Queueing Systems 26 (1997) 121–150

Asymptotic results for the superposition of a large number of data connections on an ATM link

Alain Dupuis [a], Fabrice Guillemin [a] and Bruno Sericola [b]

[a] *France Telecom/CNET, DAC/ATM, 2 Avenue Pierre Marzin, 22300 Lannion, France*
E-mail: {Alain.Dupuis,Fabrice.Guillemin}@lannion.cnet.fr
[b] *IRISA-INRIA, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France*
E-mail: Bruno.Sericola@irisa.fr

Received 21 October 1996; revised 11 April 1997

The $M/PH/\infty$ system is introduced in this paper to analyze the superposition of a large number of data connections on an ATM link. In this model, information is transmitted in bursts of data arriving at the link as a Poisson process of rate λ and burst durations are PH distributed with unit mean. Some transient characteristics of the $M/PH/\infty$ system, namely the duration θ of an excursion by the occupation process $\{X_t\}$ above the link transmission capacity C, the area V swept under process $\{X_t\}$ above C and the number of customers arriving in such an excursion period, are introduced as performance measures. Explicit methods of computing their distributions are described. It is then shown that, as conjectured in earlier studies, random variables $C\theta$, CV, and N converge in distribution as C tends to infinity while the utilization factor of the link defined by $\gamma = \lambda/C$ is fixed in $(0, 1)$, towards some transient characteristics of an $M/M/1$ queue with input rate γ and unit service rate. Further simulation results show that after adjustment of the load of the $M/M/1$ queue, a similar convergence result holds for the superposition of a large number of On/Off sources with various On and Off period distributions. This shows that some transient quantities associated with an $M/M/1$ queue can be used in the characterization of open loop multiplexing of a large number of On/Off sources on an ATM link.

Keywords: ATM, statistical multiplexing, $M/PH/\infty$, queueing system, transient characteristics

1. Introduction

Among all the statistical multiplexing schemes, which have so far been introduced in the literature for Asynchronous Transfer Mode (ATM) networks, open-loop statistical multiplexing is certainly the simplest one. Indeed, this scheme simply consists in multiplexing different connections on an ATM link by overbooking without feedback control the link utilization (i.e., the instantaneous cell arrival rate may be greater than the link rate). In general, the link is equipped with a buffer intended to absorb the amount of information arriving in excess to the link transmission capacity. The simplicity of open loop statistical multiplexing is in that only traffic parameter enforcement at network access and suitable connection acceptance control are sufficient to meet possible Quality of Service (QoS) requirements in terms of cell transfer delay and cell loss ratio (CLR), say, a CLR objective about 10^{-9}.

In the framework of open-loop statistical multiplexing on an ATM link, the critical point is actually to determine, for a given traffic configuration and network utilization objective, the size of the buffer attached to a network link so that the QoS objectives are met, especially with regard to cell loss. A huge amount of work has been devoted to this issue over the past few years (see [13], for instance). Several studies [12], however, came to the conclusion that buffer dimensioning under burst scale congestion conditions is very uncertain because the performance of a queue in terms of cell loss and cell waiting time is then highly sensitive to the characteristics of the input process. This phenomenon has also been observed by using very simple queuing models [8]. In addition, Doshi [5] proved that the worst case assumption for buffer dimensioning is not always the well-known periodic On/Off behavior.

We will suppose in this paper that the buffer attached to the link is intended to absorb cell scale congestion only. More precisely, the size r of the buffer is assessed by assuming that all traffic sources are periodic and that the cumulative peak bit rate cannot exceed the link transmission capacity. A simple finite capacity $M/D/1/r$ queue can then be used to determine the buffer size r. For instance, r may be set equal to 128 to achieve a CLR of 10^{-10} and a link utilization factor of 85%. For more general arrival processes (e.g., traffic with bursts of data), it is then assumed that all information in excess to the link rate is lost. This assumption is known as the *unbuffered assumption* in the literature. Under such conditions, Doshi [5] notably proved that in a homogeneous environment the periodic On/Off behavior represents the worst case.

In view of the above discussion, this paper is intended to study the asymptotic behavior of the superposition of identical On/Off sources. We specifically consider the following random variables, which have been introduced in [10] as performance measures for characterizing open-loop statistical multiplexing on an ATM link:

- the duration θ of an excursion above the link transmission capacity C by process $\{X_t\}$ representing the cumulative peak bit rate of the superposition of the identical On/Off sources; such an excursion will be referred to as congestion period since it corresponds to an overflow period for the system;
- the area V swept under process $\{X_t\}$ above the link transmission capacity C in a congestion period; owing to the unbuffered assumption, V represents the volume of information lost in a congestion period;
- the number N of customers arriving in a congestion period; N is the number of bursts arriving in a congestion period.

Concerning the superposition of a large number of On/Off sources, it is known [6] that the properly rescaled superposition process $\{X_t^{(M)}\}$ of M identical On/Off exponential sources converges in distribution as M tends to infinity to the occupation process of an $M/M/\infty$ queue. Now, convergence results in [10] claim that as C tends to infinity while the utilization factor of the link defined by $\gamma = \lambda/C$ is fixed in $(0,1)$, the random variables $C\theta$, CV, and N associated with the $M/M/\infty$ system with input rate λ and unit mean service time converge in distribution to the respective transient characteristics Θ, $\mathcal{V}$, and $\mathcal{N}$ of the $M/M/1$ queue with unit service rate and mean arrival rate γ, where

- Θ denotes the busy period duration of the above $M/M/1$ queue;
- $\mathcal{V}$ is the area swept under the occupation process of this $M/M/1$ queue in a busy period;
- $\mathcal{N}$ is the number of customers served in a busy period.

This convergence result is essentially due the Markov property and the Aldous local linearization property [1] satisfied by the occupation process $\{X_t\}$. In [10], it was furthermore conjectured that the above convergence should hold for more general queues of the $M/G/\infty$ type.

In this paper, we prove the conjecture for $M/PH/\infty$ queues by taking benefit of the Markovian structure of such systems and by using uniformization technique [14]. Note that in the $M/PH/\infty$ model, the input Poisson process represents the arrival of bursts at the link and that bursts (i.e., activity periods of traffic sources) have a phase type distribution. The occupation process $\{X_t\}$ of the $M/PH/\infty$ queue then describes the cumulative peak bit rate of the bursts simultaneously multiplexed on the link. Given that PH distributions are dense in the set of probability distribution functions, this allows us to have confidence in the validity of the conjecture for $M/G/\infty$ systems. Furthermore, we show through simulation that after adjustment of the input rate of the $M/M/1$ queue, the above convergence result seems to be true for the superposition of a large number of more general On/Off sources.

2. The $M/PH/\infty$ system

Consider an $M/PH/\infty$ queue with mean arrival rate λ and $PH(\beta,T)$ distributed service times. The $PH(\beta,T)$ service time distribution is constructed by considering a Markov chain on the state space $\{1,\dots,l,l+1\}$, where states $1,\dots,l$ are transient and state $l+1$ is absorbing. The initial distribution of the Markov chain is $(\beta,0)$ where $\beta = (\beta_1,\dots,\beta_l)$ with $\sum_j \beta_j = 1$ and the transition rate matrix of the Markov chain in states $1,\dots,l$ is the $l \times l$ matrix T, whose entry (i,j) is denoted by $T(i,j) = \mu_{i,j}$, for $1 \leqslant i,j \leqslant l$. The random variable X is said to follow the distribution $PH(\beta,T)$ if X has the same distribution as the sojourn time of the above Markov chain in states $1,\dots,l$ until it gets absorbed in state $l+1$ [11]. In particular,

$$\Pr\{X > t\} = \beta \mathrm{e}^{Tt}\mathbf{1},$$

where $\mathbf{1}$ is the row vector, whose each entry is equal to 1 and whose dimension is determined by the context.

The matrix T is a sub-infinitesimal generator, that is, $\mu_{i,i} < 0$ for every i, $\mu_{i,j} \geqslant 0$ for every $i \neq j$ and $\sum_{j=1}^{l} \mu_{i,j} \leqslant 0$ for every i with strict inequality for at least one value of i. Defining $\mu_{i,0} = -\sum_{j=1}^{l} \mu_{i,j}$, we have $\mu_{i,0} \geqslant 0$ for every i and $\mu_{i,0} > 0$ for at least one value of i. For the sake of simplicity, we also define $\mu_i = -\mu_{i,i}$ for every i and $\mu = \max_{i=1,\dots,l} \mu_i$.

The mean service rate ϕ of a customer in the $M/PH/\infty$ queue is given by

$$\phi = \frac{1}{-\beta T^{-1}\mathbf{1}}.$$

In the following, we assume that $\phi = 1$.

Let $\{\Lambda_t\} = \{(\Lambda_t(1), \dots, \Lambda_t(l))\}$ denote the process such that for $i = 1, \dots, l$, $\Lambda_t(i)$ is the number of customers in the ith service phase at time t. Process $\{\Lambda_t\}$ is a Markov process with state space $\mathbb{N}^l$. For $k \geqslant 0$, we denote by S_k the subset of states where k customers are in the queue, namely

$$S_k = \left\{ (s_1, \dots, s_l) \in \mathbb{N}^l \;\middle|\; \sum_{i=1}^{l} s_i = k \right\}.$$

We also denote by X_t the total number of customers in the queue at time t, defined by

$$X_t = \sum_{i=1}^{l} \Lambda_t(i).$$

The infinitesimal generator of process $\{\Lambda_t\}$ is denoted by A. Using the partition of the state space induced by the subsets S_k, the matrix A has a block tridiagonal structure given by

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} & & & & & \\ A_{1,0} & A_{1,1} & A_{1,2} & & & & \\ & A_{2,1} & A_{2,2} & A_{2,3} & & & \\ & & A_{3,2} & \cdot & \cdot & & \\ & & & \cdot & \cdot & A_{k-1,k} & \\ & & & & A_{k,k-1} & A_{k,k} & A_{k,k+1} \\ & & & & & \cdot & \cdot & \cdot \end{pmatrix},$$

where the block $A_{i,j}$ contains the transitions from subset S_i to subset S_j.

