A Quasi-Random Multirate Loss Model Supporting Elastic and Adaptive Traffic under the Bandwidth Reservation Policy

Ioannis D. Moscholios*, John S. Vardakas†, Michael D. Logothetis‡, and Michael N. Koukias‡
*Dept. of Informatics & Telecommunications, University of Peloponnese, 221 00 Tripolis, Greece.
Email: idm@uop.gr
†Iquadrat, Barcelona, Spain.
Email : jvardakas@iquadrat.com
‡Dept. of Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece.
Email: {mlogo, mkoukias}@upatras.gr

***Abstract*—In this paper, we propose a multirate teletraffic loss model of a single link that accommodates elastic and adaptive services whose calls come from a finite traffic-source population. This call arrival process is known as a quasi-random process and is used in traffic modelling when the number of users who generate traffic is relatively small compared to the system capacity. In-service elastic and adaptive calls can tolerate bandwidth compression by extending their remaining service-time (elastic calls) or not (adaptive calls). In this loss system, we study the effect of the bandwidth reservation policy on time congestion probabilities, call congestion probabilities and link utilization. The bandwidth reservation policy is considered when a certain quality of service for each service-class is required and is essential to be guaranteed. The proposed model does not have a product form solution, and therefore we propose approximate but recursive formulas for the efficient calculation of the above mentioned performance measures. The accuracy and consistency of the proposed model are verified by simulation and is found to be quite satisfactory. Finally, we generalize the proposed model to include calls from both finite and infinite number of traffic sources.**

Keywords-Markov chain; quasi-random process; elastic-adaptive traffic; recursive formula; time-call congestion; bandwidth reservation.

I. INTRODUCTION

IN contemporary communication networks, the traffic environment is composed mostly of multirate services of elastic and adaptive traffic. The co-existence of this type of services makes the call-level performance analysis and evaluation of modern telecom networks much more complicated and challenging. It is therefore essential to have proper multirate teletraffic loss models, for the call-level QoS assessment of such networks [1]. Based on analytical tools, telecom engineers can include various services in the network according to their offered traffic-load, decide on proper network dimensioning and avoid link over-dimensioning [2].

In [1], we present a multirate loss model for elastic and adaptive services with finite traffic-source population. By the term elastic traffic (e.g., file transfer), we refer to in-service calls that have the ability to compress their bandwidth, while simultaneously expanding their service time. On the other hand, adaptive traffic refers to in-service calls that tolerate bandwidth compression, but their service time does not alter (e.g., adaptive video) [3]. In both cases, we assume that bandwidth compression is permitted down to a minimum value. As far as the consideration of traffic sources with finite population is concerned, it results in a quasi-random call arrival process, which is, in many cases, a more realistic consideration than a random (Poisson) process (where infinite population of traffic sources is assumed). As the Markov chain analysis shows, the existence of the bandwidth compression/expansion mechanism destroys the Markov chain reversibility, and because of this, the model has no product form solution. Therefore, we propose approximate but recursive formulas for the efficient calculation of the call-level performance metrics, such as time and call congestion probabilities and link utilization. The consistency and the accuracy of the model are verified through simulation and found to be quite satisfactory.

In this paper, we extend [1], by studying the effect of the Bandwidth Reservation (BR) policy on time and call congestion probabilities, as well as on link utilization. The BR policy can achieve the equalization of call blocking probabilities among calls of different service-classes, or guarantee a certain QoS for each service-class. For instance, to equalize blocking probabilities between different service-classes, the BR policy ensures greater link bandwidth to high speed service-classes. Applications of the BR policy in wired (e.g., [4]-[6]), wireless (e.g., [7]-[9]) and optical networks (e.g., [10], [11]) show the policy's significant role in teletraffic engineering. As an example, in wireless networks the BR policy can ensure certain QoS for handoff traffic, while, it is worth mentioning that in optical networks, the term "bandwidth reservation" refers to "wavelength reservation" [10].

The Markov chain analysis is extensively used in the call-level analysis of communication networks. Contrary to the analysis of the complete sharing policy where the stationary probabilities have a product form solution, the BR policy cannot be analysed by the use of a product form solution. This is because one-way transitions appear in the state space of the Markov chain, which destroy the reversibility of the Markov chain [12]. Therefore, the reason why the proposed model in this paper does not have a product form solution is twofold:

not only the existence of the bandwidth compression/expansion mechanism but also the existence of the BR policy. Because of the absence of a product form solution we resort to an approximate solution, and propose a recursive formula for the efficient calculation of the link occupancy distribution. This formula simplifies the determination of all performance measures. We have evaluated the accuracy and consistency of the proposed model through simulation, and found them to be quite satisfactory. Herein, we name the model of [1], Extended Finite - Erlang Multirate Loss Model (EF-EMLM), because it is based on the classical EMLM (also known as Kaufman-Roberts model – more details are referred to the next Section) [13], [14]. Hence, the proposed new model is named EF-EMLM/BR. Furthermore, we generalize the EF-EMLM/BR to include calls from both finite and infinite number of traffic sources.

