fruitful for dimensioning an ATM multiplexer fed with the aggregation of variable bit rate sources. Furthermore, the number of sources is assumed sufficiently large and each source transmits bursts sparsely enough so that the burst arrival process can be considered as a Poisson process with intensity denoted by u. With the above assumptions, the statistical multiplexing problem addressed in this paper can be summarized as depicted in Figure 1. The instantaneous peak cell rate Λ_t created by the superposition of bursts is displayed in Figure 2.

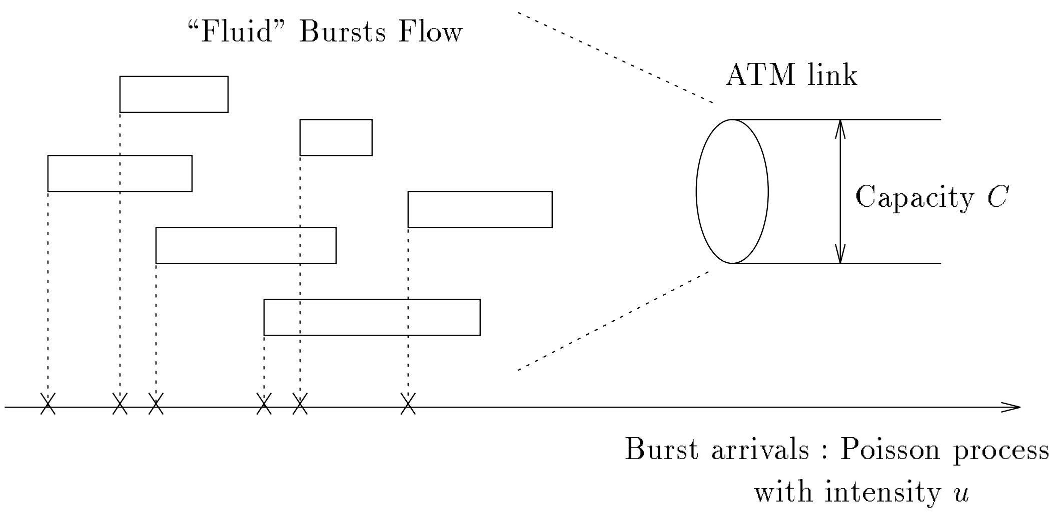

Figure 1: The fluid flow superposition of bursts on an ATM link.

Since all information in excess with respect to the link capacity is assumed to be lost when the instantaneous arrival rate is larger than the link rate, the performance of statistical multiplexing on an ATM link can be characterized by

- the duration θ of an excursion of process $\{\Lambda_t\}$ above level C. In the following, such an excursion will be referred to as congestion period as it corresponds to an overflow period for the link occupancy;
- the area V swept by process $\{\Lambda_t\}$ above level C during a congestion. In view of the fluid flow approximation, this area represents the quantity of information lost in a congestion period;
- the number N of bursts arriving in a congestion period. This quantity may be relevant for upper layer protocols which initiate burst retransmission as soon as a cell is lost during the transmission of a burst.

In the remainder of this paper, we assume that the burst duration is exponentially distributed with unit mean. Process $\{\Lambda_t\}$ can then be identified as the occupation process of an $M/M/\infty$ system with arrival rate u and unit service rate. Given the above assumptions and from the definition of performance parameters θ, V, and N, the analysis of statistical multiplexing on the ATM link leads to the analysis of such transient characteristics for an $M/M/\infty$ system. In the context of queuing theory, variables θ, V, and N correspond to the time there are more than C customers in the system, the volume of additional work entering the system and the number of consecutive customers arriving over such a period, respectively.

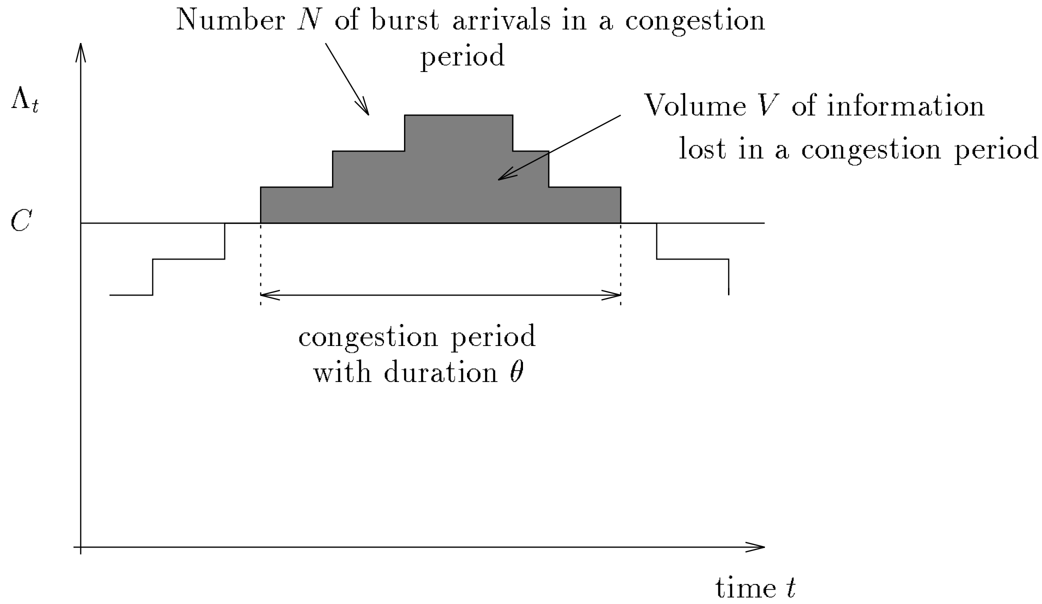

Figure 2: Instantaneous peak cell rate created by the superposition of bursts – the individual peak cell rate is taken as unity.

3 Distribution of congestion period θ

3.1 Preliminaries

The occupation process $\{\Lambda_t\}$ associated with the $M/M/\infty$ queuing system with mean arrival rate u and unit mean service rate is a continuous-time Markov process with state space $\{0, 1, 2, \ldots\}$, and infinitesimal generator $\mathcal{L}$ given by

$$\begin{aligned} &\mathcal{L}(0,0) = -u,\ \mathcal{L}(0,1) = u, \\ &\mathcal{L}(i,i-1) = i,\ \ \mathcal{L}(i,i) = -(u+i),\ \ \mathcal{L}(i,i+1) = u \end{aligned}$$

for $i \geq 1$ and $\mathcal{L}(i,j) = 0$ otherwise. Moreover, we assume that the link capacity C is an integer ≥ 1.

The congestion period θ is defined as the sojourn time of $\{\Lambda_t\}$ in the subspace $\{i \geq C + 1\}$. Using the Markov property, the successive sojourn times are independent and identically distributed and we can therefore consider the first one starting at time 0. The congestion period θ can also be viewed as the absorption time of the absorbing Markov process $\{\Lambda'_t\}$ with state space $\{0, 1, 2, 3, \ldots\}$, initial state 1, absorbing state 0 and infinitesimal generator $\mathcal{L}'$ whose non zero entries are

$$\mathcal{L}'(i,i-1) = C+i,\ \ \mathcal{L}'(i,i) = -(u+C+i),\ \ \mathcal{L}'(i,i+1) = u\ \text{ for } i \geq 1.$$

The congestion period θ can then be written as

$$\theta = \int_0^\infty \mathbb{1}_{\{\Lambda'_u \geq 1\}} du = \sum_{i=1}^{\infty} \int_0^\infty \mathbb{1}_{\{\Lambda'_u = i\}} du$$

where $\mathbb{1}_A$ denotes the indicator function of set A.