Let the load ρ of the $M/PH/\infty$ queue be defined by

$$\rho \overset{\text{def}}{=} \frac{\lambda}{\phi} = \lambda,$$

since $\phi = 1$, and assume in the following that the $M/PH/\infty$ queue is in the stationary regime.

The stationary distribution of the process $\{\Lambda_t\}$ is denoted by π and is given by

$$\pi(s_1, \ldots, s_l) = \prod_{i=1}^{l} \mathrm{e}^{-\rho_i} \frac{\rho_i^{s_i}}{s_i!} = \mathrm{e}^{-\rho} \prod_{i=1}^{l} \frac{\rho_i^{s_i}}{s_i!}, \tag{1}$$

where

$$\rho_i = \lambda\bigl(-\beta T^{-1}\bigr)(i) \quad \text{and} \quad \rho = \sum_{i=1}^{l} \rho_i = \frac{\lambda}{\phi}.$$

We assume that for every $i = 1, \ldots, l$, we have $\rho_i > 0$. Indeed, if $\rho_i = 0$ for some i then the customers in service never visit the phase i so that this phase could be removed from the service distribution.

The distribution π can be decomposed in subvectors as

$$\pi = (\pi_0, \pi_1, \pi_2, \ldots),$$

where for $k \geqslant 0$, the subvector π_k is the projection of vector π over the subset S_k. Note that for every $k \geqslant 0$, we have

$$\pi_k \mathbf{1} = \mathrm{e}^{-\rho} \frac{\rho^k}{k!}. \tag{2}$$

Let the link transmission capacity C be a positive integer. We denote by B and B' the following subsets:

$$B = \bigcup_{k=0}^{C} S_k \quad \text{and} \quad B' = \bigcup_{k=C+1}^{\infty} S_k.$$

Using the partition $\{B, B'\}$, the infinitesimal generator A can be written as

$$A = \begin{pmatrix} A_B & A_{BB'} \\ A_{B'B} & A_{B'} \end{pmatrix}$$

and the stationary distribution π as

$$\pi = (\pi_B, \pi_{B'}).$$

With regard to the excursions by $\{\Lambda_t\}$ above C, a key quantity is the row vector probability distribution v over subset B' given by

$$v = (v_{C+1}, 0, 0, \ldots),$$

where for $s \in B'$, $v_{C+1}(s)$ is the probability that the excursion by process $\{\Lambda_t\}$ above C starts in state $s = (s_1, \ldots, s_l) \in S_{C+1}$, given that the $M/PH/\infty$ queue is in stationary regime. In [15], it is shown that the vector v is given by

$$v = \frac{\pi_B A_{BB'}}{\pi_B A_{BB'} \mathbf{1}}. \tag{3}$$

Using relations (3) and (1), we can state the following result for v_{C+1}, whose proof is given in the appendix.

Proposition 1. The probability that an excursion by process $\{\Lambda_t\}$ above C starts in state $s = (s_1, \ldots, s_l) \in S_{C+1}$ is given by

$$v_{C+1}(s) = \sum_{i=1}^{l} \beta_i C! \frac{(\rho_i/\rho)^{s_i - 1}}{(s_i - 1)!} \mathbf{1}_{\{s_i > 0\}} \prod_{k \neq i} \frac{(\rho_k/\rho)^{s_k}}{s_k!}. \tag{4}$$

3. Distribution of the volume V of lost information

Using the results of [4,15], the distribution of random variable V satisfies

$$\Pr\{V > t\} = v \mathrm{e}^{Mt} \mathbf{1}, \tag{5}$$

where v is given by proposition 1 and matrix M is defined by

$$M = R^{-1} A_{B'}, \tag{6}$$

with reward matrix R being a diagonal matrix over subset B', such that for every $k \geqslant C+1$, $R(s,s) = k - C$ if $s \in S_k$. Relation (5) can be justified as follows. We consider the reward Markov chain corresponding to the Markov chain $\{\Lambda_t\}$ with rewards given by matrix R and during an excursion in subset B'. Then, the infinitesimal generator of this reward Markov chain is given by M and relation (5) simply states that starting from state v, the reward Markov chain is still in subset B' at time t.

To compute the distribution of V given by relation (5), we first need to describe the transition rates of process $\{\Lambda_t\}$. Let e_i denote the ith canonical row vector of the state space $\mathbb{N}^l$; e_i has l entries with the ith entry being equal to 1 and the other ones being equal to 0. For $s \in \mathbb{N}^l$, the non-zero transition rates of matrix A from state s are

$$\begin{aligned} s &\to s + e_i \quad \text{with rate } \beta(i)\lambda, \\ s &\to s - e_i \quad \text{with rate } s_i \mu_{i,0} \mathbf{1}_{\{s_i \geqslant 1\}}, \\ s &\to s + e_j - e_i \quad \text{with rate } s_i \mu_{i,j} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{j \neq i\}}. \end{aligned}$$

By definition of matrix M, its non-zero transition rates from a state $s \in S_{C+n}$, $n \geqslant 1$, are

$$\begin{aligned} s &\to s + e_i \quad \text{with rate } \frac{\beta(i)\lambda}{n}, \\ s &\to s - e_i \quad \text{with rate } \frac{s_i \mu_{i,0}}{n} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{n \geqslant 2\}}, \\ s &\to s + e_j - e_i \quad \text{with rate } \frac{s_i \mu_{i,j}}{n} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{j \neq i\}}. \end{aligned} \tag{7}$$

The distribution of V can be obtained by uniformization [14] as follows. Let us denote by ν the uniformization rate associated with matrix M and given by

$$\nu = \sup_{s \in B'} \left| M(s,s) \right| = \sup_{n \geqslant 1} \nu_{C+n},$$

where ν_{C+n} is defined for $n \geqslant 1$ by

$$\nu_{C+n} = \sup_{s \in S_{C+n}} \left\{ \sum_{i=1}^{l} \frac{\beta(i)\lambda}{n} + \sum_{i=1}^{l} \frac{s_i \mu_{i,0}}{n} \mathbf{1}_{\{s_i \geqslant 1\}} + \sum_{i=1}^{l} \sum_{j=1}^{l} \frac{s_i \mu_{i,j}}{n} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{j \neq i\}} \right\}.$$

Since $\sum_{j=1}^{l} \mu_{i,j} \mathbf{1}_{\{j \neq i\}} = -\mu_{i,i} - \mu_{i,0} = \mu_i - \mu_{i,0}$,

$$\nu_{C+n} = \sup_{s \in S_{C+n}} \frac{1}{n} \left(\lambda + \sum_{i=1}^{l} s_i \mu_i \mathbf{1}_{\{s_i \geqslant 1\}} \right).$$

Using the definition of μ, which is the maximum of the μ_i's, we obtain

$$\nu_{C+n} = \frac{1}{n} \big(\lambda + (C+n)\mu \big),$$

and then,

$$\nu = \lambda + (C+1)\mu.$$

From equation (5), we have

$$\Pr\{V > t\} = v \mathrm{e}^{Mt} \mathbf{1} = \sum_{k=0}^{\infty} \mathrm{e}^{-\nu t} \frac{(\nu t)^k}{k!} v P^k \mathbf{1}, \tag{8}$$

where $P = \mathbf{I} + M/\nu$ is a sub-stochastic matrix over subset B'. The non-zero transition probabilities of matrix P from a state $s \in S_{C+n}$, $n \geqslant 1$, are

$$\begin{aligned}
s \to s + e_i \quad & \text{with probability } \frac{\beta(i)\lambda}{n\nu}, \\
s \to s - e_i \quad & \text{with probability } \frac{s_i \mu_{i,0}}{n\nu} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{n \geqslant 2\}}, \\
s \to s + e_j - e_i \quad & \text{with probability } \frac{s_i \mu_{i,j}}{n\nu} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{j \neq i\}}, \\
s \to s \quad & \text{with probability } 1 - \frac{\lambda + \sum_{i=1}^{l} s_i \mu_i \mathbf{1}_{\{s_i \geqslant 1\}}}{n\nu}.
\end{aligned} \tag{9}$$

Using relation (8), we can compute the distribution of V with an arbitrary error tolerance ε. Indeed, let

$$K = \min \left\{ k \in \mathbb{N} \;\middle|\; \sum_{j=0}^{k} \mathrm{e}^{-\nu t} \frac{(\nu t)^j}{j!} \geqslant 1 - \varepsilon \right\}. \tag{10}$$

We then have

$$\Pr\{V > t\} = \sum_{k=0}^{K} \mathrm{e}^{-\nu t} \frac{(\nu t)^k}{k!} v P^k \mathbf{1} + e(K), \tag{11}$$

where $e(K)$ is the rest of the series which verifies $e(K) \leqslant \varepsilon$.

Matrix P has the same structure as matrices A'_B and M. Defining the row vectors U_k over subset B' as $U_k = vP^k$, we have the recurrence relation $U_k = U_{k-1}P$ for $k \geqslant 1$ and $U_0 = v$. As in the case of vector v, the vector U_k can be decomposed over the partition $\{S_{C+n},\ n \geqslant 1\}$ of subset B' as $U_k = (U_{k,C+1}, U_{k,C+2}, \ldots, U_{k,C+n}, \ldots)$. It is easily checked that for $k \geqslant 0$, we have $U_{k,C+n} = 0$. This is due to the particular structure of vector v (namely $v_{C+n} = 0$ for $n \geqslant 2$) and to the block tridiagonal structure of matrix P.

We thus obtain the following recurrence relations for $k \geqslant 1$

$$\begin{aligned}
U_{k,C+1} &= U_{k-1,C+1}P_{C+1,C+1} + U_{k-1,C+2}P_{C+2,C+1},\\
U_{k,C+n} &= U_{k-1,C+n-1}P_{C+n-1,C+n} + U_{k-1,C+n}P_{C+n,C+n} + U_{k-1,C+n+1}P_{C+n+1,C+n}\\
&\qquad \text{for } 2 \leqslant n \leqslant k+1,\\
U_{k,C+n} &= 0 \quad \text{for } n \geqslant k+2.
\end{aligned}$$

Using relation (11), we have for every $t \geqslant 0$ and for every $\varepsilon > 0$,

$$0 \leqslant \Pr\{V > t\} - \sum_{k=0}^{K} \mathrm{e}^{-\nu t} \frac{(\nu t)^k}{k!} U_{k,C+1}\mathbf{1} \leqslant \varepsilon.$$

If we set $x_k = U_{k,C+1}\mathbf{1}$, the distribution of V can be obtained, for a fixed value of t and a fixed value ε of the error tolerance, by using the algorithm, whose pseudocode is given below.