Potential applications of the proposed new model are mostly in the area of cellular networks, where calls come from finite sources and their bandwidth can be compressed (e.g., [15]–[20]). More precisely, a Wideband Code Division Multiple Access (W-CDMA) network, like Universal Mobile Telecommunications System (UMTS), supports not only streaming traffic of voice service, but also data traffic of an elastic or adaptive type (transferring messages or images) that can tolerate bandwidth compression. A single base station of this network can be modelled as a multirate loss system. The number of users in a cell is rather realistic to be considered finite, due to the limited coverage of a cell; this is even more realistic in the case of microcells. The BR policy can be applied to the system in order to reserve a part of the cell resources especially for handoff incoming traffic, which has to be serviced with a higher priority than the traffic originated inside the cell.

This paper is organised as follows: Section II contains related work. In Section III, we: a) present the basic assumptions and the call admission control, b) show the recursive formula for the link occupancy distribution and c) provide formulas for the various performance measures of the EF-EMLM. In Section IV, we: a) consider the application of the BR policy in the EF-EMLM, b) show the recursive formula for the link occupancy distribution, c) show how the EF-EMLM is related to other multirate loss models and d) provide formulas for the calculation of various performance measures of the EF-EMLM/BR. In Section V, we provide numerical results whereby the new model is compared to existing models and evaluated through simulation results. In Section VI, we generalize the proposed model to include calls from both finite and infinite number of sources. We conclude in Section VII. In Appendix A, we prove the recursive formula for the link occupancy distribution. Finally, we tabulate (as Appendix B) all the symbols used in this paper.

II. Related Work

Multirate teletraffic loss models of a single link that accommodates elastic and adaptive calls have been proposed in [3], [21]-[24]. In [3], the call arrival process is Poisson. A new call is accepted in the link with its peak bandwidth requirement when the occupied link bandwidth together with the peak bandwidth of that call does not exceed the capacity of the link. Otherwise, the new call is accepted in the link by compressing its peak-bandwidth, as well as the bandwidth of all in-service calls (of all service-classes). Call blocking occurs when the minimum bandwidth requirement of a call (achieved after the maximum bandwidth compression) is higher than the link's available bandwidth. The minimum bandwidth requirement of a call is a proportion of its peak-bandwidth; this proportion is common to all service-classes. When an in-service call departs from the system, then the remaining in-service calls, whose bandwidth has been compressed, expand their bandwidth in proportion to their peak-bandwidth requirement. The Markov chain analysis of this model shows that the bandwidth compression/expansion mechanism destroys the Markov chain reversibility, and therefore the model has no product form solution. However, according to [3], a reversible Markov chain that describes the model in an approximate way does exist, and leads to a recursive formula for the determination of link occupancy distribution and, consequently, call blocking probabilities and link utilization. This formula resembles the classical Kaufman-Roberts formula used in the EMLM, where Poisson arriving calls of different service-classes have fixed bandwidth requirements (stream traffic), and compete for the available link bandwidth under the complete sharing policy [13], [14]; thus, we name the model of [3], Extended EMLM (E-EMLM). In [21], the E-EMLM is extended to include retrials, i.e., blocked calls may retry one or more times to be serviced with reduced bandwidth. In [22], new calls, upon their arrival, may reduce their bandwidth according to the occupied link bandwidth. In [23], [24], the E-EMLM is further extended to include the Batched Poisson call arrival process, which is used to approximate arrival processes that are more "peaked"and "bursty" than the Poisson process.

In [1], we consider an extension of the E-EMLM, the EF-EMLM, whereby calls arrive in the link according to a quasi-random process [25]. The latter is smoother than the Poisson process and is used in traffic modelling when the number of users (sources) who generate traffic is finite. As an application example of the quasi-random process one may think of a microcell in a cellular network, where the number of users roaming in the cell's vicinity can be considered finite [19], [26]. Recently, a multirate loss model that includes Poisson calls of stream, elastic and adaptive traffic has been proposed in [20]; however, the presence of stream traffic prohibits the recursive calculation of link occupancy distribution or the other call-level performance measures.

III. The Extended Finite EMLM (EF-EMLM)

In the following subsections, we present basic assumptions for call admission control, the recursive calculation of the link occupancy distribution and the consequent calculation of the call-level performance measures of the EF-EMLM, under the complete sharing policy (without QoS guarantee).