In the following, we denote by e_1, e, and $\mathbb{I}$ the row vector having the first entry equal to 1 and other entries equal to 0, the column vector with all entries equal to 1, and the identity matrix, respectively. Their dimension will be defined by the context. Zero (row or column) vectors are simply denoted by 0. We decompose the transition rate matrix $\mathcal{L}'$ with respect to the partition $\{0\}$, $\{1,2,3,\ldots\}$ of state space $\{0,1,2,\ldots\}$ as

$$\mathcal{L}' = \begin{pmatrix} 0 & 0 \\ -Qe & Q \end{pmatrix}$$

where Q is the submatrix of transition rates between states of $\{1,2,3,\ldots\}$. With this notation and from [10], the distribution of the congestion period θ can be written as

$$\Pr\{\theta > t\} = e_1 e^{Qt} e. \tag{3.1}$$

To compute the latter expression, however, the basic uniformization technique presented in [9] cannot be applied in the present case since we obviously have

$$\sup_{i\geq 1} |Q(i,i)| = \infty.$$

This can be circumvented by introducing random variables $\theta_n^{\inf}$ and $\theta_n^{\sup}$ associated with process $\{\Lambda'_t\}$ and defined by

$$\theta_n^{\inf} = \sum_{i=1}^{n} \int_0^\infty \mathbb{I}_{\{\Lambda'_u = i\}} du,$$

$$\theta_n^{\sup} = \sum_{i=1}^{n} \int_0^\infty \mathbb{I}_{\{\Lambda'_u = i\}} du + \sum_{i=n+1}^{\infty} i \int_0^\infty \mathbb{I}_{\{\Lambda'_u = i\}} du.$$

We clearly have $\theta_n^{\inf} \leq \theta \leq \theta_n^{\sup}$ for any integer n and it is straightforward that $\theta_n^{\inf} \to \theta$ a.s. when $n \to \infty$. Moreover, write $\theta_n^{\sup} = \theta_n^{\inf} + \tilde{\theta}_n^{\sup}$ with

$$\tilde{\theta}_n^{\sup} = \sum_{i=n+1}^{\infty} i \int_0^\infty \mathbb{I}_{\{\Lambda'_u = i\}} du$$

and note that the volume V of lost information over a congestion period, defined by the area swept by process $\{\Lambda_t\}$ above level C on $[0,\theta]$, can be expressed as

$$V = \int_0^\theta (\Lambda_u - C) \mathbb{I}_{\{\Lambda_u > C\}} du$$

or equivalently, using process $\{\Lambda'_t\}$,

$$V = \sum_{i=1}^{\infty} i \int_0^\infty \mathbb{I}_{\{\Lambda'_u = i\}} du. \tag{3.2}$$

As V has a finite mean [11], the latter series is a.s. converging. This entails that its remainder to order n a.s. tends to 0 as n tends to infinity, or equivalently, $\tilde{\theta}_n^{\sup} \to 0$ a.s. when $n \to \infty$. It follows that $\theta_n^{\sup} \to \theta$ a.s. when $n \to \infty$. As a consequence, the distribution of θ can be approximated by those of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ for large enough n.

3.2 Distribution of $\theta_n^{\inf}$

Define the subsets U_n and D_n by

$$U_n = \{1 \le i \le n\} \quad \text{and} \quad D_n = \{i \ge n+1\}$$

for every $n \ge 1$. With this notation, $\theta_n^{\inf}$ is the total time spent by process $\{\Lambda'_t\}$ in the subset U_n until absorption. Now, decompose the transition rate matrix Q with respect to the subsets U_n and D_n of $\{1, 2, 3, \ldots\}$ as

$$Q = \begin{pmatrix} Q_{U_n} & Q_{U_n,D_n} \\ Q_{D_n,U_n} & Q_{D_n} \end{pmatrix}.$$

The submatrix Q_{U_n} (resp. Q_{U_n,D_n}, Q_{D_n,U_n}, Q_{D_n}) corresponds to transitions inside U_n (resp. from U_n to D_n, from D_n to U_n, inside D_n).

Using then the result given in [3] for distributions of sojourn times of Markov reward processes, with rewards set to 1 for states in U_n and to 0 for all other states, the distribution of $\theta_n^{\inf}$ can be expressed as

$$\Pr\{\theta_n^{\inf} > t\} = e_1 e^{Q_n^{\inf} t} e$$

where

$$Q_n^{\inf} = Q_{U_n} - Q_{U_n,D_n} Q_{D_n}^{-1} Q_{D_n,U_n}.$$

It can be easily shown that the n-dimensional square matrix $Q_{U_n,D_n} Q_{D_n}^{-1} Q_{D_n,U_n}$ has only one non-zero entry, corresponding to transition (n, n) and equal to u. The non-zero entries of matrix $Q_n^{\inf}$ are then

$$\begin{cases} Q_n^{\inf}(i, i-1) = C + i, \ \ Q_n^{\inf}(i, i) = -(u + C + i), \ \ Q_n^{\inf}(i, i+1) = u \text{ for } 1 \le i \le n-1, \\ \\ Q_n^{\inf}(n, n-1) = C + n, \ \ Q_n^{\inf}(n, n) = -(C + n). \end{cases}$$

To compute the distribution of $\theta_n^{\inf}$, we can now proceed by uniformization [9]. Define $\nu_n = u + C + n$ and introduce the matrix

$$P_n^{\inf} = \mathbb{I} + \frac{Q_n^{\inf}}{\nu_n}.$$

This definition relation along with the above expression of $\Pr\{\theta_n^{\inf} > t\}$ in terms of matrix $Q_n^{\inf}$ enable us to write

$$\Pr\{\theta_n^{\inf} > t\} = \sum_{k=0}^{\infty} e^{-\nu_n t} \frac{(\nu_n t)^k}{k!} e_1 (P_n^{\inf})^k e$$

for any fixed t. Noting that $P_n^{\inf}$ is a substochastic matrix, we have

$$\Pr\{\theta_n^{\inf} > t\} = \sum_{k=0}^{K} e^{-\nu_n t} \frac{(\nu_n t)^k}{k!} e_1 (P_n^{\inf})^k e + \varepsilon_K \tag{3.3}$$

where ε_K satisfies

$$\varepsilon_K = \sum_{k=K+1}^{\infty} e^{-\nu_n t} \frac{(\nu_n t)^k}{k!} e_1 (P_n^{\inf})^k e \le \sum_{k=K+1}^{\infty} e^{-\nu_n t} \frac{(\nu_n t)^k}{k!} = 1 - \sum_{k=0}^{K} e^{-\nu_n t} \frac{(\nu_n t)^k}{k!}.$$

The truncation step K can then be computed for any given error tolerance ε as

$$K = \min\left\{ h \geq 0 \;\middle|\; \sum_{j=0}^{h} e^{-\nu_n t}\frac{(\nu_n t)^j}{j!} \geq 1-\varepsilon \right\}. \tag{3.4}$$

The only remaining problem for calculating $\Pr\{\theta_n^{\inf} > t\}$ is to compute the sequence $w_k^{\inf}$, $0 \leq k \leq K$, defined by $w_k^{\inf} = e_1 (P_n^{\inf})^k e$. If $z_k^{\inf}$ is the column vector with length n given by $z_k^{\inf} = (P_n^{\inf})^k e$, $w_k^{\inf}$ is the first component of vector $z_k^{\inf}$, that is, $w_k^{\inf} = z_k^{\inf}(1)$. Introduce then

$$\begin{cases} p_i = \dfrac{u}{\nu_n} = P_n^{\inf}(i,i+1) \text{ for } 1 \leq i \leq n-1, \\[2ex] r_i = 1 - \dfrac{u+C+i}{\nu_n} = P_n^{\inf}(i,i) \text{ for } 1 \leq i \leq n-1, \\[2ex] r_n = 1 - \dfrac{C+n}{\nu_n} = P_n^{\inf}(n,n), \\[2ex] q_i = \dfrac{C+i}{\nu_n} = P_n^{\inf}(i,i-1) \text{ for } 2 \leq i \leq n. \end{cases} \tag{3.5}$$

Using the relation $z_k^{\inf} = P_n^{\inf} z_{k-1}^{\inf}$, the components of vector $z_k^{\inf}$ can be recursively computed as

$$\begin{cases} z_k^{\inf}(1) = r_1 z_{k-1}^{\inf}(1) + p_1 z_{k-1}^{\inf}(2), \\[2ex] z_k^{\inf}(i) = q_i z_{k-1}^{\inf}(i-1) + r_i z_{k-1}^{\inf}(i) + p_i z_{k-1}^{\inf}(i+1) \text{ for } 2 \leq i \leq n-1, \\[2ex] z_k^{\inf}(n) = q_n z_{k-1}^{\inf}(n-1) + r_n z_{k-1}^{\inf}(n) \end{cases} \tag{3.6}$$

with initial condition $z_0^{\inf} = e$.