PSEUDOCODE OF THE ALGORITHM FOR COMPUTING THE DISTRIBUTION OF RANDOM VARIABLE V

Step 0. Compute K using relation (10)
Step 1. $x_0 = 1$, $U_{1,C+1} = v_{C+1}P_{C+1,C+1}$, $U_{1,C+2} = v_{C+1}P_{C+1,C+2}$, $x_1 = U_{1,C+1}\mathbf{1}$
Step 2. for $k = 2$ to K do

$$\begin{aligned}
&U_{k,C+1} = U_{k-1,C+1}P_{C+1,C+1} + U_{k-1,C+2}P_{C+2,C+1}\\
&\text{for } n = 2 \text{ to } k-1 \text{ do}\\
&\quad U_{k,C+n} = U_{k-1,C+n-1}P_{C+n-1,C+n} + U_{k-1,C+n}P_{C+n,C+n}\\
&\qquad\qquad + U_{k-1,C+n+1}P_{C+n+1,C+n}\\
&\text{endfor}\\
&U_{k,C+k} = U_{k-1,C+k-1}P_{C+k-1,C+k} + U_{k-1,C+k}P_{C+k,C+k}\\
&U_{k,C+k+1} = U_{k-1,C+k}P_{C+k,C+k+1}\\
&x_k = U_{k,C+1}\mathbf{1}
\end{aligned}$$

endfor

Step 3. $\Pr\{V > t\} = \sum_{k=0}^{K} \mathrm{e}^{-\nu t} \frac{(\nu t)^k}{k!} x_k$

4. Distribution of the congestion duration θ

Using again the results of [15], the distribution of random variable θ is given by

$$\Pr\{\theta > t\} = v\mathrm{e}^{A_{B'}t}\mathbf{1} \quad \text{for } t \geqslant 0,$$

where v is defined in proposition 1. The uniformization technique invoked previously to compute the distribution of random variable V is unfortunately not applicable for the computation of the distribution of random variable θ. This is due to the fact that the uniformization factor

$$\sup_{s\in B'} \left|A_{B'}(s,s)\right| = \infty.$$

To overcome this problem, we introduce as in [9] two auxiliary random variables, denoted by $\theta_n^{\inf}$ and $\theta_n^{\sup}$, such that

$$\Pr\{\theta_n^{\inf} > t\} \leqslant \Pr\{\theta > t\} \leqslant \Pr\{\theta_n^{\sup} > t\},$$
$$\lim_{n\to\infty} \Pr\{\theta_n^{\inf} > t\} = \lim_{n\to\infty} \Pr\{\theta_n^{\sup} > t\} = \Pr\{\theta > t\},$$

and for which the uniformization technique applies.

4.1. *Distribution of random variable* $\theta_n^{\inf}$

For every $n \geqslant 1$, we consider the $M/PH/C+n/C+n$ loss system with $C+n$ servers. We denote by $\{\Lambda_t^n\}$ the occupation process of that queue. This process is a Markov process over the finite state space $S^{\inf}$ given by

$$S^{\inf} = \bigcup_{k=0}^{C+n} S_k.$$

Its infinitesimal generator $A_n^{\inf}$ is

$$A_n^{\inf} = \begin{pmatrix} A_{0,0} & A_{0,1} & & & \\ A_{1,0} & A_{1,1} & & & \\ & A_{2,1} & \cdot & & \\ & & \cdot & \cdot & \\ & & \cdot & \cdot & A_{C+n-1,C+n} \\ & & & A_{C+n,C+n-1} & A_{C+n,C+n}^{\inf} \end{pmatrix},$$

where submatrices $A_{i,j}$ are defined in section 2 and matrix $A_{C+n,C+n}^{\inf}$ corresponds to matrix $A_{C+n,C+n}$ in which parameter λ has been taken equal to 0.

We take as initial distribution of process $\{\Lambda_t^n\}$ the row vector α given by

$$\alpha = (0,\ldots,0,v_{C+1},0,\ldots,0).$$

We then define $\theta_n^{\inf}$ as the duration of an excursion by process $\{\Lambda_t^n\}$ in subset $S^{[n]} = S_{C+1} \cup \cdots \cup S_{C+n}$, given that the queue is in the stationary regime.

As usual, the distribution of $\theta_n^{\inf}$ is given by

$$\Pr\{\theta_n^{\inf} > t\} = v^{[n]} e^{B_n^{\inf} t} \mathbf{1},$$

where $v^{[n]}$ is the row vector over subset $S^{[n]}$ defined by

$$v^{[n]} = (v_{C+1}, 0, \ldots, 0)$$

and matrix $B_n^{\inf}$ is given by

$$B_n^{\inf} = \begin{pmatrix} A_{C+1,C+1} & A_{C+1,C+2} & & & \\ A_{C+2,C+1} & A_{C+2,C+2} & A_{C+2,C+3} & & \\ & A_{C+3,C+2} & A_{C+3,C+3} & & \\ & & A_{C+4,C+3} & \cdot & \\ & & & \cdot & A_{C+n-1,C+n} \\ & & & A_{C+n,C+n-1} & A^{\inf}_{C+n,C+n} \end{pmatrix}.$$

$\Pr\{\theta_n^{\inf} > t\}$, which depends on the truncation level n, is increasing function of n and verifies

$$\lim_{n\to\infty} \Pr\{\theta_n^{\inf} > t\} = \Pr\{\theta > t\}.$$

$\Pr\{\theta_n^{\inf} > t\}$ is actually an approximant from below of $\Pr\{\theta > t\}$. The key point is that for any value of n the distribution of $\theta_n^{\inf}$ can be computed by uniformization techniques.

4.2. *Distribution of random variable* $\theta_n^{\sup}$

Random variable $\theta_n^{\sup}$ is obtained by modifying the reward matrix used in the computation of random variable V, namely by replacing some reward coefficients by 1 so as to obtain information on the time spent in some states during an excursion by the occupation process $\{X_t\}$ of the $M/PH/\infty$ queue above level C. Given that process $\{\Lambda_t\}$ starts an excursion in subspace B' at time 0, $\theta_n^{\sup}$ is specifically defined by

$$\theta_n^{\sup} = \sum_{i=1}^{n} \sum_{s\in S_{C+i}} \int_0^\infty \mathbf{1}_{\{\tilde{\Lambda}_u = s\}}\,\mathrm{d}u + \sum_{i=n+1}^{\infty} \sum_{s\in S_{C+i}} i \int_0^\infty \mathbf{1}_{\{\tilde{\Lambda}_u = s\}}\,\mathrm{d}u, \qquad (12)$$

where $\{\tilde{\Lambda}_t\}$ is the Markov process corresponding to process $\{\Lambda_t\}$ absorbed in subspace B (i.e., $\tilde{\Lambda}_t = \Lambda_{t\wedge\theta}$).

It is straightforward that

$$\theta'_n \stackrel{\text{def}}{=} \sum_{i=1}^{n} \sum_{s\in S_{C+i}} \int_0^\infty \mathbf{1}_{\{\tilde{\Lambda}_u = s\}}\,\mathrm{d}u \to \theta \text{ a.s.} \quad \text{when} \quad n\to\infty.$$

Moreover, write $\theta_n^{\sup} = \theta'_n + \tilde{\theta}_n^{\sup}$ with

$$\tilde{\theta}_n^{\sup} = \sum_{i=n+1}^{\infty} \sum_{s \in S_{C+i}} i \int_0^{\infty} \mathbf{1}_{\{\tilde{\Lambda}_u = s\}} \mathrm{d}u$$

and note that the volume V of lost information over a congestion period can be expressed as

$$V = \sum_{i=1}^{\infty} \sum_{s \in S_{C+i}} \int_0^{\theta} i \mathbf{1}_{\{\Lambda_u = s\}} \mathrm{d}u$$

or equivalently, using the process $\{\tilde{\Lambda}_t\}$,

$$V = \sum_{i=1}^{\infty} \sum_{s \in S_{C+i}} \int_0^{\infty} i \mathbf{1}_{\{\tilde{\Lambda}_u = s\}} \mathrm{d}u. \tag{13}$$

A classical result in stochastic point process theory (namely the mean value formula [3]) entails that the mean value of random variable V is given by

$$\mathrm{E}[V] = \frac{1}{\lambda_C} \sum_{i=1}^{\infty} i \pi_{C+i} \mathbf{1} = \frac{\mathrm{e}^{-\rho}}{\lambda_C} \sum_{i=1}^{\infty} \frac{i}{(C+i)!} \rho^{C+i},$$

where λ_C is the mean intensity of the point process counting the excursions of process $\{\Lambda_t\}$ in subspace B' given by

$$\lambda_C = \lambda \pi_C \mathbf{1} = \lambda \frac{\rho^C}{C!} \mathrm{e}^{-\rho},$$

and then

$$\mathrm{E}[V] = \frac{1}{\lambda} \sum_{i=1}^{\infty} i \frac{C!}{(C+i)!} \rho^i \mathrm{e}^{-\rho}.$$

Note that the mean value of V for the $M/PH/\infty$ queue has the same value of the corresponding variable for the $M/M/\infty$ queue given in [10]. This property is due to the fact that the respective occupation processes in both systems have the same stationary distribution.

It follows that $\mathrm{E}[V] < \infty$ and that the series (13) is a.s. converging. This entails in particular that its remainder to order n tends a.s. to 0 as n tends to infinity, or equivalently, $\tilde{\theta}_n^{\sup} \to 0$ a.s. when $n \to \infty$. It follows that $\theta_n^{\sup}$ is a decreasing sequence of random variables with respect to n and that $\theta_n^{\sup} \to \theta$ a.s. when $n \to \infty$.