A. Notation, basic assumptions and call admission

We study a link of capacity C bandwidth units that accommodates elastic and adaptive calls of K different service-classes. Let K_e and K_a be the set of elastic and adaptive service-classes ($K_e + K_a = K$), respectively. Calls of service-class k ($k = 1, \ldots, K$) come from a finite source population N_k and request b_k bandwidth units (peak-bandwidth requirement). The mean arrival rate of service-class k idle sources is $\lambda_k = (N_k - n_k)v_k$ where n_k is the number of in-service calls and v_k is the arrival rate per idle source. This call arrival process is a quasi-random process [25]. If $N_k \to \infty$ for $k = 1, \ldots, K$ and the total offered traffic-load remains constant, then a Poisson process arises. To introduce bandwidth compression, the occupied link bandwidth j may virtually exceed C up to T bandwidth units.

The description of call admission is based on a new service-class k call that arrives in the system when the occupied link bandwidth is j bandwidth units. Then:

i) If $j + b_k \leq C$, the call is accepted in the system with b_k bandwidth units for an exponentially distributed service time with mean μ_k^{-1}.

ii) If $j + b_k > T$, the call is blocked and lost.

iii) If $T \geq j + b_k > C$, the call is accepted in the system by compressing its peak-bandwidth requirement, as well as the assigned bandwidth of all in-service calls (of all service-classes). After compression has taken place, all calls share the C bandwidth units in proportion to their peak-bandwidth requirement, while the link operates at its full capacity C. This is the processor sharing discipline [27], [28].

When $T \geq j + b_k > C$, the compressed bandwidth b_k^{comp} of the newly accepted service-class k call, is calculated by:

$$b_k^{\text{comp}} = r b_k = \frac{C}{j'} b_k \tag{1}$$

where $r = \frac{C}{j'}$ denotes the compression factor and $j' = j + b_k$.

Since $j = \sum_{k=1}^{K} n_k b_k = \mathbf{nb}$, $\mathbf{n} = (n_1, n_{2,...,} n_K)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_K)$, the values of r are expressed by $r \equiv r(\mathbf{n}) = \frac{C}{\mathbf{nb}+b_k}$. The bandwidth of all in-service calls is also compressed by the same factor $r(\mathbf{n})$ and becomes equal to $b_i^{\text{comp}} = \frac{C}{j'} \cdot b_i$ for $i = 1, \ldots, K$. After bandwidth compression, the occupied link bandwidth is $j = C$. All adaptive calls do not alter their service time. On the other hand, all elastic calls increase their service time so that the product *service time* by *bandwidth* remains constant. The minimum bandwidth that a call of service-class k ($k = 1, \ldots, K$) tolerates, is:

$$b_{k,\min}^{\text{comp}} = r_{\min} b_k = \frac{C}{T} b_k \tag{2}$$

where $r_{\min} = \frac{C}{T}$ is the minimum proportion of the required peak-bandwidth and is common to all calls.

When an in-service call of service-class k, with bandwidth b_k^{comp}, departs from the system, then the remaining in-service calls of each service-class i ($i = 1, \ldots, K$), expand their bandwidth to b_i^{expan}, in proportion to their peak-bandwidth b_i:

$$b_i^{\text{expan}} = \min\left(b_i,\ b_i^{\text{comp}} + \frac{b_i}{\sum\limits_{k=1}^{K} n_k b_k} b_k^{\text{comp}} \right) \tag{3}$$

To illustrate the previous bandwidth compression mechanism consider the following simple example. Let $C = 3$ bandwidth units, $T = 5$ bandwidth units, $K = 2$ service-classes, $b_1 = 1$ bandwidth unit, $b_2 = 2$ bandwidth units, $N_1 = N_2 = 10$ sources, $v_1 = v_2 = 0.1$ and $\mu_1^{-1} = \mu_2^{-1} = 1$ time unit. The first service-class is elastic while the second service-class is adaptive. The permissible states $\mathbf{n} = (n_1, n_2)$ of the system are 12; they are presented in Table I together with the occupied link bandwidth, $j = n_1 b_1 + n_2 b_2$, before and after compression has been applied. Note that compression is applied if $T \geq j > C$ (bold values of the 3^{rd} column of Table I). After compression has been applied, we have that $j = C$ (bold values of the 4^{th} column of Table I). For example, assume that a new 2^{nd} service-class call arrives while the system is in state $(n_1, n_2) = (1, 1)$ and $j = C = 3$ bandwidth units. The new call is accepted in the system, since $j' = j + b_2 = T = 5$ bandwidth units, after bandwidth compression has been applied to all calls (new and in-service calls). The new state of the system is now $(n_1, n_2) = (1, 2)$. In this state, and based on (2), calls of the 1^{st} and 2^{nd} service-class compress their bandwidth to the following values:

$$b_{1,\min}^{\text{comp}} = r_{\min} b_1 = \tfrac{3}{5} b_1 = 0.6\ ,\ b_{2,\min}^{\text{comp}} = r_{\min} b_2 = \tfrac{3}{5} b_2 = 1.2$$

so that $j = n_1 b_{1,\min}^{\text{comp}} + n_2 b_{2,\min}^{\text{comp}} = 0.6 + 2.4 = 3 = C$. The value of μ_1^{-1} becomes $\frac{\mu_1^{-1}}{r_{\min}}$ so that $b_1 \mu_1^{-1}$ remains constant, while the value of μ_2^{-1} does not alter.

Consider now that the system is in state $(n_1, n_2) = (1, 2)$ and a 2^{nd} service-class call departs from the system. Then, its assigned bandwidth $b_{2,\min}^{\text{comp}} = 1.2$ is shared to the remaining calls in proportion to their peak-bandwidth requirement. Thus, in the new state $(n_1, n_2) = (1, 1)$ the 1^{st} service-class call expands its bandwidth to $b_1 = 1$ bandwidth unit and the 2^{nd} service-class call to $b_2 = 2$ bandwidth units. Thus, $j = n_1 b_1 + n_2 b_2 = C = 3$ bandwidth units. Furthermore, the service time of elastic calls only is decreased to $\mu_1^{-1} = 1$ time unit.

In Fig. 1, we present the state transition diagram of this example. If we consider the four adjacent states (n_1, n_2) : $(2, 0), (2, 1), (3, 1)$ and $(3, 0)$ and apply the Kolmogorov' s criterion (*flow clockwise = flow counter-clockwise*) [27], it is obvious that this criterion does not hold. This means that the Markov chain is not reversible. Generally speaking, the bandwidth compression mechanism destroys reversibility in the proposed model and therefore no PFS exists. To circumvent this problem, we use, in the following subsection, state dependent multipliers per service-class k, $\phi_k(\mathbf{n})$, which have a similar role with $r(\mathbf{n})$ and lead to a reversible Markov chain.

B. Determination of link occupancy distribution

Let $\mathbf{\Omega}$ be the system's state space $\mathbf{\Omega} = \{\mathbf{n} : 0 \leq \mathbf{nb} \leq T\}$. The fact that the system cannot be described by a reversible Markov chain means that local balance does not exist between adjacent states of $\mathbf{\Omega}$. Therefore, the steady-state distribution $P(\mathbf{n})$ does not have a product form solution. To derive an approximate but recursive formula for

TABLE I. STATE SPACE AND OCCUPIED LINK BANDWIDTH

n_1	n_2	j (before compression) $0 \leq j \leq T$	j (after compression) $0 \leq j \leq C$
0	0	0	0
0	1	2	2
0	2	**4**	**3**
1	0	1	1
1	1	3	3
1	2	**5**	**3**
2	0	2	2
2	1	**4**	**3**
3	0	3	3
3	1	**5**	**3**
4	0	**4**	**3**
5	0	**5**	**3**

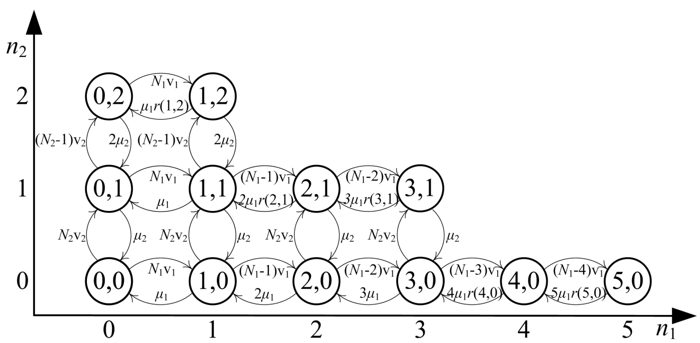

Figure 1. State transition diagram of the tutorial example.