3.3 Distribution of $\theta_n^{\sup}$

For every $n \geq 1$, define the infinite diagonal reward matrix R_n with entries

$$R_n(i,i) = \begin{cases} 1 & \text{if } 1 \leq i \leq n, \\ i & \text{if } i \geq n+1, \end{cases}$$

and $R(i,j) = 0$ otherwise. With this reward structure, $\theta_n^{\sup}$ is the reward accumulated by process $\{\Lambda'_t\}$ until absorption. Using again the result given in [3] for sojourn times of Markov reward processes, the distribution of $\theta_n^{\sup}$ can be expressed as $\Pr\{\theta_n^{\sup} > t\} = e_1 e^{Q_n^{\sup} t} e$ where $Q_n^{\sup} = R_n^{-1} Q$. This distribution can be evaluated by uniformization since

$$\sup_{i \geq 0} |Q_n^{\sup}(i,i)| < u + C + n = \nu_n < \infty.$$

Proceeding as in Section 3.2, define the matrix

$$P_n^{\sup} = \mathbb{I} + \frac{Q_n^{\sup}}{\nu_n}.$$

With this definition relation, we can write

$$\Pr\{\theta_n^{\sup} > t\} = \sum_{k=0}^{\infty} e^{-\nu_n t}\frac{(\nu_n t)^k}{k!} e_1 (P_n^{\sup})^k e.$$

Matrix $P_n^{\rm sup}$ being substochastic, we have

$$\Pr\{\theta_n^{\rm sup} > t\} = \sum_{k=0}^{K} e^{-\nu_n t}\frac{(\nu_n t)^k}{k!} e_1 (P_n^{\rm sup})^k e + \varepsilon'_K, \tag{3.7}$$

where $\varepsilon'_K \leq \varepsilon$ if K is chosen for a given error tolerance as in relation (3.4). Successively introducing $w_k^{\rm sup} = e_1(P_n^{\rm sup})^k e$ for $0 \leq k \leq K$ and the column vector $z_k^{\rm sup} = (P_n^{\rm sup})^k e$ with infinite length, $w_k^{\rm sup}$ is the first component of vector $z_k^{\rm sup}$, that is $w_k^{\rm sup} = z_k^{\rm sup}(1)$. As above, we define

$$\begin{cases} p'_i = \dfrac{u}{\nu_n} = P_n^{\rm sup}(i, i+1) \text{ for } 1 \leq i \leq n, \\[2ex] p'_i = \dfrac{u}{i\nu_n} \text{ for } i \geq n+1, \\[2ex] r'_i = 1 - \dfrac{u + C + i}{\nu_n} = P_n^{\rm sup}(i, i) \text{ for } 1 \leq i \leq n, \\[2ex] r'_i = 1 - \dfrac{u + C + i}{i\nu_n} = P_n^{\rm sup}(i, i) \text{ for } i \geq n+1, \\[2ex] q'_i = \dfrac{C + i}{\nu_n} = P_n^{\rm sup}(i, i-1) \text{ for } 2 \leq i \leq n, \\[2ex] q'_i = \dfrac{C + i}{i\nu_n} = P_n^{\rm sup}(i, i-1) \text{ for } i \geq n+1. \end{cases} \tag{3.8}$$

Using the relation $z_k^{\rm sup} = P_n^{\rm sup} z_{k-1}^{\rm sup}$ and noting that, for every $k \geq 0$, $z_k^{\rm sup}(i) = 1$ if $i \geq k+1$, the components of vector $z_k^{\rm sup}$ can be computed by using the recurrence relations

$$\begin{cases} z_k^{\rm sup}(1) = r'_1 z_{k-1}^{\rm sup}(1) + p'_1 z_{k-1}^{\rm sup}(2), \\[1ex] z_k^{\rm sup}(i) = q'_i z_{k-1}^{\rm sup}(i-1) + r'_i z_{k-1}^{\rm sup}(i) + p'_i z_{k-1}^{\rm sup}(i+1) \text{ for } 2 \leq i \leq k-2, \\[1ex] z_k^{\rm sup}(k-1) = q'_{k-1} z_{k-1}^{\rm sup}(k-2) + r'_{k-1} z_{k-1}^{\rm sup}(k-1) + p'_{k-1}, \\[1ex] z_k^{\rm sup}(k) = q'_k z_{k-1}^{\rm sup}(k-1) + r'_k + p'_k. \end{cases} \tag{3.9}$$

3.4 Algorithms for the Computation of the Distributions of $\theta_n^{\rm inf}$ and $\theta_n^{\rm sup}$

Using the recurrence relations given in Sections 3.2 and 3.3, the distributions of $\theta_n^{\rm inf}$ and $\theta_n^{\rm sup}$ can be evaluated at some points, say, $t_1, \ldots, t_M$ for some integer $M \geq 1$. For each of them, the input parameters are the offered load u, the link capacity C, the truncation error ε, level $n \geq C$, and the points $t_1, \ldots, t_M$. The truncation level K, which depends on n, is computed by using eq. (3.4) for $t = t_M$.

Remark : The truncation level K defined in 3.4 is in fact a function of t, say $K(t)$. In order to compute $\Pr\{\theta_n^{\rm inf} > t_m\}$ for each m, we need the first $K(t_m)$ values of the sequence $w_k^{\rm inf}$, but these values are independent of the value of t_m. It follows that for a fixed value of ε, if $t_1 < \cdots < t_M$ then we have $K(t_1) < \cdots < K(t_M)$ and so it suffices to store the first $K(t_M)$

values of the sequence $w_k^{\inf}$ in order to get $\Pr\{\theta_n^{\inf} > t_m\}$ for each m. More precisely, we have for each m,

$$\begin{aligned}\Pr\{\theta_n^{\inf} > t_m\} &= \sum_{k=0}^{\infty} e^{-\nu_n t_m}\frac{(\nu_n t_m)^k}{k!} e_1 (P_n^{\inf})^k e \\ &= \sum_{k=0}^{K(t_M)} e^{-\nu_n t_m}\frac{(\nu_n t_m)^k}{k!} e_1 (P_n^{\inf})^k e + \sum_{k=K(t_M)+1}^{\infty} e^{-\nu_n t_m}\frac{(\nu_n t_m)^k}{k!} e_1 (P_n^{\inf})^k e\end{aligned}$$

and, moreover

$$\sum_{k=K(t_M)+1}^{\infty} e^{-\nu_n t_m}\frac{(\nu_n t_m)^k}{k!} e_1 (P_n^{\inf})^k e \le \sum_{k=K(t_m)+1}^{\infty} e^{-\nu_n t_m}\frac{(\nu_n t_m)^k}{k!} e_1 (P_n^{\inf})^k e \le \varepsilon.$$

This remark will also be used next for the computation of the distributions of random variables $\theta_n^{\inf}$ and V. ■

The evaluation of the distribution of $\theta_n^{\inf}$ (resp. $\theta_n^{\sup}$) consists in computing the coefficients p_i, r_i, and q_i (resp. p_i', r_i', and q_i') defined by eq.(3.5) (resp. eq. (3.8)) and in using the recurrence relation (3.6) (resp. (3.9)). This procedure yields coefficients $w_k^{\inf}$ (resp. $w_k^{\sup}$), which are used in eq. (3.3) (resp. eq. (3.7)) to compute the distribution of $\theta^{\inf}$ (resp. $\theta^{\sup}$) at point t_m for $m = 1, \ldots, M$. The pseudocode of the algorithms are given in Tables 1 and 2, respectively.

input : u, C, n, ε, $t_1 < \cdots < t_M$
output : $\Pr\{\theta_n^{\inf} > t_1\}, \ldots, \Pr\{\theta_n^{\inf} > t_M\}$
$\nu = u + C + n$
Compute $K = \min\left\{ h \ge 0 \mid \sum_{j=0}^{h} e^{-\nu t_M}\frac{(\nu t_M)^j}{j!} \ge 1-\varepsilon \right\}$, % using for instance [2]
for $i = 1$ **to** $n-1$ **do** $p_i = u/\nu$ **endfor**
for $i = 1$ **to** $n-1$ **do** $r_i = (n-i)/\nu$ **endfor**
$r_n = u/\nu$.
for $i = 2$ **to** n **do** $q_i = (C+i)/\nu$ **endfor**
$z^{\inf}$, $v^{\inf}$: array of dimension n, $z^{\inf} = (z_1^{\inf}, \ldots, z_n^{\inf})$, $v^{\inf} = (v_1^{\inf}, \ldots, v_n^{\inf})$
% $v^{\inf}$ is a dummy variable
$w^{\inf}$: array of dimension $K+1$, $w^{\inf} = (w_0^{\inf}, \ldots, w_K^{\inf})$
for $i = 1$ **to** n **do** $z_i^{\inf} = 1$ **endfor**
$w_0^{\inf} = 1$
for $k = 1$ **to** K **do**
  $v_1^{\inf} = r_1 z_1^{\inf} + p_1 z_2^{\inf}$
  for $i = 2$ **to** $n-1$ **do** $v_i^{\inf} = q_i z_{i-1}^{\inf} + r_i z_i^{\inf} + p_i z_{i+1}^{\inf}$ **endfor**
  $v_n^{\inf} = q_n z_{n-1}^{\inf} + r_n z_n^{\inf}$
  for $i = 1$ **to** n **do** $z_i^{\inf} = v_i$ **endfor**
  $w_k^{\inf} = z_1^{\inf}$
endfor
for $m = 1$ **to** M **do** $\Pr\{\theta_n^{\inf} > t_m\} = \sum_{k=0}^{K} e^{-\nu t_m}\frac{(\nu t_m)^k}{k!} w_k^{\inf}$ **endfor**