By taking as initial distribution the row vector v for the process $\{\tilde{\Lambda}_t\}$, the distribution of $\theta_n^{\sup}$ is given as in [9] by

$$\Pr\{\theta_n^{\sup} > t\} = v \mathrm{e}^{M_n t} \mathbf{1},$$

where matrix M_n is defined by

$$M_n = R_n^{-1} A_{B'}$$

with the reward matrix R_n being a diagonal matrix over subset B', such that for every $k \geqslant C+1$

$$\begin{aligned} R_n(s,s) &= 1 && \text{if } s \in S_k \text{ with } k \leqslant C+n, \\ R_n(s,s) &= k-C && \text{if } s \in S_k \text{ with } k > C+n. \end{aligned}$$

The computation of the distribution of $\theta_n^{\sup}$ is similar to that of the distribution of V. The only difference is that the reward matrix R used in the computation of the distribution of V has to be replaced with matrix R_n.

It follows that the distribution of θ can be approximated by those of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ for large enough n, for which uniformization techniques apply. To evaluate the distribution of θ, we have to compute the distribution of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ for a sufficiently large value of n so that the difference $\Pr\{\theta_n^{\sup} > t\} - \Pr\{\theta_n^{\inf} > t\}$ is less than a given error tolerance. For that purpose the value of n can be first arbitrarily chosen and then increased until the difference becomes small enough. The distributions of $\theta_n^{\inf}$ and $\theta_n^{\sup}$ are computed by using an algorithm similar to that given for random variable V.

5. Distribution of the number N of bursts in a congestion period

For convenience, instead of directly dealing with random variable N, we consider in a first step random variable N' describing the number of customers arriving during a sojourn of process $\{\Lambda_t\}$ in the subset B'. We obviously have $N = N' + 1$. Moreover, since

$$\Pr\{N = k\} = \sum_{s \in S_{C+1}} v_{C+1}(s) \Pr\bigl\{N = k \mid \Lambda_0 = s\bigr\}, \tag{14}$$

we are led to evaluate the conditional probabilities $\Pr\{N' = k \mid \Lambda_0 = s\}$.

For this purpose, we consider the embedded Markov chain $\{Z_r\}_{r \geqslant 1}$ at the jump instants of process of $\{\Lambda_t\}$, with $Z_0 = \Lambda_0$. We denote by Q the transition probability matrix of $\{Z_r\}$. The restriction of matrix Q over subset B' is denoted by $Q_{B'}$.

The non-zero transition probabilities of matrix $Q_{B'}$ are given for every $n \geqslant 1$ and $s \in S_{C+n}$ by

$$\begin{aligned} s &\to s + e_i && \text{with probability } \frac{\beta(i)\lambda}{\lambda + \sum_{i=1}^{l} s_i \mu_i}, \\ s &\to s - e_i && \text{with probability } \frac{s_i \mu_{i,0} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{n \geqslant 2\}}}{\lambda + \sum_{i=1}^{l} s_i \mu_i}, \\ s &\to s + e_j - e_i && \text{with probability } \frac{s_i \mu_{i,j} \mathbf{1}_{\{s_i \geqslant 1\}} \mathbf{1}_{\{j \neq i\}}}{\lambda + \sum_{i=1}^{l} s_i \mu_i}. \end{aligned} \tag{15}$$

Matrix $Q_{B'}$ has the same tridiagonal block structure as matrix $A_{B'}$. In a similar fashion, the blocks of matrix $Q_{B'}$ are denoted by $Q_{C+1,C+1}$, $Q_{C+1,C+2}$, and for $n \geqslant 2$, $Q_{C+n,C+n-1}$, $Q_{C+n,C+n}$, and $Q_{C+n,C+n+1}$.

Since for $s \in S_{C+n}$

$$\begin{aligned}\Pr\{N' = k \mid Z_0 = s\} = &\sum_{s' \in S_{C+n-1}} Q(s,s')\Pr\{N' = k \mid Z_0 = s'\} \\ &+ \sum_{s' \in S_{C+n}} Q(s,s')\Pr\{N' = k \mid Z_0 = s'\} \\ &+ \sum_{s' \in S_{C+n+1}} Q(s,s')\Pr\{N' = k-1 \mid Z_0 = s'\}, \quad (16)\end{aligned}$$

with the initial condition $\Pr\{N' = k \mid Z_0 = s\} = \mathbf{1}_{\{k=0\}}$ for $s \in S_C$, we have

$$U_k(n) = Q_{C+n,C+n-1}U_k(n-1) + Q_{C+n,C+n}U_k(n) + Q_{C+n,C+n+1}U_{k-1}(n+1), \quad (17)$$

where for $k \geqslant 0$ and $n \geqslant 0$, $U_k(n)$ is a column vector containing the probabilities $\Pr\{N' = k \mid Z_0 = s\}$ when $s \in S_{C+n}$ and the initial condition is

$$U_k(0) = \begin{cases} \mathbf{1} & \text{if } k = 0, \\ 0 & \text{otherwise,} \end{cases}$$

and $U_{-1}(n) = 0$ for $n \geqslant 0$.

With this notation, we get from equation (14) $\Pr\{N = 0\} = 0$ and for every $k \geqslant 1$

$$\Pr\{N = k\} = v_{C+1}U_{k-1}(1). \quad (18)$$

We now introduce the matrices $P_{n,n-1}$ and $P_{n,n+1}$ defined for $n \geqslant 1$ by

$$\begin{aligned}P_{n,n-1} &= (\mathbf{I} - Q_{C+n,C+n})^{-1}Q_{C+n,C+n-1}, \\ P_{n,n+1} &= (\mathbf{I} - Q_{C+n,C+n})^{-1}Q_{C+n,C+n+1},\end{aligned}$$

where $\mathbf{I}$ is the identity matrix.

Relation (17) can also be written for $k, n \geqslant 1$ as

$$U_k(n) = P_{n,n-1}U_k(n-1) + P_{n,n+1}U_{k-1}(n+1). \quad (19)$$

To compute the $K+1$ first values of the distribution of N, we need to compute the vectors $U_0(1), \ldots, U_K(1)$ and then use relation (18). These computations can be done recursively from relation (19) as shown below.

ALGORITHM FOR COMPUTING THE DISTRIBUTION OF RANDOM VARIABLE N

Step 0. $U_0(0) = \mathbf{1}$
Step 1. for $k = 0$ to K do
    for $n = 1$ to $K - k + 1$ do
        $U_k(n) = P_{n,n-1}U_k(n-1) + P_{n,n+1}U_{k-1}(n+1)$
        $\Pr\{N = k+1\} = v_{C+1}U_k(1)$
    endfor
endfor

6. Asymptotic results

We prove in this section the validity of the conjecture invoked in the Introduction on the asymptotic behavior of random variables $C\theta$, CV, and N as C tends to infinity while the link utilization factor $\gamma = \lambda/C$ is fixed in $(0,1)$. First of all, let us state the following technical lemma, which shows that the mean service rate ϕ (taken equal to 1) of the phase type service time distribution can be written as a convex linear combination of coefficients $\mu_{i,0}$.

Lemma 2. The mean service rate ϕ in the $M/PH/\infty$ queue can be expressed in terms of coefficients $\mu_{i,0}$ as

$$\phi = \sum_{i=1}^{l} \frac{\rho_i}{\rho} \mu_{i,0}. \tag{20}$$

Proof. We have

$$\phi = \frac{1}{-\beta T^{-1}\mathbf{1}}.$$

Let T^0 be the column vector of dimension l whose ith entry is equal to $\mu_{i,0}$. This vector verifies $T^0 = -T\mathbf{1}$ and then,

$$\sum_{i=1}^{l} \mu_{i,0}\bigl(\beta T^{-1}\bigr)(i) = \beta T^{-1}T^0 = -1.$$

Hence, by definition of ρ_i and ρ, we get

$$\sum_{i=1}^{l} \frac{\rho_i}{\rho}\mu_{i,0} = \sum_{i=1}^{l} \frac{(\beta T^{-1})(i)}{\beta T^{-1}\mathbf{1}}\mu_{i,0} = \frac{1}{-\beta T^{-1}\mathbf{1}} = \phi,$$

and the result follows. □

To establish the convergence result for CV, we need the following technical lemmas concerning the convergence of conditional probabilities, whose proofs can be found in the appendix. Moreover, to ensure convergence, we redefine for $n \geqslant 0$ the subset S_{C+n} as

$$S_{C+n} = \left\{ (s_1, \ldots, s_l) \in \mathbb{N}^l \,\middle|\, \sum_{i=1}^{l} s_i = C + n \text{ and } s_i/C \text{ converges when } C \to \infty \right\}.$$

Lemma 3. For every $s = (s_1, \ldots, s_l) \in S_{C+1}$, we have,

$$\Pr\bigl\{CV > t \mid \Lambda_0 = s\bigr\} \to \Pr\left\{\mathcal{V}\left(\gamma, \sum_{i=1}^{l} r_i \mu_{i,0}\right) > t\right\},$$

as C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0,1)$, where $r_i = \lim_{C\to\infty} s_i/C$, for $i = 1,\ldots,l$ and $\mathcal{V}(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0})$ is the area swept in a busy period under the occupation process of the $M/M/1$ queue with input rate γ and mean service rate $\sum_{i=1}^{l} r_i\mu_{i,0}$.

Lemma 4. For every $i = 1,\ldots,l$, we have

$$\max_{s\in S_C}\left|\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s+e_i) - \Pr\left\{\mathcal{V}\left(\gamma, \sum_{j=1}^{l}\frac{s_j}{C}\mu_{j,0}\right) > t\right\}\right| \to 0$$

as C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0,1)$, where $\mathcal{V}(\gamma, \sum_{j=1}^{l} s_j\mu_{j,0}/C)$ is the area swept in a busy period under the occupation process of an $M/M/1$ queue with input rate γ and mean service rate $\sum_{j=1}^{l} s_j\mu_{j,0}/C$.

These two technical lemmas enable us to state the convergence result for the asymptotic behavior of random variable CV. The proof of the following theorem is also given in the appendix.