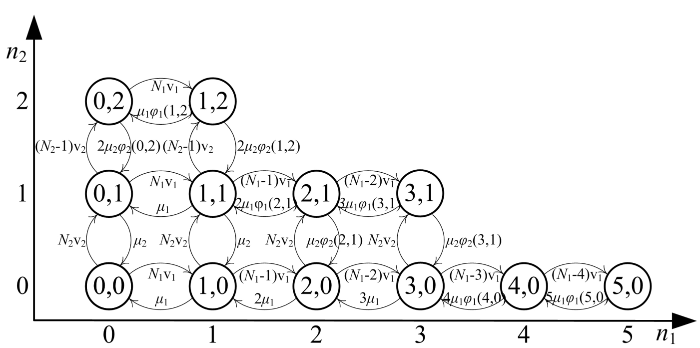

Figure 2. Modified state transition diagram of the tutorial example.

the efficient calculation of the link occupancy distribution, $G(j),\ j = 0, 1, \ldots, T$, we construct a reversible Markov chain that approximates the system by using state multipliers for all states $\mathbf{n} \in \mathbf{\Omega}$. The local balance equations between the adjacent states $\mathbf{n}_k^{-1} = (n_1, n_2, \ldots, n_k - 1, \ldots, n_K)$ and $\mathbf{n} = (n_1, n_2, \ldots, n_k, \ldots, n_K)$ have the form:

$$P(\mathbf{n}_k^{-1})(N_k - n_k + 1)v_k = P(\mathbf{n})\phi_k(\mathbf{n})\mu_k n_k\ ,\ k \in K_e \quad (4)$$

$$P(\mathbf{n}_k^{-1})(N_k - n_k + 1)v_k = P(\mathbf{n})\phi_k(\mathbf{n})\mu_k n_k\ ,\ k \in K_a \quad (5)$$

where $\phi_k(\mathbf{n})$ is a state-dependent multiplier and is defined as:

$$\phi_k(\mathbf{n}) = \begin{cases} 1 & ,\ \text{when } \mathbf{nb} \leq \mathrm{C} \text{ and } \mathbf{n} \in \mathbf{\Omega} \\ \frac{x(\mathbf{n}_k^{-1})}{x(\mathbf{n})}, & \text{when } C < \mathbf{nb} \leq T \text{ and } \mathbf{n} \in \mathbf{\Omega} \\ 0 & ,\ \text{otherwise} \end{cases} \quad (6)$$

and

$$x(\mathbf{n}) = \begin{cases} 1 & ,\ \text{when } \mathbf{nb} \leq C,\ \mathbf{n} \in \mathbf{\Omega} \\ \frac{1}{C}\left(\sum\limits_{k \in K_e} n_k b_k x(\mathbf{n}_k^{-1}) + r(\mathbf{n}) \sum\limits_{k \in K_a} n_k b_k x(\mathbf{n}_k^{-1})\right) & ,\ \text{when } C < \mathbf{nb} \leq T,\ \mathbf{n} \in \mathbf{\Omega} \\ 0 & ,\ \text{otherwise} \end{cases} \quad (7)$$

where $r(\mathbf{n}) = \frac{C}{\mathbf{nb}}$.

When $C < \mathbf{nb} \leq T$ and $\mathbf{n} \in \mathbf{\Omega}$ the bandwidth of all in-service calls is compressed by a factor $\phi_k(\mathbf{n})$ so that:

$$\sum_{k \in K_e} n_k b_k^{\text{comp}} + \sum_{k \in K_a} n_k b_k^{\text{comp}} = C \quad (8)$$

To derive (7), we keep the product *service time* by *bandwidth* of service-class k calls (elastic or adaptive) in state $\mathbf{n}$ of the irreversible Markov chain equal to the corresponding product in the same state $\mathbf{n}$ of the reversible Markov chain. This means that:

$$\frac{b_k r(\mathbf{n})}{\mu_k r(\mathbf{n})} = \frac{b_k^{\text{comp}}}{\mu_k \phi_k(\mathbf{n})} \Rightarrow b_k^{\text{comp}} = b_k \phi_k(\mathbf{n}),\ k \in K_e \quad (9)$$

and

$$\frac{b_k r(\mathbf{n})}{\mu_k} = \frac{b_k^{\text{comp}}}{\mu_k \phi_k(\mathbf{n})} \Rightarrow b_k^{\text{comp}} = b_k \phi_k(\mathbf{n}) r(\mathbf{n}),\ k \in K_a \quad (10)$$

Equation (7) results by substituting (9), (10) and (6), into (8).

Figure 2 shows the modified state transition diagram of our tutorial example, due to the introduction of the state-dependent multipliers $\phi_k(\mathbf{n})$'s. Now, if we consider again the four adjacent states $(n_1, n_2) : (2,0), (2,1), (3,1)$ and $(3,0)$, the Kolmogorov' s criterion (*flow clockwise = flow counter-clockwise*) holds, since:

$$\begin{aligned} & N_2 v_2 (N_1 - 2) v_1 \mu_2 \phi_2(3,1) 3\mu_1 = \\ & = (N_1 - 2) v_1 N_2 v_2 3\mu_1 \phi_1(3,1) \mu_2 \phi_2(2,1) \Rightarrow \\ & \phi_2(3,1) = \phi_1(3,1)\phi_2(2,1) \Rightarrow \\ & x(3,0) = x(2,0) = 1 \end{aligned}$$

In order to prove a recursive formula for the calculation of $G(j)$'s, we consider two cases: i) states where $0 \leq j \leq C$ (bandwidth compression does not occur) and ii) states where $C < j \leq T$ (bandwidth compression occurs).