Table 1: Algorithm for the computation of the distribution of $\theta_n^{\inf}$.

input : u, C, n, ε, $t_1 < \cdots < t_M$
output : $\Pr\{\theta_n^{\sup} > t_1\}, \ldots, \Pr\{\theta_n^{\sup} > t_M\}$
$\nu = u + C + n$
Compute $K = \min\left\{ h \geq 0 \;\middle|\; \sum_{j=0}^{h} e^{-\nu t_M} \frac{(\nu t_M)^j}{j!} \geq 1 - \varepsilon \right\}$, % using for instance [2]
for $i = 1$ **to** $\min(n, K)$ **do**
  $p'_i = u/\nu$
  $r'_i = (n-i)/\nu$
  $q'_i = (C+i)/\nu$
endfor
for $i = n+1$ **to** K **do**
  $p'_i = u/(i\nu)$
  $r'_i = ((i-1)(u+C) + (n-1)i)/(i\nu)$
  $q'_i = (C+i)/(i\nu)$
endfor
$z^{\sup}$, $v^{\sup}$: array of dimension K, $z^{\sup} = (z_1^{\sup}, \ldots, z_K^{\sup})$, $v^{\sup} = (v_1^{\sup}, \ldots, v_K^{\sup})$
% $v^{\sup}$ is a dummy variable
$w^{\sup}$: array of dimension $K+1$, $w^{\sup} = (w_0^{\sup}, \ldots, w_K^{\sup})$
for $i = 1$ **to** n **do** $z_i^{\sup} = 1$ **endfor**
$w_0^{\sup} = 1$
for $k = 1$ **to** K **do**
  $v_1^{\sup} = r'_1 z_1^{\sup} + p'_1 z_2^{\sup}$
  for $i = 2$ **to** $k-2$ **do** $v_i^{\sup} = q'_i z_{i-1}^{\sup} + r'_i z_i^{\sup} + p'_i z_{i+1}^{\sup}$ **endfor**
  $v_{k-1}^{\sup} = q'_{k-1} z_{k-2}^{\sup} + r'_{k-1} z_{k-1}^{\sup} + p'_{k-1}$
  $v_k^{\sup} = q'_k z_{k-1}^{\sup} + r'_k + p'_k$
  for $i = 1$ **to** k **do** $z_i^{\sup} = v_i^{\sup}$ **endfor**
  $w_k^{\sup} = z_1^{\sup}$
endfor
for $m = 1$ **to** M **do** $\Pr\{\theta_n^{\sup} > t_m\} = \sum_{k=0}^{K} e^{-\nu t_m} \frac{(\nu t_m)^k}{k!} w_k^{\sup}$ **endfor**

Table 2: Algorithm for the computation of the distribution of $\theta_n^{\sup}$.

To evaluate the distribution of θ, we compute the distributions of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ for a sufficiently large value of n so that the difference is less than a given error tolerance. Since the distribution of $\theta_n^{\inf}$ (resp. $\theta_n^{\sup}$) is computed with an error ε_K (resp. ε'_K) less than ε as shown by relation (3.4), it follows that the difference between the distributions of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ is less than 2ε. Hence, the difference between the distributions of $\theta_n^{\inf}$ and θ, as well as θ and $\theta_n^{\sup}$, will be itself less than 2ε.

Figures 3 and 4 depict the distribution of θ for different values of C. Mean arrival rate u has been set to 240, mean service rate to 1 and error tolerance ε to 10^{-5}. The value of n has been fixed to $n = 500$ so that the difference between the distributions of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ is less than 10^{-7}.

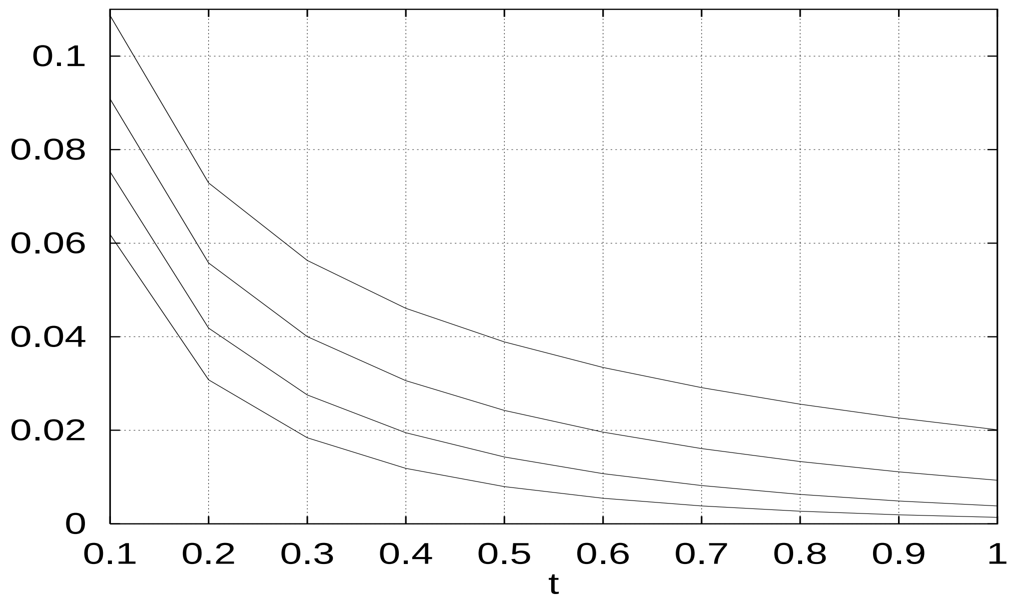

Figure 3: Distribution of the congestion period θ for $u = 240$, $\varepsilon = 10^{-5}$ and different values of C – from top to bottom : $\Pr\{\theta > t\}$ for $C = 240, 250, 260$ and 270.

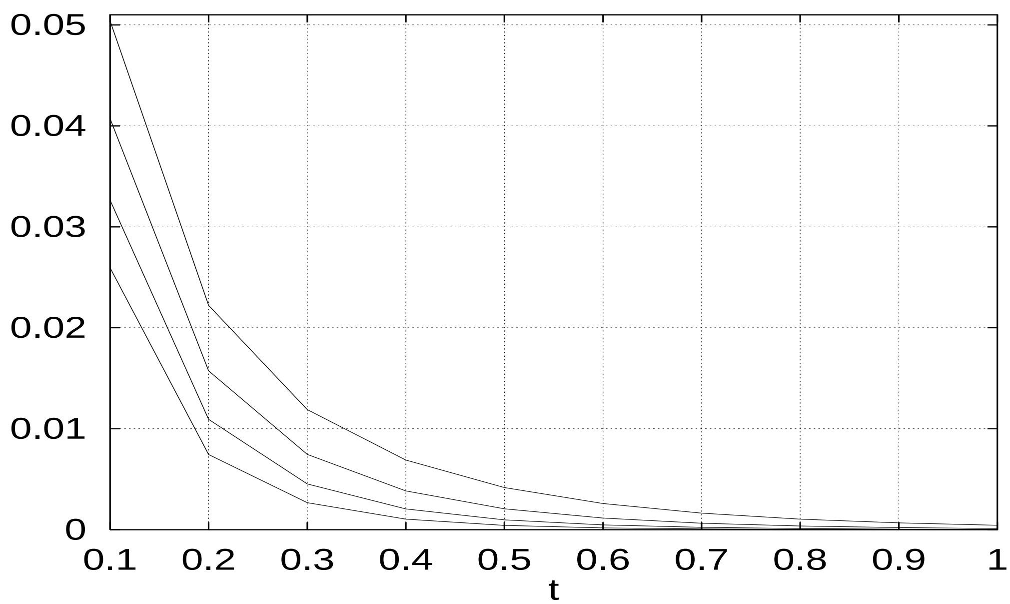

Figure 4: Distribution of the congestion period θ for $u = 240$, $\varepsilon = 10^{-5}$ and different values of C – from top to bottom : $\Pr\{\theta > t\}$ for $C = 280, 290, 300$ and 310.