Theorem 5. When C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0,1)$,

$$\text{for } t \geqslant 0, \quad \Pr\{CV > t\} \to \Pr\{\mathcal{V} > t\},$$

where $\mathcal{V}$ is the area swept in a busy period under the occupation process of the $M/M/1$ queue with input rate γ and unit service rate.

Using representation (14) for the distribution of random variable N, the same arguments as those used to obtain relation (38) for CV lead to

$$\Pr\{N = k\} = \sum_{i=1}^{l}\beta_i \sum_{s\in S_C}\frac{C!}{s_1!\cdots s_l!}\prod_{j=1}^{l}\left(\frac{\rho_j}{\rho}\right)^{s_j}\Pr\bigl\{N = k \mid \Lambda_0 = s + e_i\bigr\}.$$

Before stating the convergence result for random variable N, we need the following lemma, whose proof can be found in the appendix.

Lemma 6. For every $s = (s_1,\ldots,s_l) \in S_{C+1}$, we have,

$$\text{for } k \geqslant 0, \quad \Pr\bigl\{N = k \mid \Lambda_0 = s\bigr\} \to \Pr\left\{\mathcal{N}\left(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0}\right) = k\right\}$$

as $C \to \infty$ with $\gamma = \lambda/C$ fixed in $(0,1)$, where $r_i = \lim_{C\to\infty} s_i/C$, for $i = 1,\ldots,l$ and $\mathcal{N}(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0})$ is the number of customers served in the busy of the $M/M/1$ queue with arrival rate γ and mean service rate $\sum_{i=1}^{l} r_i\mu_{i,0}$.

Invoking the same arguments as those used to obtain lemma 4 (i.e., the maximum is reached for a value $s^* \in S_{C+1}$ and the function $\Pr\{\mathcal{N}(\gamma, y) = k\}$ is continuous with respect to $y \in [0, \infty[$, an easy consequence of the previous lemma is the following result.

Lemma 7. For every $k \geqslant 0$, we have

$$\max_{s \in S_{C+1}} \left| \Pr\{N = k \mid \Lambda_0 = s\} - \Pr\left\{ \mathcal{N}\left(\gamma, \sum_{j=1}^{l} \frac{s_j}{C} \mu_{j,0}\right) = k \right\} \right| \to 0$$

as C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0, 1)$, where $\mathcal{N}(\gamma, \sum_{j=1}^{l} s_j \mu_{j,0}/C)$ is the number of customers served in a busy period of an $M/M/1$ queue with input rate γ and mean service rate $\sum_{j=1}^{l} s_j \mu_{j,0}/C$.

The same arguments as in the proof of theorem 5 along with lemma 7 and Weierstrass's theorem allow us to state the following result.

Theorem 8. When C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0, 1)$,

$$\text{for } k \geqslant 0, \quad \Pr\{N = k\} \to \Pr\{\mathcal{N} = k\}, \tag{21}$$

where $\mathcal{N}$ is the number of customers served in a busy period of the $M/M/1$ queue with input rate γ and unit service rate.

Finally, replacing M, V, and $\mathcal{V}$ with $A_{B'}$, θ, and Θ, respectively, a straightforward adaptation of the proofs of lemmas 3 and 4, and theorem 5 allows us to state the following results for θ.

Lemma 9. For every $s = (s_1, \ldots, s_l) \in S_{C+1}$, we have

$$\Pr\{C\theta > t \mid \Lambda_0 = s\} \to \Pr\left\{ \Theta\left(\gamma, \sum_{i=1}^{l} r_i \mu_{i,0}\right) > t \right\}$$

as C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0, 1)$, where $r_i = \lim_{C\to\infty} s_i/C$, for $i = 1, \ldots, l$ and

$$\Theta\left(\gamma, \sum_{i=1}^{l} r_i \mu_{i,0}\right)$$

is the duration of a busy period of the $M/M/1$ queue with input rate γ and mean service rate $\sum_{i=1}^{l} r_i \mu_{i,0}$.

Lemma 10. For every $i = 1, \ldots, l$, we have

$$\max_{s \in S_C} \left| \left(\mathrm{e}^{Mt/C} \mathbf{1}\right)(s + e_i) - \Pr\left\{ \Theta\left(\gamma, \sum_{j=1}^{l} \frac{s_j}{C} \mu_{j,0}\right) > t \right\} \right| \to 0 \tag{22}$$

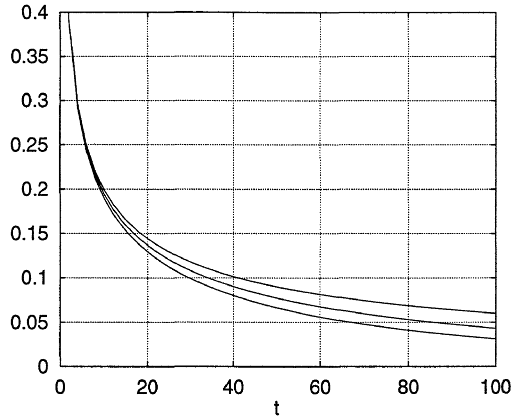

Figure 1. From top to the bottom: $\Pr\{\mathcal{V} > t\}$, $\Pr\{100V > t\}$, and $\Pr\{200V > t\}$.

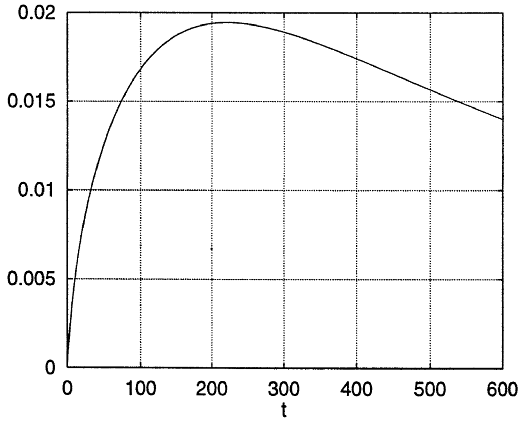

Figure 2. $\Pr\{\mathcal{V} > t\} - \Pr\{200V > t\}$.

as C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0,1)$, where $\Theta(\gamma, \sum_{j=1}^{l} s_j \mu_{j,0}/C)$ is the duration of a busy period of the $M/M/1$ queue with input rate γ and mean service rate $\sum_{j=1}^{l} s_j \mu_{j,0}/C$.

Theorem 11. When C tends to ∞ with $\gamma = \lambda/C$ fixed in $(0,1)$,

$$\text{for } t \geqslant 0, \quad \Pr\{C\theta > t\} \to \Pr\{\Theta > t\}, \tag{23}$$

where Θ is duration of a busy period of the $M/M/1$ queue with input rate γ and unit service rate.

To illustrate from a numerical point of view the above convergence results, consider the $M/PH/\infty$ queue with mean arrival rate λ and $PH(\beta, T)$ service times distribution given by

$$\beta = (1, 0) \quad \text{and} \quad T = \begin{pmatrix} -2 & 2 \\ 0 & -2 \end{pmatrix}.$$

Note that this distribution is an Erlang distribution with 2 phases and mean 1. It follows that the mean service rate is $\phi = 1$. We consider two cases: $C = 100$ and $C = 200$ and we fix the value of γ to $\gamma = \lambda/C = 0.85$. For $C = 100$, we have $\lambda = 85$ and for $C = 200$, $\lambda = 170$.

Figure 1 shows the distributions of random variables $100V$ and $200V$ and the distribution of random variable $\mathcal{V}$ corresponding to the area swept under the occupation process during a busy period of the $M/M/1$ with mean arrival rate γ and unit mean service time. The distribution of $\mathcal{V}$ clearly appears to be the limit of the distributions of random variables $100V$ and $200V$.

Figure 2 shows the difference between the distributions of random variables $200V$ and $\mathcal{V}$. In both figures, the value of the error tolerance ε has been chosen equal to 10^{-5}.

7. Further simulation results

Since the introduction in the literature of the celebrated Anick, Mitra, and Sondhi model [2] for the analysis of a system handling multiple data sources, a huge amount of work has been devoted to the study of the superposition of a large number of On/Off sources on an ATM link, especially under the buffered assumption (see [13], for instance). In this section, going further in the investigations on the asymptotic behavior of random variables $C\theta$, CV, and N, we consider under the unbuffered assumption the superposition on an ATM link of a large number of On/Off sources with various On and Off distributions.

In the previous sections, it has been shown that the local linearization property conjectured by Aldous in [1] for $M/M/\infty$ queues and rigorously proved via Laplace transform analysis in [10], holds for $M/PH/\infty$ type queues. Similarly, if we consider now the superposition of On/Off sources with general arrival processes and general burst durations, we first examine the case of S exponential On/Off sources. Assuming that the mean burst duration and the peak bit rate of a source are equal to 1, the infinitesimal generator of process $\{\Lambda_t\}$ representing at time t the number of active sources (or equivalently the cumulative peak bit rate of the bursts simultaneously multiplexed on the link) is given by

$$\begin{pmatrix} -S\lambda & S\lambda & 0 & 0 & \\ 1 & -(S-1)\lambda - 1 & (S-1)\lambda & 0 & \\ 0 & 2 & -(S-2)\lambda - 2 & \ddots & \\ 0 & 0 & 3 & \ddots & \lambda \\ & & & S & -S \end{pmatrix}, \tag{24}$$

where $1/\lambda$ is the mean silence duration and the number S of On/Off sources is assumed to be much greater than the link transmission capacity C.

Using the same arguments as in [1], the excursion process above level C associated with process $\{\Lambda_t\}$ can be approximated via local approximation by process $\{C\Lambda_t'\}$ with infinitesimal generator

$$\begin{pmatrix} -\left(\frac{S}{C}-1\right)\lambda & \left(\frac{S}{C}-1\right)\lambda & 0 & \dots & & & \\ 1 & -\left(\frac{S}{C}-1\right)\lambda-1 & \left(\frac{S}{C}-1\right)\lambda & 0 & \dots & & \\ 0 & 1 & -\left(\frac{S}{C}-1\right)\lambda-1 & \left(\frac{S}{C}-1\right)\lambda & 0 & \dots & \\ & & \ddots & \ddots & \ddots & & \end{pmatrix}, \quad (25)$$

which is the infinitesimal generator of the process describing the number of customers in the $M/M/1$ queue with unit service rate and input rate

$$\gamma' = \left(\frac{S}{C} - 1\right)\lambda = \gamma - \frac{1-\gamma}{b-1}, \quad (26)$$

where b is the peak to mean rate coefficient of a source, which satisfies $b = (1+\lambda)/\lambda$, and γ is the link utilization factor defined by

$$\gamma = \frac{Sm}{C} = \frac{S}{bC} \quad (27)$$

with m denoting the mean rate of an individual source.