When $0 \leq j \leq C$, then $\phi_k(\mathbf{n}) = 1$ and based on (4) and (5), it is proved that [29]:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{j} \sum\limits_{k \in K} (N_k - n_k + 1)\alpha_k b_k G(j - b_k) & \\ & \text{for } j = 1, \dots, C \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where: $\alpha_k = \frac{v_k}{\mu_k}$ is the offered traffic-load (in erl) per idle source of service-class k.

Note that (11) comes in fact from the Engset multirate loss model [29], in which calls come from finite sources and compete for the available link bandwidth under the complete sharing policy. The name "Engset" is justified by the fact that for $K = 1$ service-class, (11) can be used to calculate time congestion probabilities (by taking the sum of $G(j)$'s over all blocking states) whose values are the same with the classical Engset formula [30].

When $C < j \leq T$, it can be proved (see Appendix A) that:

$$\sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k G(j - b_k) + \frac{C}{j} \sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k G(j - b_k) = CG(j) \quad (12)$$

The combination of (11) and (12) gives the approximate recursive formula of *G(j)*'s, when $1 \leq j \leq T$:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{\min(j,C)} \sum\limits_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k G(j - b_k) & \\ + \frac{1}{j} \sum\limits_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k G(j - b_k) & \\ & \text{for } j = 1, \dots, T \\ 0 & \text{for } j < 0 \end{cases} \quad (13)$$

C. Relation of the EF-EMLM to other multirate loss models

(a) When $N_k \to \infty$ for $k = 1, \dots, K$ and the total offered traffic-load remains constant, then the call arrival process is Poisson and the E-EMLM results. In that case, the formula of $G(j)$'s is given by [3]:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{\min(j,C)} \sum\limits_{k \in K_e} \alpha_{k,\text{inf}} b_k G(j - b_k) + & \\ \frac{1}{j} \sum\limits_{k \in K_a} \alpha_{k,\text{inf}} b_k G(j - b_k) & \\ & \text{for } j = 1, \dots, T \\ 0 & \text{for } j < 0 \end{cases} \quad (14)$$

where $\alpha_{k,\text{inf}} = \frac{\lambda_{k,\text{inf}}}{\mu_k}$ (in erl) and $\lambda_{k,\text{inf}}$ is the arrival rate of calls of service-class k.

(b) When $N_k \to \infty$ for $k = 1, \dots, K$ and the total offered traffic-load remains constant, and $T = C$, then no bandwidth compression is allowed and the classical EMLM arises. In that case, the formula of $G(j)$'s is given by the Kaufman-Roberts recursion [13], [14]:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{j} \sum\limits_{k=1}^{K} \alpha_{k,\text{inf}} b_k G(j - b_k) & \text{for } j = 1, \dots, C \\ 0 & \text{for } j < 0 \end{cases} \quad (15)$$

D. Determination of various performance measures

The calculation of $G(j)$'s in (13) requires the value of n_k which is unknown. In other finite multirate loss models (e.g., [29]–[31]) there exist calculation methods for the determination of n_k in each state j through the use of an equivalent stochastic system, with the same traffic description parameters and exactly the same set of states. However, the state space determination of the equivalent system is complex, especially for large capacity systems that serve many service-classes. Thus, we avoid such methods and approximate n_k in state j, $n_k(j)$, as the mean number of service-class k calls in state j, $y_k(j)$, when Poisson arrivals are considered, i.e., $n_k(j) \approx y_k(j)$. Such approximations are common in the literature and induce little error (e.g., [32]–[34]). The values of $y_k(j)$ are given by (16) and (17) in the case of elastic and adaptive service-classes, respectively [3]:

$$\begin{aligned} & y_k(j)G(j) = \\ & = \frac{1}{\min(C,j)} \alpha_{k,\text{inf}} b_k G(j - b_k)\left(y_k(j - b_k) + 1\right) + \\ & + \frac{1}{\min(C,j)} \sum_{i=1 \wedge i \neq k}^{K_e} \alpha_{i,\text{inf}} b_i G(j - b_i) y_k(j - b_i) + \\ & + \frac{1}{j} \sum_{i=1}^{K_\alpha} \alpha_{i,\text{inf}} b_i G(j - b_i) y_k(j - b_i) \end{aligned} \quad (16)$$