4 Distribution of Volume of Lost Information V

4.1 Uniformization Procedure

As expressed in (3.2), random variable V can be interpreted as a performability measure and using the result given in [3], with reward set to value i for state i, we obtain

$$\Pr\{V > t\} = e_1 e^{Mt} e \tag{4.1}$$

where

$$M = R^{-1}Q,$$

R being the diagonal reward matrix with entries $R(i,i) = i$, $i \geq 1$. In contrast with the computation of the distribution of θ processed in the previous section by means of intermediate variables, the distribution of V can here be directly evaluated by uniformization since

$$\nu = \sup_{i \geq 0} |M(i,i)| = u + C + 1 < \infty. \tag{4.2}$$

Defining the substochastic matrix $P = \mathbb{I} + \frac{M}{\nu}$, we have $\Pr\{V > t\} = \sum_{k=0}^{\infty} e^{-\nu t} \frac{(\nu t)^k}{k!} e_1 P^k e$. As above, we define the truncation step K by writing

$$\Pr\{V > t\} = \sum_{k=0}^{K} e^{-\nu t} \frac{(\nu t)^k}{k!} e_1 P^k e + \varepsilon''_K \tag{4.3}$$

where K is chosen so that $\varepsilon''_K \leq \varepsilon$, that is

$$K = \min\left\{ h \geq 0 \;\middle|\; \sum_{j=0}^{h} e^{-\nu t} \frac{(\nu t)^j}{j!} \geq 1 - \varepsilon \right\}. \tag{4.4}$$

Successively denoting by w_k^V, $0 \leq k \leq K$, the numbers $w_k^V = e_1 P^k e$ and by z_k^V the column vector $z_k^V = P^k e$ with infinite length, w_k^V is the first component of vector z_k^V, that is, $w_k^V = z_k^V(1)$. As previously, we define

$$\begin{cases} p''_i = \dfrac{u}{i\nu} = P(i, i+1) \text{ for } i \geq 1, \\[2ex] r''_i = \dfrac{(i-1)(u+C)}{i\nu} = P(i,i) \text{ for } i \geq 1, \\[2ex] q''_i = \dfrac{C+i}{i\nu} = P(i, i-1) \text{ for } i \geq 2. \end{cases}$$

Using the relation $z_k^V = P z_{k-1}^V$ and noting that, for every $k \geq 0$, $z_k^V(i) = 1$ if $i \geq k+1$, the components of vector z_k^V can be recursively computed as (note that $r''_1 = 0$)

$$\begin{cases} z_k^V(1) = p''_1 z_{k-1}^V(2), \\[2ex] z_k^V(i) = q''_i z_{k-1}^V(i-1) + r''_i z_{k-1}^V(i) + p''_i z_{k-1}^V(i+1) \text{ for } 2 \leq i \leq k-2, \\[2ex] z_k^V(k-1) = q''_{k-1} z_{k-1}^V(k-2) + r''_{k-1} z_{k-1}^V(k-1) + p''_{k-1}, \\[2ex] z_k^V(k) = q''_k z_{k-1}^V(k-1) + r''_k + p''_k. \end{cases}$$

4.2 Algorithm for the Computation of the Distribution of V

The algorithm used to compute the distribution of random variable V is similar to those used for $\theta^{\inf}$ and $\theta^{\sup}$. The pseudo-code of the algorithm is given in Table 3. This algorithm directly gives the distribution of V with a closing error ε.

input : u, C, ε, $t_1 < \cdots < t_M$
output : $\Pr\{V > t_1\}, \ldots, \Pr\{V > t_M\}$
$\nu = u + C + 1$
Compute $K = \min\left\{ h \geq 0 \;\middle|\; \sum_{j=0}^{h} e^{-\nu t_M} \frac{(\nu t_M)^j}{j!} \geq 1 - \varepsilon \right\}$, % using for instance [2]
for $i = 1$ **to** K **do**
    $p''_i = u/(i\nu)$
    $r''_i = (i-1)(u+C)/(i\nu)$
    $q''_i = (C+i)/(i\nu)$
endfor
z^V: array of dimension $K+1$, $z^V = (z^V_1, \ldots, z^V_{K+1})$,
v^V: array of dimension K, $v^V = (v^V_1, \ldots, v^V_K)$
w^V: array of dimension $K+1$, $w^V = (w^V_0, \ldots, w^V_K)$
for $i = 1$ **to** $K+1$ **do** $z^V_i = 1$ **endfor**
$w^V_0 = 1$
for $k = 1$ **to** K **do**
    $v^V_1 = p''_1 z^V_2$
    for $i = 2$ **to** k **do** $v^V_i = q''_i z^V_{i-1} + r''_i z^V_i + p''_i z^V_{i+1}$ **endfor**
    for $i = 1$ **to** k **do** $z^V_i = v^V_i$ **endfor**
    $w^V_k = z^V_1$
endfor
for $m = 1$ **to** M **do** $\Pr\{\theta^{\sup}_n > t_m\} = \sum_{k=0}^{K} e^{-\nu t_m} \frac{(\nu t_m)^k}{k!} w^V_k$ **endfor**

Table 3: Algorithm for the computation of the distribution of V.

For given values of the mean arrival rate u and the closing error ε, the algorithm given in Table 3 has been used to compute the distribution of V for different values of level C. Fig. 5 and Fig. 6 depict the distribution of V for $u = 240$, $\varepsilon = 10^{-5}$ and different values of C.

5 Distribution of Number N of Customers Arriving in a Congestion Period

The number N of customers arriving in a congestion period is distributed as the number N_A of increasing steps of process $\{\Lambda'_t\}$ starting from state 1. Defining $P_i(n)$ as

$$P_i(n) = \Pr\{N_A = n / \Lambda'_0 = i\},$$

we want to compute $\Pr\{N = n\} = P_1(n)$ for $n \in \mathbb{N}$.

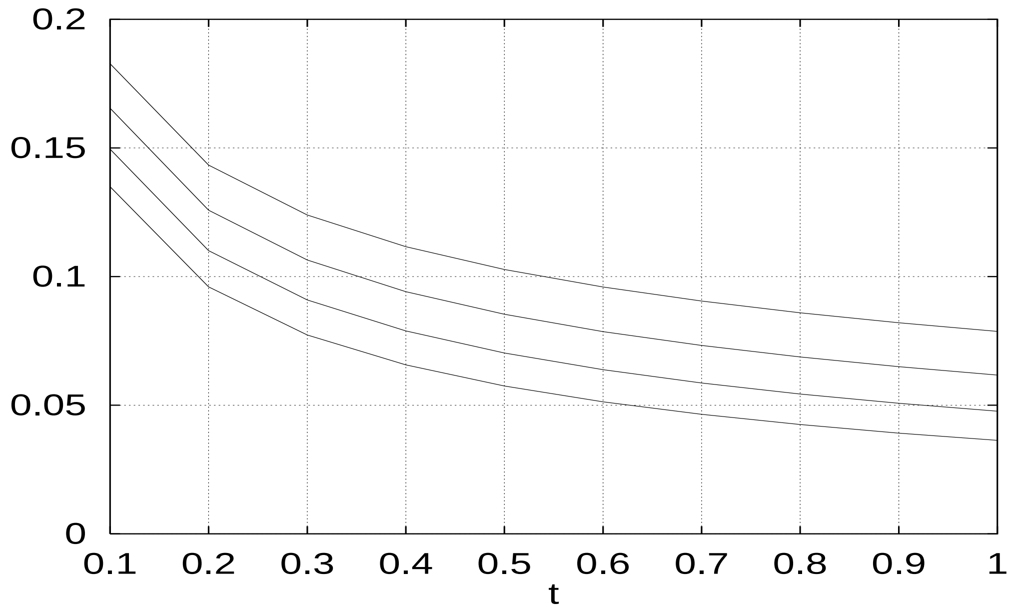

Figure 5: Distribution of the volume V of lost information for $u = 240$, $\varepsilon = 10^{-5}$ and different values of C – from top to bottom : $\Pr\{V > t\}$ for $C = 240, 250, 260$ and 270.