Similarly to the $M/M/\infty$ case, one may conjecture that the transient characteristics $C\theta$, CV, and N associated with the superposition of S On/Off sources can be approximated by the respective transient characteristics Θ, $\mathcal{V}$, and $\mathcal{N}$ of the $M/M/1$ queue with input rate γ' defined by equation (26) and unit service rate. In the following, we show via simulation that this conjecture seems to be valid for the superposition of a large number of On/Off sources.

For this purpose, we consider an ATM link of transmission capacity $c = 600$ Mbps, $S = 1400$ On/Off sources with an individual mean rate of about $m = 364$ Kbps and an individual peak bit rate of $h = 2$ Mbps, which result in a link utilization factor $\gamma = 85\%$ and a peak to mean ratio $b \sim 5.5$. Note that under the above assumptions, $C = 300$. An On/Off source is furthermore characterized by the mean silence and burst durations, denoted by ES and EB, respectively. In the following, we will assume that the mean burst duration is 1 ms so that the mean volume of information in a burst is 2 Kbits (256 bytes). The mean silence duration is given by

$$ES = (b-1)EB \sim 4.5 \text{ ms}. \quad (28)$$

Table 1
Mean values of random variables θ, V, and N.

On and Off duration distributions	$\bar{\theta}$	$\overline{V}$	$\overline{N}$
Exponential Off and On periods	1.69×10^{-2}	7.85×10^{-2}	5.20
Constant On periods and exponential Off periods	1.63×10^{-2}	7.85×10^{-2}	5.00
Exponential On periods and Constant Off periods	1.61×10^{-2}	7.77×10^{-2}	4.91
Hyperexponential ($Cv^2 = 5$) On and Off periods	1.91×10^{-2}	1.12×10^{-1}	5.81
Approximations	1.82×10^{-2}	9.93×10^{-2}	5.45

The stationary probability $P_{\text{cong.}}$ that the instantaneous input rate exceeds the link transmission capacity C is given by the survivor function of the Bernoulli distribution of parameter $1/b$, namely

$$P_{\text{cong.}} = \sum_{k=C+1}^{S} \binom{S}{k} \left(\frac{1}{b}\right)^{k} \left(1 - \frac{1}{b}\right)^{S-k} \sim 10^{-3}. \tag{29}$$

The instantaneous number of active sources is described by process $\{X_t\}$ and the instantaneous peak bit rate is $\{hX_t\}$. In the following, we consider the random variable V introduced in the previous sections so that the volume of information in bits lost in a congestion period is $v = hV$.

For different distributions of On and Off durations the mean values of random variables θ, V, and N, denoted by $\bar{\theta}$, $\overline{V}$, and $\overline{N}$, respectively are compared in table 1 with the respective theoretical mean values obtained via the approximation by the $M/M/1$ queue with input rate γ' and unit service time [10], namely

$$\bar{\theta}_{\text{approx}} = \frac{1}{C(1-\gamma')}, \tag{30}$$

$$\overline{V}_{\text{approx}} = \frac{1}{C(1-\gamma')^2}, \tag{31}$$

$$\overline{N}_{\text{approx}} = \frac{1}{(1-\gamma')}. \tag{32}$$

Results given in table 1 show that simulation and approximation results are in good agreement.

The distributions of $C\theta$, CV, and N for constant On and exponential Off period distributions are depicted in figures 3, 4, and 5, respectively. These figures show that the distributions of $C\theta$, CV, and N can be well approximated by those of variables Θ, $\mathcal{V}$, and $\mathcal{N}$ associated with the $M/M/1$ queue with input rate γ and unit service rate. Further simulation results can be found in [7].

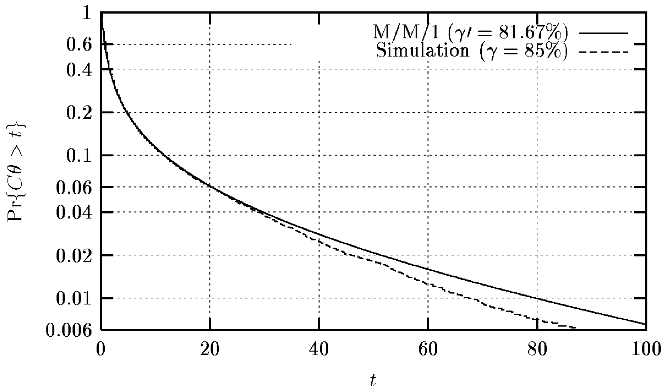

Figure 3. Distribution of $C\theta$ when On periods are constant and Off periods are exponentially distributed.

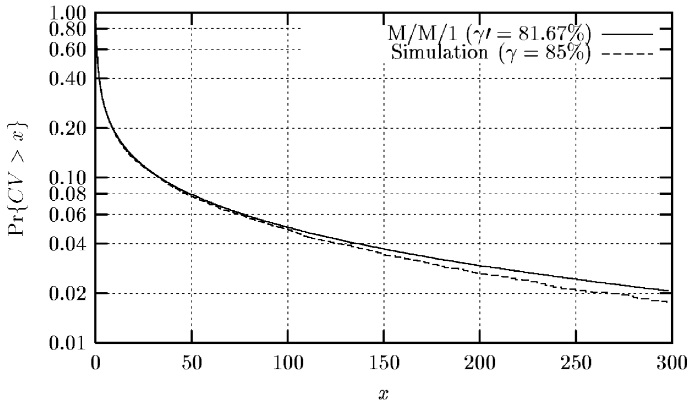

Figure 4. Distribution of CV when On periods are constant and Off periods are exponentially distributed.

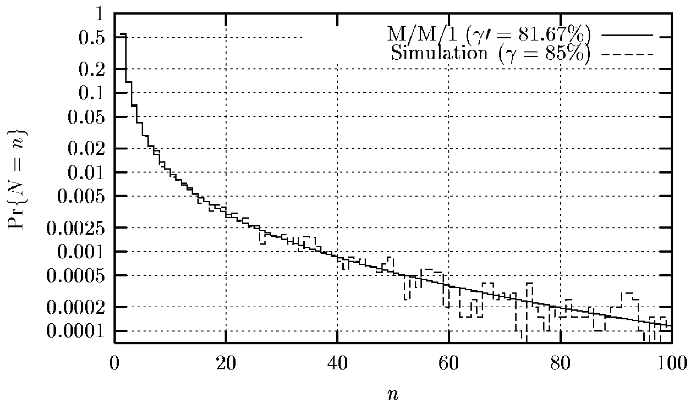

Figure 5. Distribution of N when On periods are constant and Off periods are exponentially distributed.

Now, it is worthwhile to note that the freeze-out fraction, which represents the fraction of lost information, is given by

$$\bar{\pi}_{\text{loss}} = \frac{\mathrm{E}[h(\Lambda_t - C)^+]}{\mathrm{E}[h\Lambda_t]} = \frac{P_{\text{cong.}}}{\rho C} \sim 4.6 \times 10^{-6} \tag{33}$$

and the mean volume $\bar{v}$ of information lost in a congestion period is

$$\bar{v} \sim h \times \frac{1}{C(1-\gamma')^2} \sim 296.3 \text{ Kbits} \sim 700 \text{ cells} \tag{34}$$

by using the convergence of CV to $\mathcal{V}$, whose mean value is given by $1/C(1-\gamma')^2$.

It follows that for a given low stationary congestion probability, the freeze-out fraction $\bar{\pi}_{\text{loss}}$ may take a small value but the volume of information lost in a congestion period may be large. This can be shown by considering the mean values. The situation is still worse when we consider remote quantiles. For instance, the $1 - 10^{-5}$-quantile of the distribution of $\mathcal{V}/C$ is equal to 42. It thus appears that the freeze-out fraction defined as the ratio of two long term average quantities gives only poor information on the quality of service actually offered to the users because in congestion periods a large amount of information is lost. This phenomenon is definitely not reflected by the value of the freeze-out fraction.

8. Conclusion

The $M/PH/\infty$ model has been introduced in this paper to model the superposition of a large number of data bursts on an ATM link. Taking benefit of the Markovian structure of the system, explicit methods have been described to compute the distributions of some transient characteristics related to excursions of the occupation process above the link transmission capacity C. Given that such excursions represent overflow periods under the unbuffered assumption, the transient characteristics considered are the congestion duration θ, the volume V of lost information, and the number N of bursts arriving in a congestion period.

Furthermore, still using the Markovian property of the system, some asymptotic results have been established on the behavior of $C\theta$, CV, and N when the transmission capacity C tends to infinity while the link utilization factor γ is fixed in $(0, 1)$. It turns out that $C\theta$, CV and N converge in distribution to the duration Θ of the busy period in the $M/M/1$ queue with unit service rate and input rate γ, the area $\mathcal{V}$ swept under the occupation process of this $M/M/1$ queue during a busy period, and the number $\mathcal{N}$ of customers served in a busy period of this $M/M/1$ queue, respectively. This shows that the local linearization conjecture stated by Aldous for $M/M/\infty$ queues holds for $M/PH/\infty$ queues. One may then reasonably suppose that this conjecture is also valid for $M/G/\infty$ type queues, given that PH distributions are dense in the set of probability distributions.

Further simulation results show that when considering the superposition of a large number of On/Off sources, $C\theta$, CV, and N can be quite accurately approximated, after adjustment of the input rate of the limiting $M/M/1$ queue, by Θ, $\mathcal{V}$, and $\mathcal{N}$, respectively. It thus appears that random variables Θ, $\mathcal{V}$, and $\mathcal{N}$ can be used to obtain robust estimates of some QoS parameters related to statistical multiplexing on an ATM link.