$$\begin{aligned} & y_k(j)G(j) = \\ & = \frac{1}{j} \alpha_{k,\text{inf}} b_k G(j - b_k)(y_k(j - b_k) + 1) + \\ & + \frac{1}{j} \sum_{i=1 \wedge i \neq k}^{K_a} \alpha_{i,\text{inf}} b_i G(j - b_i) y_k(j - b_i) + \\ & + \frac{1}{\min(C,j)} \sum_{i=1}^{K_e} \alpha_{i,\text{inf}} b_i G(j - b_i) y_k(j - b_i) \end{aligned} \quad (17)$$

where the values of $G(j)$'s are calculated by (14).

Having determined $G(j)$'s according to (13), we calculate the following performance measures:

1) The time congestion probabilities of service-class k, denoted as P_{b_k}, which is the probability that at least $T - b_k + 1$ bandwidth units are occupied:

$$P_{b_k} = \sum_{j=T-b_k+1}^{T} G^{-1} G(j) \quad (18)$$

where: $G = \sum\limits_{j=0}^{T} G(j)$ is a normalization constant.

Time congestion probabilities are determined by the proportion of time the system is congested.

2) The call congestion probabilities of service-class k, denoted as C_{b_k}, which is the probability that a new service-class k call is blocked and lost:

$$C_{b_k} = \sum_{j=T-b_k+1}^{T} G^{-1}G(j) \quad (19)$$

where $G(j)$'s are determined for a system with $N_k - 1$ traffic sources.
Call congestion probabilities are determined by the proportion of arriving calls that find the system congested.

3) The link utilization, denoted as U:

$$U = \sum_{j=1}^{C} jG^{-1}G(j) + \sum_{j=C+1}^{T} CG^{-1}G(j) \quad (20)$$

IV. The EF-EMLM Under The BR Policy (EF-EMLM/BR)

In the following subsections, we describe the EF-EMLM/BR, we provide the recursive determination of the link occupancy distribution, we show the relation of the EF-EMLM/BR to other multirate loss models and determine the various call-level performance measures of this model, which assures QoS guarantee (regarding congestion probabilities).

A. Description of the proposed EF-EMLM/BR

The BR policy is used to guarantee a certain QoS for calls of each service-class or attain equalization of call blocking probabilities among different service-classes that share a link by a proper selection of the BR parameters. If, for example, equalization of call blocking probabilities is required between calls of two service-classes with $b_1 = 1$ and $b_2 = 10$ bandwidth units, respectively, then $t(1) = 9$ bandwidth units and $t(2) = 0$ bandwidth units so that $b_1+t(1) = b_2+t(2)$. Note that $t(1) = 9$ bandwidth units means that 9 bandwidth units are reserved to benefit calls of the 2^{nd} service-class. Similarly, if a link accommodates calls of three service-classes with $b_1 = 1, b_2 = 5$ and $b_3 = 10$ bandwidth units, respectively, and equalization of call blocking probabilities is required between calls of the first two service-classes, then $t(1) = 4$ and $t(2) = 0$ bandwidth units so that $b_1 + t(1) = b_2 + t(2)$.

The application of the BR policy in a single link multirate loss model is based on the assumption that the number of calls of certain service-classes is negligible in those states j that form the so-called reservation space. More precisely, in the proposed EF-EMLM/BR the number of service-class k calls is negligible in states $j > T - t(k)$ and is incorporated in the calculation of $G(j)$'s (see (21) below) by the variable $D_k(j - b_k)$ given in (22). Generally speaking, the population of calls of service-class k in the reservation space may not be negligible. In [35] and [36], a complex procedure is implemented in order to take into account this population and increase the accuracy of call blocking probability results in the EMLM and Engset multirate state-dependent loss models, respectively. However, according to [36] this procedure may not always increase the accuracy of the call blocking probability results compared to simulation.

B. Determination of the link occupancy distribution

To apply the BR policy to the EF-EMLM we consider the method of [37]. In that case, the formula for the approximate calculation of $G(j)$ takes the form:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(j,C)}\sum\limits_{k\in K_e}(N_k-n_k+1)\alpha_k D_k(j-b_k)G(j-b_k) & \\ +\frac{1}{j}\sum\limits_{k\in K_a}(N_k-n_k+1)\alpha_k D_k(j-b_k)G(j-b_k) & \\ & \text{for } j=1,\ldots,T \\ 0 & \text{for } j<0 \end{cases} \quad (21)$$

where:

$$D_k(j-b_k) = \left\langle \begin{array}{ll} b_k & \text{for } j \le T - t(k) \\ 0 & \text{for } j > T - t(k) \end{array} \right\rangle \quad (22)$$

and $t(k)$ is the reserved bandwidth (BR parameter) for service-class k calls (elastic or adaptive).