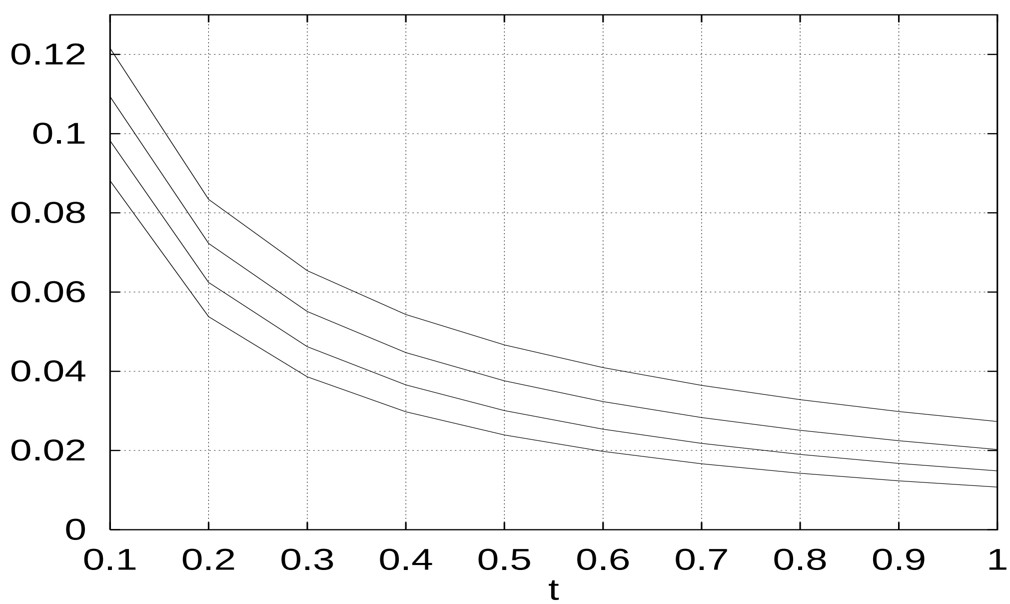

Figure 6: Distribution of the volume V of lost information for $u = 240$, $\varepsilon = 10^{-5}$ and different values of C – from top to bottom : $\Pr\{V > t\}$ for $C = 280, 290, 300$ and 310.

Let us introduce the following notation:

$$\begin{cases} p_i''' = \dfrac{u}{u+C+i} \text{ for } i \geq 1, \\ \\ q_i''' = \dfrac{C+i}{u+C+i} \text{ for } i \geq 1. \end{cases} \tag{5.1}$$

The numbers $P_i(n)$ verify the recurrence relations

$$P_i(n) = p_i''' P_{i+1}(n-1) + q_i''' P_{i-1}(n),$$

for $i \geq 1$ and $n \geq 0$, with initial condition $P_0(n) = \mathbb{1}_{\{n=0\}}$ (note that $P_i(-1) = 0$ since $N \geq 0$).

These recurrence relations can also be written as

$$\begin{cases} P_{n+1}(0) = q_{n+1}''' P_n(0), \\ \\ P_i(n-i+1) = p_i''' P_{i+1}(n-i) + q_i''' P_{i-1}(n-i+1) \text{ for } 2 \leq i \leq n-1, \\ \\ P_1(n) = p_1''' P_2(n-1) \end{cases} \tag{5.2}$$

for every $n \geq 1$ and with initial condition $P_1(0) = q_1'''$. These relations provide an algorithmic scheme to compute $P_1(n)$, whose pseudo-code is given in Table 4.

input : u, C, M
output : $\Pr\{N = n\}$, $n = 0, \ldots, M$
for $i = 1$ **to** M **do**
    $p_i''' = u/(u+C+i)$
    $q_i''' = (C+i)/(u+C+i)$
endfor
$q_{M+1}''' = (C+M+1)/(u+C+M+1)$
v^N : array of dimension $M+1$, $v^N = (v_0^N, \ldots, v_M^N)$
$v_0^N = q_1'''$
for $n = 1$ **to** M **do**
    $v_0^N = q_{n+1}''' v_0^N$
    for $i = 1$ **to** $n-1$ **do** $v_i^N = p_{n-i+1}''' v_{i-1}^N + q_{n-i+1}''' v_i^N$ **endfor**
    $v_n^N = p_1''' v_{n-1}^N$
    $\Pr\{N = n\} = v_n^N$
endfor

Table 4: Algorithm for the computation of the distribution of N.

6 Limit Distributions for $C\theta$, CV and N

In this section, we study the limit behavior of random variables $C\theta$, CV and N when C tends to infinity and $\rho = u/C \in (0,1)$ is fixed. Let $\theta_{M/M/1}$, $V_{M/M/1}$ and $N_{M/M/1}$ be the busy period, the area swept by the occupation process and the number of customers arriving in a busy period of an M/M/1 queue with arrival rate ρ and service rate equal to 1, respectively.

Theorem 6.1 *When C tends to infinity and for fixed $\rho = u/C \in (0,1)$, we have*

$$\Pr\{C\theta > t\} \longrightarrow \Pr\{\theta_{M/M/1} > t\},$$

$$\Pr\{CV > t\} \longrightarrow \Pr\{V_{M/M/1} > t\},$$

$$\Pr\{N = n\} \longrightarrow \Pr\{N_{M/M/1} = n\}.$$

Proof. Using the results of [10] and [3], the distribution of variables $\theta_{M/M/1}$ and $V_{M/M/1}$ can be expressed as

$$\Pr\{\theta_{M/M/1} > t\} = e_1 e^{Ht} e \text{ and } \Pr\{V_{M/M/1} > t\} = e_1 e^{R^{-1}Ht} e$$

where H is the infinitesimal generator associated with the occupation process of the above defined $M/M/1$ queue. When C tends to infinity and for fixed $\rho = u/C$, we have

$$\frac{Q}{C} \longrightarrow H$$

with Q introduced in (3.1). It follows that, by using relations (3.1) and (4.1) and by continuity,

$$\Pr\{C\theta > t\} \longrightarrow e_1 e^{Ht} e,$$

and

$$\Pr\{CV > t\} \longrightarrow e_1 e^{R^{-1}Ht} e,$$

which completes the proof for variables $C\theta$ and CV. When C tends to infinity and for fixed $\rho = u/C$, we have from relation (5.1) for every i,

$$p_i''' \longrightarrow p = \frac{\rho}{\rho + 1} \text{ and } q_i''' \longrightarrow q = \frac{1}{\rho + 1}.$$

If we replace all p_i''''s with p and all q_i''''s by q in recurrence relation (5.2), we obtain a recurrence relation whose solution $P_1(n)$ is equal to $\Pr\{N_{M/M/1} = n\}$. In this case, this recurrence relation can be solved to give

$$\Pr\{N_{M/M/1} = n\} = \frac{\binom{2n}{n}}{n+1} p^n q^{n+1}$$

where $\dfrac{\binom{2n}{n}}{n+1}$, $n \geq 0$, are known as Catalan's numbers. ■

For different values of level C, the distributions of random variables $C\theta$, CV, N and their limit distributions are depicted in Figure 7 and Figure 8 for $\rho = 1$ and in Figure 9 and Figure 10 for $\rho = 0.8$. As far as means of θ, V and N are concerned, the latter theorem confirms the so-called "local approximation" of a process near a boundary which has been introduced by Aldous [1] as a heuristic. In the present case, this heuristic amounts to replacing the local behavior of process $\{\Lambda_t\}$ near boundary C by a simpler process, namely the $M/M/1$ occupation process, whose characteristics are simply derived from those of $\{\Lambda_t\}$ by linearisation near state C. Here we control that such an approximation scheme is also applicable to the distributions of such transient variables.