From this study, it turns out that the transient characteristics of a simple $M/M/1$ queue can be used to characterize statistical multiplexing of a large number of On/Off data sources on an ATM link. This approach notably allows the analysis of transient phenomena occurring in the case of congestion. The next step is to investigate how the local linearization property exhibited in this paper and the Poisson clumping heuristic of Aldous could be used to perform buffer dimensioning. Indeed, the Poisson clumping heuristic claims that congestion periods occur as a Poisson process with very small intensity when C tends to infinity. Then, one may conjecture that attaching to the link a buffer with a capacity equal to a remote quantile of the volume of information arriving in excess to the link transmission capacity, which can be approximated by $\mathcal{V}/C$, may be sufficient to significantly eliminate loss of information.

Appendix. Proofs of technical results

Proof of proposition 1. The non-zero entries of matrix $A_{C,C+1}$ are given for $s \in S_{C+1}$ and $i = 1, \dots, l$, by

$$A_{C,C+1}(s - e_i, s) = \lambda\beta(i)\mathbf{1}_{\{s_i \geqslant 1\}},$$

or equivalently, for $s \in S_C$ and $i = 1, \dots, l$, by

$$A_{C,C+1}(s, s + e_i) = \lambda\beta(i).$$

It follows that for every $s \in S_C$,

$$(A_{C,C+1}\mathbf{1})(s) = \sum_{i=1}^{l} A_{C,C+1}(s, s + e_i) = \lambda,$$

that is

$$A_{C,C+1}\mathbf{1} = \lambda\mathbf{1}.$$

We then get by relation (2)

$$\pi_C A_{C,C+1}\mathbf{1} = \lambda\pi_C\mathbf{1} = \lambda \mathrm{e}^{-\rho}\frac{\rho^C}{C!}.$$

Introducing $u_{C+1} = \pi_C A_{C,C+1}$, we have for every $s \in S_{C+1}$,

$$u_{C+1}(s) = \sum_{i=1}^{l} \pi_C(s - e_i)A_{C,C+1}(s - e_i, s) = \lambda\sum_{i=1}^{l}\beta(i)\pi_C(s - e_i)\mathbf{1}_{\{s_i \geqslant 1\}}.$$

From relation (1), we have for every $s \in S_{C+1}$ and $i = 1, \ldots, l$

$$\pi_C(s - e_i)\mathbf{1}_{\{s_i \geqslant 1\}} = \pi_{C+1}(s)\frac{s_i}{\rho_i},$$

that is

$$u_{C+1}(s) = \lambda \pi_{C+1}(s) \sum_{i=1}^{l} \frac{\beta(i)s_i}{\rho_i},$$

and hence, for every $s \in S_{C+1}$,

$$v_{C+1}(s) = \frac{\pi_{C+1}(s)\sum_{i=1}^{l}(\beta(i)s_i/\rho_i)}{\mathrm{e}^{-\rho}\rho^C/C!}.$$

Replacing now $\pi_{C+1}(s)$ by its expression given by relation (1), we obtain

$$v_{C+1}(s) = \frac{\left(\prod_{i=1}^{l}(\rho_i^{s_i}/s_i!)\right)\left(\sum_{i=1}^{l}(\beta(i)s_i/\rho_i)\right)}{\rho^C/C!},$$

which can be rewritten as in equation (4). This completes the proof. □

Proof of lemma 3. Let $s = (s_1, \ldots, s_l) \in S_{C+1}$, such that for $i = 1, \ldots, l$, $s_i/C \to r_i$ when $C \to \infty$. By definition, we have

$$\Pr\{CV > t \mid \Lambda_0 = s\} = \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s).$$

Let $n \geqslant 1$ be fixed. Starting from a state $s^{(n)} \in S_{C+n}$, process $\{\Lambda_t\}$ will be after $k \geqslant 0$ transitions in a state $s' = (s'_1, \ldots, s'_l)$ such that for $i = 1, \ldots, l$,

$$\max\left(0, s_i^{(n)} - k\right) \leqslant s'_i \leqslant s_i^{(n)} + k.$$

When $C \to \infty$, $s_i^{(n)}/C \to r_i$ and so $s'_i/C \to r_i$.

Let us define the tridiagonal block matrix H over subset B', whose non-zero transition rates are given for every $s \in S_{C+n}$ ($n \geqslant 1$) by

$$\begin{aligned} s &\to s + e_i && \text{with rate } \frac{\beta(i)\gamma}{n}, \\ s &\to s - e_i && \text{with rate } \frac{r_i\mu_{i,0}}{n}1_{\{r_i \geqslant 1\}}1_{\{n \geqslant 2\}}, \\ s &\to s + e_j - e_i && \text{with rate } \frac{r_i\mu_{i,j}}{n}1_{\{r_i \geqslant 1\}}1_{\{j \neq i\}}, \end{aligned} \tag{35}$$

where for every $i = 1, \ldots, l$, $r_i = \lim_{C\to\infty} s_i/C$.

From the definition of matrix M given in relation (7), we have for every $n \geqslant 1$, $s^{(n)} \in S_{C+n}$, and $s' \in B'$,

$$\frac{M(s, s')}{C} \to H\left(s, s'\right) \quad \text{when } C \to \infty.$$

It follows that for every $s \in S_{C+1}$, we have

$$\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s) \to \left(\mathrm{e}^{Ht}\mathbf{1}\right)(s) \quad \text{when } C \to \infty. \tag{36}$$

We denote by $H_{C+n,C+n-1}$ for $n \geqslant 2$, $H_{C+n,C+n}$ for $n \geqslant 1$, and $H_{C+n,C+n+1}$ for $n \geqslant 1$, the blocks of the tridiagonal matrix H. From the definition of matrix H given by relation (35), we have

$$\begin{aligned}
H_{C+n,C+n-1}\mathbf{1} &= \frac{\sum_{i=1}^{l} r_i\mu_{i,0}}{n}\mathbf{1},\\
H_{C+n,C+n}\mathbf{1} &= -\frac{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}}{n}\mathbf{1},\\
H_{C+n,C+n+1}\mathbf{1} &= \frac{\gamma}{n}\mathbf{1}.
\end{aligned}$$

Let us now define the infinite tridiagonal matrix Q_r over the set $\{1, 2, \ldots\}$ by

$$Q_r(1,1) = -\frac{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}}{n}, \qquad Q_r(1,2) = \frac{\gamma}{n},$$

and for $n \geqslant 2$,

$$Q_r(n,n-1) = \frac{\sum_{i=1}^{l} r_i\mu_{i,0}}{n}, \quad Q_r(n,n) = -\frac{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}}{n}, \quad Q_r(n,n+1) = \frac{\gamma}{n}.$$

It follows that for every $s \in S_{C+1}$, we have

$$\left(\mathrm{e}^{Ht}\mathbf{1}\right)(s) = d\mathrm{e}^{Q_r t}\mathbf{1},$$

where $d = (1, 0, 0, \ldots)$. We then obtain, from relation (36), for every $s \in S_{C+1}$,

$$\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s) \to d\mathrm{e}^{Q_r t}\mathbf{1} \quad \text{when } C \to \infty. \tag{37}$$

Noting that we precisely have (see, for instance, [9])

$$\Pr\left\{\mathcal{V}\left(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0}\gamma\right) > t\right\} = d\mathrm{e}^{Q_r t}\mathbf{1},$$

the proof is done. □

Proof of lemma 4. The maximum is reached for a value $s \in S_C$ denoted by $s^* = (s_1^*, \ldots, s_l^*)$. By definition of S_C, we have for $j = 1, \ldots, l$, s_j^*/C converges when $C \to \infty$. We denote by r_j^* the limit of s_j^*/C. We then have

$$\left|\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s^* + e_i) - \Pr\left\{\mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{s_i^*}{C}\mu_{j,0}\right) > t\right\}\right| \leqslant Z_1 + Z_2,$$

where

$$Z_1 = \left| \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s^* + e_i) - \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} r_j^* \mu_{j,0}\right) > t \right\} \right|$$

and

$$Z_2 = \left| \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} r_j^* \mu_{j,0}\right) > t \right\} - \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{s_j^*}{C} \mu_{j,0}\right) > t \right\} \right|.$$

From lemma 3, Z_1 tends to 0 when $C \to \infty$. Z_2 also tends to 0 when $C \to \infty$, due to the continuity of the function $\Pr\{\mathcal{V}(\gamma, y) > t\}$ with respect to $y \in [0, \infty[$. $\square$

Proof of theorem 5. First, note that by using relation (4) with $s = (s_1, \ldots, s_l)$, we can write

$$\begin{aligned}
\Pr\{CV > t\} &= v\mathrm{e}^{Mt/C}\mathbf{1} = \sum_{s \in S_{C+1}} v_{C+1}(s)\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s) \\
&= \sum_{s \in S_{C+1}} \sum_{i=1}^{l} \beta_i C! \frac{(\rho_i/\rho)^{s_i - 1}}{(s_i - 1)!} \mathbf{1}_{\{s_i > 0\}} \prod_{k \neq i} \frac{(\rho_k/\rho)^{s_k}}{s_k!} \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s) \\
&= \sum_{i=1}^{l} \beta_i \sum_{s \in S_{C+1}} C! \frac{(\rho_i/\rho)^{s_i - 1}}{(s_i - 1)!} \mathbf{1}_{\{s_i > 0\}} \prod_{k \neq i} \frac{(\rho_k/\rho)^{s_k}}{s_k!} \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s).
\end{aligned}$$

By the variable change $s_i \to s_i + 1$, we have

$$\Pr\{CV > t\} = \sum_{i=1}^{l} \beta_i \sum_{s \in S_C} \frac{C!}{s_1! \cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j} \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s + e_i). \tag{38}$$

From relation (38) and the fact that $\phi = 1$, we have

$$\begin{aligned}
&\Pr\{CV > t\} - \Pr\{\mathcal{V} > t\} \\
&\quad = \sum_{i=1}^{l} \beta_i \sum_{s \in S_C} \frac{C!}{s_1! \cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j} \left(\left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s + e_i) - \Pr\{\mathcal{V} > t\}\right).
\end{aligned}$$

The module of the above quantity is such that

$$\left|\Pr\{CV > t\} - \Pr\{\mathcal{V} > t\}\right| \leqslant \Sigma_1 + \Sigma_2,$$

where

$$\Sigma_1 = \sum_{i=1}^{l} \beta_i \sum_{s \in S_C} \frac{C!}{s_1! \cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j}$$

$$\times \left| \left(\mathrm{e}^{Mt/C}\mathbf{1}\right)(s+e_i) - \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{s_j}{C}\mu_{j,0}\right) > t \right\} \right|,$$

and

$$\Sigma_2 = \sum_{i=1}^{l} \beta_i \sum_{s\in S_C} \frac{C!}{s_1!\cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j} \left| \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{s_j}{C}\mu_{j,0}\right) > t \right\} - \Pr\{\mathcal{V} > t\} \right|.$$

Since

$$\sum_{i=1}^{l} \beta_i \sum_{s\in S_C} \frac{C!}{s_1!\cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j} = 1,$$

it is easy to show by using lemma 4 and Lebesgue dominated convergence theorem that the term Σ_1 tends to 0 as $C \to \infty$ under the assumptions of theorem 5.