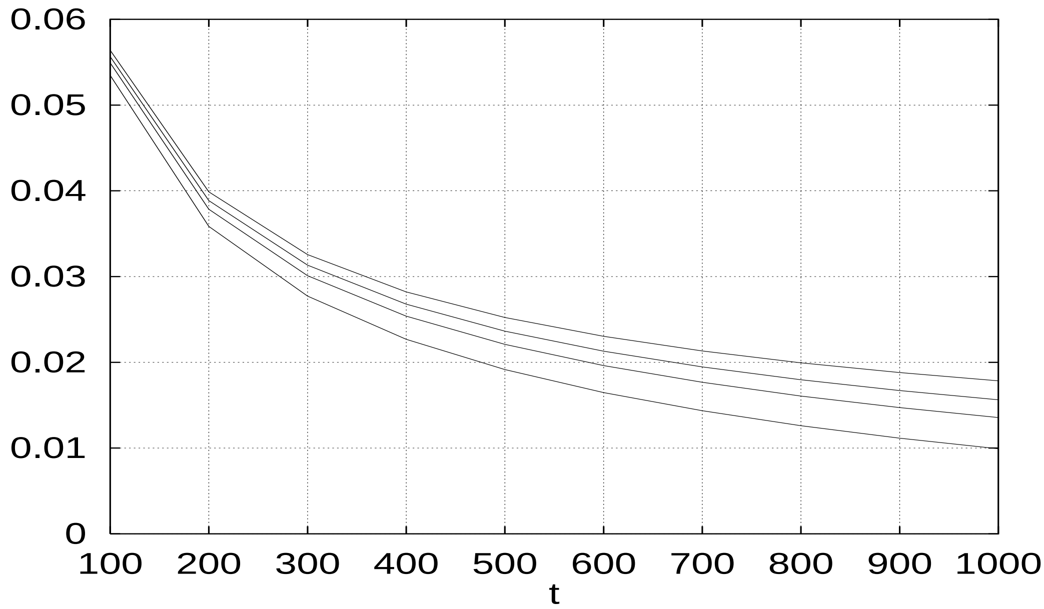

Figure 7: Distribution of $C\theta$ – from top to bottom : $\Pr\{\theta_{M/M/1} > t\}$, $\Pr\{4000\theta > t\}$, $\Pr\{2000\theta > t\}$, $\Pr\{1000\theta > t\}$ when $\rho = 1$.

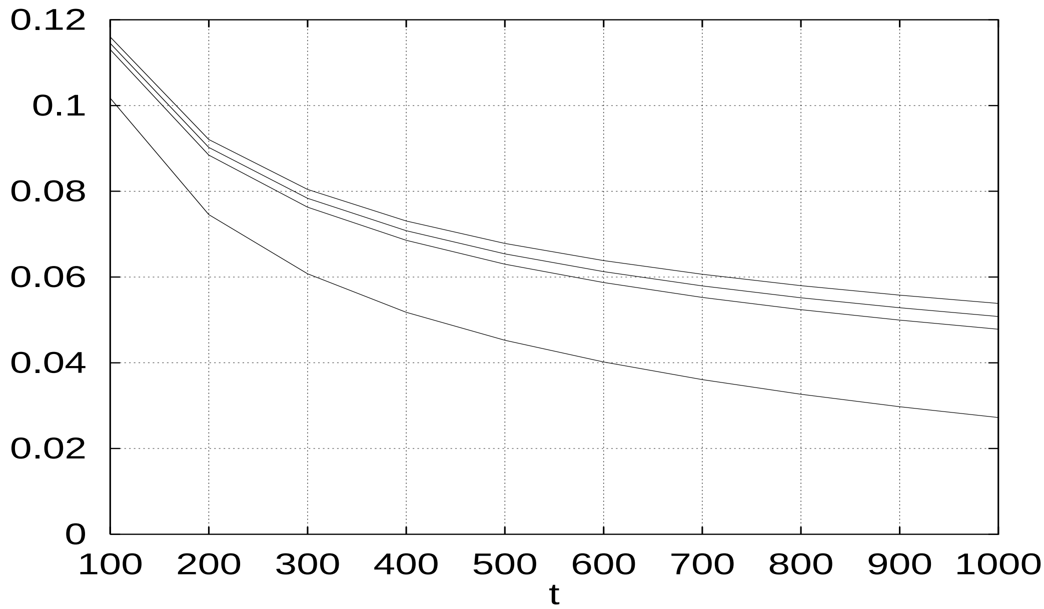

Figure 8: Distribution of CV – from the top to bottom : $\Pr\{V_{M/M/1} > t\}$, $\Pr\{1000\theta > t\}$, $\Pr\{500V > t\}$, $\Pr\{100V > t\}$ when $\rho = 1$.

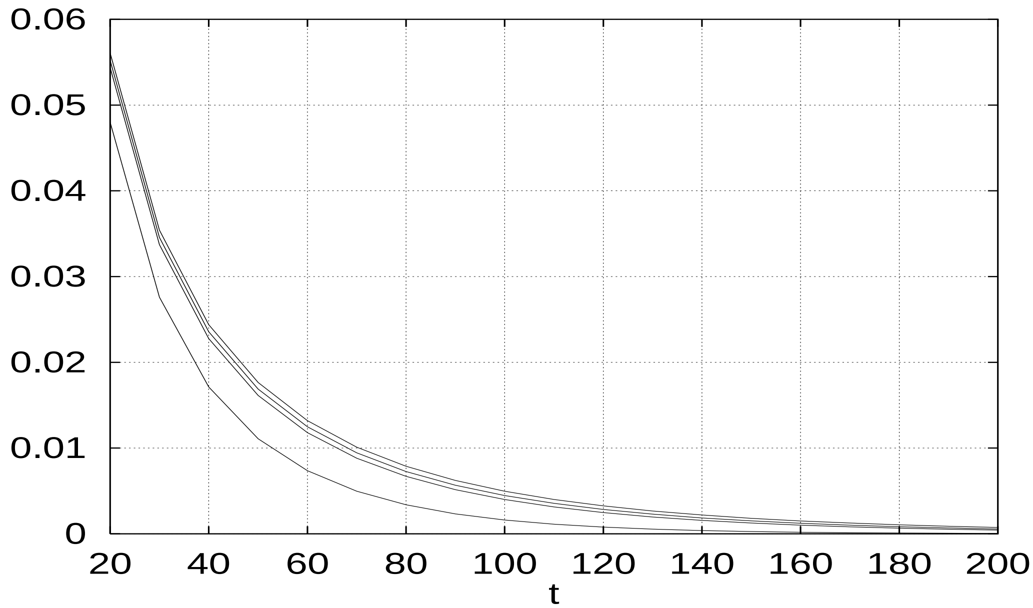

Figure 9: Distribution of $C\theta$ – from top to bottom : $\Pr\{\theta_{M/M/1} > t\}$, $\Pr\{1000\theta > t\}$, $\Pr\{500\theta > t\}$, $\Pr\{100\theta > t\}$ when $\rho = 0.8$.

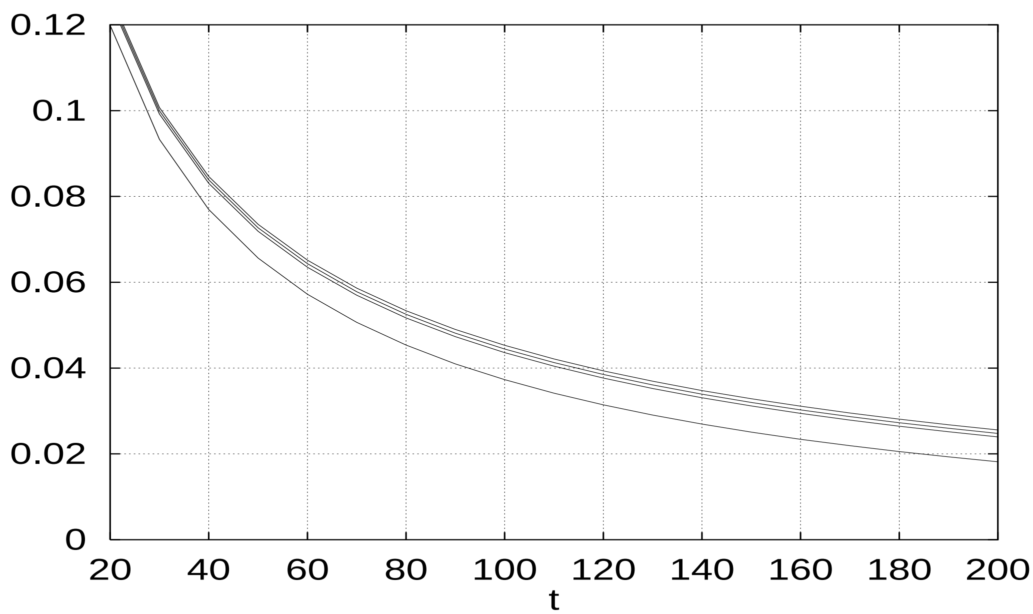

Figure 10: Distribution of CV – from top to bottom : $\Pr\{V_{M/M/1} > t\}$, $\Pr\{1000\theta > t\}$, $\Pr\{500V > t\}$, $\Pr\{100V > t\}$ when $\rho = 0.8$.

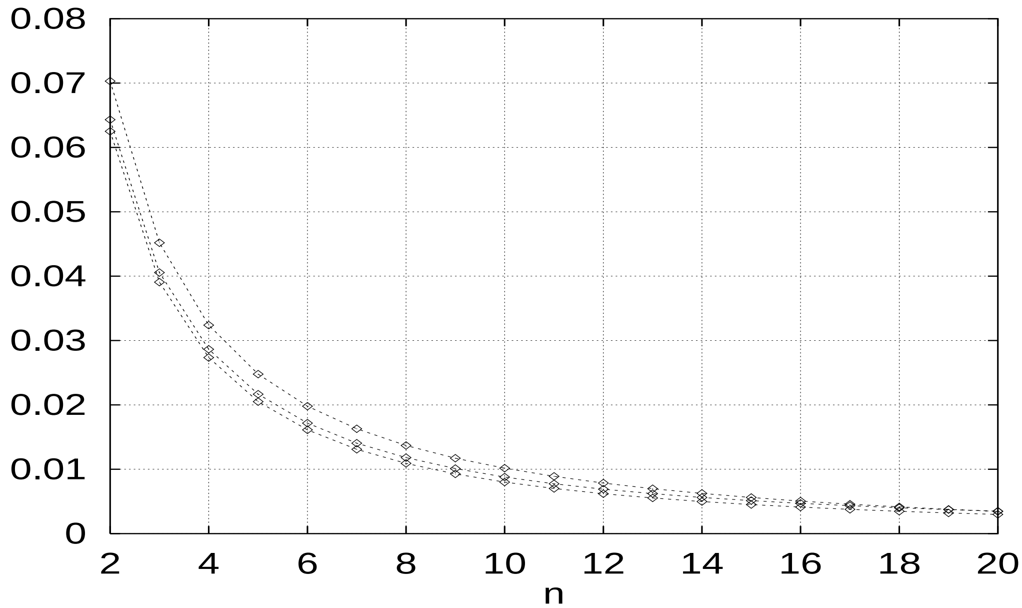

Figure 11: Distribution of N for $\rho = 1$ – from top to bottom : $\Pr\{N = n\}$ for $C = 10$, $\Pr\{N = n\}$ for $C = 50$, $\Pr\{N_{M/M/1} = n\}$.

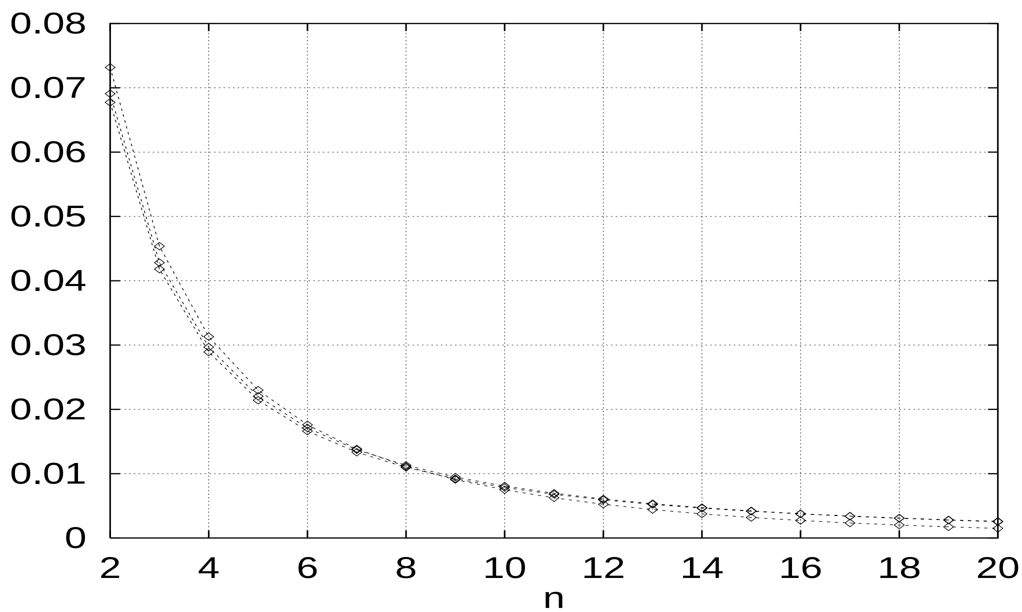

Figure 12: Distribution of N for $\rho = 0.8$ – from top to bottom : $\Pr\{N = n\}$ for $C = 10$, $\Pr\{N = n\}$ for $C = 50$, $\Pr\{N_{M/M/1} = n\}$.

7 Concluding remarks

The superposition of identical On/Off sources on an ATM link using open-loop statistical multiplexing and with no buffering to absorb burst scale congestion has been modeled by an $M/M/\infty$ queuing system. Some transient characteristics of this system were considered as relevant QoS parameters to characterize such a statistical multiplexing scheme. Using the Markov property of the occupation process, a uniformization technique has been invoked to compute the distribution of these characteristics. In addition, this technique allowed us in this particular context to prove the local approximation heuristic of Aldous, valid as $C \to \infty$ with mean load $u = \rho C$ for some $\rho \in (0, 1)$. The condition $C \to \infty$ is equivalent to assume that the actual transmission capacity C of the link is much greater than the individual source peak bit rate, thus ensuring a very small stationary overflow probability.

Another asymptotic result, namely the *Poisson Clumping Heuristic* [1], which has been proved in [4] for the $M/M/\infty$ system, claims that excursions above level C occur as a Poisson process with rate $\lambda = u^{C+1}e^{-u}/C!$ as C becomes very large. Given these asymptotic results, the next step in this line of investigation is to attach a buffer to the link in order to absorb overflow. Specifically, the problem is to dimension a single-server queue in a fluid flow approach where customers arrive as a Poisson process with rate λ. Upon each arrival, a volume $\mathcal{A}/C$ of information enters the system, $\mathcal{A}$ corresponding to the area swept by the occupation process during a busy period of an $M/M/1$ queue with arrival rate ρ and unit mean service time. The occupation process $\{\Lambda_t\}$ being Markovian, successive volumes of information are independent and identically distributed, with Laplace transform given in [4]. The buffer size should be chosen so that the proportion of rejected information is less than a given small threshold. In this approximation framework, the server rate is set equal to C because just before congestion occurs, the occupation process is close to the link capacity C. Since C is assumed to be large, it is reasonable to approximate by C the number of active servers, which determines the service rate in the $M/M/\infty$ system.

We would thus obtain a universal queue dimensioning of network buffers in an open-loop statistical multiplexing scheme when the link transmission capacities are much greater than the individual source bit rates. Above approximations should be obviously validated by simulation against realistic traffic models, which in particular account for the discrete-time nature of ATM cell traffic.

References

[1] D. Aldous : *Probability approximations via the Poisson Clumping Heuristic.* Springer Verlag, Vol. 77, 1989.

[2] P. N. Bowerman, R. G. Nolty and E. M. Scheuer : *Calculation of the Poisson cumulative distribution function.* IEEE Transactions on Reliability, 39(10), pp. 158-161. 1990.

[3] G. Ciardo, R. Marie, B. Sericola and K. S. Trivedi : *Performability analysis using semi-Markov reward processes.* IEEE Trans. Computers C-39, pp. 1251-1264, October 1990.

[4] F. Guillemin and A. Simonian : *Transient characteristics of an $M/M/\infty$ system.* Advances in Applied Probability, to appear, 1994.

[5] F. Guillemin, R. Mazumdar and A. Simonian : *On heavy traffic approximation for transient characteristics of $M/M/\infty$ queues.* Submitted, 1994.

[6] H. Kröner, T. Theimer, and U. Briem : *Queueing models for ATM systems - a comparison.* ITC Specialist Seminar, Morristown, October 1990.

[7] I. Norros, J. Roberts, A. Simonian and J. Virtamo : *The superposition of Variable Bit Rate sources in an ATM multiplexer.* IEEE J. on Sel. Areas in Commun. 9(3), pp. 378387, April 1991.

[8] J. Roberts : *Variable bit rate traffic control in B-ISDN.* IEEE Communications Magazine, pp. 50-56, September 1991.

[9] S. M. Ross : *Stochastic Processes.* J. Wiley, 1983.

[10] G. Rubino and B. Sericola : *Sojourn times in Markov processes.* J. Appl. Prob. 27, pp. 744-756, 1989.

[11] A. Simonian, F. Guillemin, J.R. Louvion and L. Romoeuf : *Transient analysis of statistical multiplexing on an ATM link.* Globecom'92, Orlando, December 1992.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur

INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)

ISSN 0249-6399