To show that $\Sigma_2 \to 0$, we consider a continuous function ψ on $[0,1]^l$. Weierstrass's theorem states, in particular, that, if $x_1 + \cdots + x_l = 1$ then

$$\sum_{s\in S_C} \frac{C!}{s_1!\cdots s_l!} \prod_{j=1}^{l} x_j^{s_j} \psi\left(\frac{s_1}{C}, \ldots, \frac{s_l}{C}\right) \to \psi(x_1, \ldots, x_l) \quad \text{when } C \to \infty.$$

If we take

$$\psi(x_1, \ldots, x_l) = \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{i=1}^{l} x_i \mu_{i,0}\right) > t \right\},$$

which is continuous on $[0,1]^l$, we get

$$\sum_{s\in S_C} \frac{C!}{s_1!\cdots s_l!} \prod_{j=1}^{l} \left(\frac{\rho_j}{\rho}\right)^{s_j} \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{s_j}{C}\mu_{j,0}\right) > t \right\} \to \Pr\left\{ \mathcal{V}\left(\gamma, \sum_{j=1}^{l} \frac{\rho_j}{\rho}\mu_{i,0}\right) \right\}.$$

Now since by lemma 2

$$\sum_{j=1}^{l} \frac{\rho_j}{\rho}\mu_{i,0} = \phi = 1 \quad \text{and} \quad \sum_{i=1}^{l} \beta_i = 1,$$

we obtain by dominated convergence $\Sigma_2 \to 0$ when $C \to \infty$, and the proof is done. □

Proof of lemma 6. Let $s = (s_1, \ldots, s_l) \in S_{C+1}$ such that for $i = 1, \ldots, l$, $s_i/C \to r_i$ when $C \to \infty$. By definition, we have

$$\Pr\{N' = k \mid \Lambda_0 = s\} = \left(U_k(1)\right)(s). \tag{39}$$

Let us now define the tridiagonal block matrix F over subset B', whose non zero transition probabilities are given for every $s \in S_{C+n}$ ($n \geqslant 1$) by

$$\begin{aligned} s \to s + e_i \quad &\text{with probability } \frac{\beta(i)\gamma}{\gamma + \sum_{i=1}^{l} r_i\mu_i}, \\ s \to s - e_i \quad &\text{with probability } \frac{r_i\mu_{i,0}\mathbf{1}_{\{r_i>0\}}\mathbf{1}_{\{n\geqslant 2\}}}{\gamma + \sum_{i=1}^{l} r_i\mu_i}, \\ s \to s + e_j - e_i \quad &\text{with probability } \frac{r_i\mu_{i,j}\mathbf{1}_{\{r_i>0\}}\mathbf{1}_{\{j\neq i\}}}{\gamma + \sum_{i=1}^{l} r_i\mu_i}. \end{aligned} \tag{40}$$

From the definition of matrix $Q_{B'}$ given by relation (15), we have, for every $n \geqslant 1$, $s \in S_{C+n}$, and $s' \in B'$,

$$Q(s, s') \to F(s, s') \quad \text{when } C \to \infty.$$

It follows that $U_k(n) \to V_k(n)$, where $V_k(n)$ is given by $V_k(0) = U_k(0)$, $V_0(n) = U_0(n)$, $V_{-1}(n) = 0$, and

$$V_k(n) = (\mathbf{I} - F_{C+n,C+n})^{-1}\big(F_{C+n,C+n-1}V_k(n-1) + F_{C+n,C+n+1}V_{k-1}(n+1)\big). \tag{41}$$

From relation (40), matrix F satisfies

$$F_{C+n,C+n-1}\mathbf{1} = \frac{\sum_{i=1}^{l} r_i\mu_{i,0}}{\gamma + \sum_{i=1}^{l} r_i\mu_i}\mathbf{1},$$

$$F_{C+n,C+n}\mathbf{1} = \frac{\sum_{i=1}^{l} r_i(\mu_i - \mu_{i,0})}{\gamma + \sum_{i=1}^{l} r_i\mu_i}\mathbf{1},$$

$$F_{C+n,C+n+1}\mathbf{1} = \frac{\gamma}{\gamma + \sum_{i=1}^{l} r_i\mu_i}\mathbf{1},$$

and hence,

$$(\mathbf{I} - F_{C+n,C+n})^{-1}F_{C+n,C+n-1}\mathbf{1} = \frac{\sum_{i=1}^{l} r_i\mu_{i,0}}{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}}\mathbf{1},$$

$$(\mathbf{I} - F_{C+n,C+n})^{-1}F_{C+n,C+n+1}\mathbf{1} = \frac{\gamma}{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}}\mathbf{1}.$$

This implies by recurrence that, for k and n fixed, all the entries of the vector $V_k(n)$ are equal, that is $V_k(n)$ can be written as

$$V_k(n) = v_k(n)\mathbf{1},$$

where $v_k(n)$ is a real number. We then have from relation (39), for every $s \in S_{C+1}$,

$$\Pr\{N' = k \mid \Lambda_0 = s\} = \big(U_k(1)\big)(s) \to \big(V_k(1)\big)(s) = v_k(1) \quad \text{when } C \to \infty. \tag{42}$$

Recurrence relation (41) becomes

$$v_k(n) = \frac{\sum_{i=1}^{l} r_i\mu_{i,0}}{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}} v_k(n-1) + \frac{\gamma}{\gamma + \sum_{i=1}^{l} r_i\mu_{i,0}} v_{k-1}(n+1). \tag{43}$$

Consider now the $M/M/1$ queue with input rate γ and mean service rate $\sum_{i=1}^{l} r_i\mu_{i,0}$. Let Y_t be the number of customers in that queue at time t. Let $\mathcal{N}'(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0})$ denote the number of customers arriving during a sojourn of Y_t in the subset $\{1, 2, \ldots\}$. It is easy to verify (see for instance [9]) that the conditional probabilities $\Pr\{\mathcal{N}'(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0}) = k \mid Y_0 = n\}$ satisfy relation (43). Hence,

$$v_k(n) = \Pr\left\{\mathcal{N}'\left(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0}\right) = k \;\middle|\; Y_0 = n\right\}.$$

Thus, from relation (42), we get for every $s \in S_{C+1}$ and $k \geqslant 0$,

$$\Pr\{N' = k \mid \Lambda_0 = s\} \to \Pr\left\{\mathcal{N}'\left(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0}\right) = k \;\middle|\; Y_0 = 1\right\}$$

when $C \to \infty$ and hence, the proof is done by definition of $\mathcal{N}(\gamma, \sum_{i=1}^{l} r_i\mu_{i,0})$. $\square$

References

[1] D. Aldous, *Probability Approximations via the Poisson Clumping Heuristic* (Springer, Berlin, 1989).

[2] D. Anick, D. Mitra and M. Sondhi, Stochastic theory of a data-handling system with multiple sources, Bell Sys. Tech. J. 61(8) (1982) 1872–1894.

[3] P. Brémaud and F. Baccelli, *Palm Probabilities and Stationary Queues*, Lecture Notes in Statistics 41 (Springer, Berlin, 1987).

[4] G. Ciardo, R. Marie, B. Sericola and K.S. Trivedi, Performability analysis using semi-Markov reward processes, IEEE Trans. Computers 39 (1990) 1251–1264.

[5] B.T. Doshi, Deterministic rule based traffic descriptors for broadband ISDN: worst case behavior and connection acceptance control, in: *Proc. Globecom '93* (December 1993).

[6] N.G. Duffield and D.J. Daley, Bounds and comparison of the loss ratio in queues driven by an $M/M/\infty$ source, Report DIAS-APG-94-25, Dublin Institute for Advanced Studies (1994).

[7] A. Dupuis, F. Guillemin and B. Sericola, Asymptotic results for the superposition of a large number of data connections on an ATM link, Technical Report 3010, INRIA (September 1996).

[8] F. Guillemin, J. Boyer and A. Dupuis, Burstiness in broadband integrated networks, Performance Evaluation 15(3) (1992) 163–176.

[9] F. Guillemin, G. Rubino, B. Sericola and A. Simonian, Transient analysis of statistical multiplexing of data connections on an ATM link, in: *Proc. ITC '15*, Washington, DC (June 1997).

[10] F. Guillemin and A. Simonian, Transient characteristics of an $M/M/\infty$ system, Adv. Appl. Probab. 27 (1995) 862–888.

[11] V. Ramaswami and G. Latouche, Modeling packet arrivals from asynchronous input lines, in: *Proc. ITC '12*, Torino (June 1988).

[12] J. Roberts, Variable-bit-rate traffic control in B-ISDN, IEEE Communications Magazine (September 1991) 50–56.

[13] J. Roberts, ed., Performance evaluation and design of multiservice networks, Commission of the European Communities, Cost 224 Final Report (October 1992).
[14] S.M. Ross, *Stochastic Processes* (J. Wiley, 1983).
[15] G. Rubino and B. Sericola, Sojourn times in Markov processes, J. Appl. Probab. 27 (1989) 744–756.