C. Relation of the EF-EMLM/BR to other multirate loss models

(a) When $N_k \to \infty$ for $k = 1, \ldots, K$, and the total offered traffic-load remains constant, then we have the Poisson arrival process and the formula of $G(j)$'s is given by [38]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(j,C)}\sum\limits_{k\in K_e}\alpha_k D_k(j-b_k)G(j-b_k)+ & \\ +\frac{1}{j}\sum\limits_{k\in K_a}\alpha_k D_k(j-b_k)G(j-b_k) & \\ & \text{for } j=1,\ldots,T \\ 0 & \text{for } j<0 \end{cases} \quad (23)$$

where $\alpha_k = \frac{\lambda_k}{\mu_k}$ (in erl), λ_k is the arrival rate of calls of service-class k and the values of $D_k(j)$ are given by (22).

(b) When all service-classes are elastic and $N_k \to \infty$ for $k = 1, \ldots, K$ and the total offered traffic-load remains constant, then the formula of $G(j)$'s is determined by [39]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(j,C)}\sum\limits_{k\in K}\alpha_k D_k(j-b_k)G(j-b_k) & \\ & \text{for } j=1,\ldots,T \\ 0 & \text{for } j<0 \end{cases} \quad (24)$$

(c) When $N_k \to \infty$ for $k = 1, \ldots, K$, and the total offered traffic-load remains constant, and $T = C$, then no bandwidth compression is allowed and the EMLM under the BR policy results. In that case, the formula of $G(j)$'s is given by the Roberts'recursion [37]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k=1}^{K}\alpha_{k,\inf} D_k(j-b_k)G(j-b_k) & \\ & \text{for } j=1,\ldots,C \\ 0 & \text{for } j<0 \end{cases} \quad (25)$$

D. Determination of various performance measures

The calculation of $G(j)$'s in (21) requires the value of n_k which is unknown. We approximate n_k in state j, $n_k(j)$, as the mean number of service-class k calls in state j, $y_k(j)$, when Poisson arrivals are considered, i.e., $n_k(j) \approx y_k(j)$. The values of $y_k(j)$ are given by (26) and (27) in the case of elastic and adaptive service-classes, respectively:

$$y_k(j)G(j) = \frac{1}{\min(C,j)}\alpha_{k,\inf}D_k(j-b_k)G(j-b_k)\left(y_k(j-b_k)+1\right) + \frac{1}{\min(C,j)}\sum_{i=1 \wedge i \neq k}^{K_e}\alpha_{i,\inf}D_i(j-b_i)G(j-b_i)y_k(j-b_i) + \frac{1}{j}\sum_{i=1}^{K_\alpha}\alpha_{i,\inf}D_i(j-b_i)G(j-b_i)y_k(j-b_i) \quad (26)$$

$$y_k(j)G(j) = \frac{1}{j}\alpha_{k,\inf}D_k(j-b_k)G(j-b_k)\left(y_k(j-b_k)+1\right) + \frac{1}{j}\sum_{i=1 \wedge i \neq k}^{K_a}\alpha_{i,\inf}D_i(j-b_i)G(j-b_i)y_k(j-b_i) + \frac{1}{\min(C,j)}\sum_{i=1}^{K_e}\alpha_{i,\inf}D_i(j-b_i)G(j-b_i)y_k(j-b_i) \quad (27)$$

where the values of $D_k(j)$ are given by (22) and the values of $G(j)$'s by (23).

Note that in (26) and (27), the mean number of service-class k calls in state j, $y_k(j) = 0$, if $j > T - t(k)$ as it is implied by (22).

Having determined the values of $G(j)$'s according to (21), we can calculate the link utilization based on (20) and the time and call congestion probabilities as follows:

$$P_{b_k} = \sum_{j=T-b_k-t_k+1}^{T} G^{-1}G(j) \quad (28)$$

$$C_{b_k} = \sum_{j=T-b_k-t_k+1}^{T} G^{-1}G(j) \quad (29)$$

where: $G = \sum_{j=0}^{T} G(j)$ is a normalization constant.

Note that, in order to calculate the call congestion probabilities of service-class k, $G(j)$'s, the system should be determined with $N_k - 1$ traffic sources – hence, the similarity between (28) and (29).

V. EVALUATION

In this section, we present an application example and compare the analytical results of the Time Congestion (TC) probabilities and link utilization obtained from the E-EMLM, the EF-EMLM and the EF-EMLM/BR. The corresponding simulation results, presented for the EF-EMLM and the EF-EMLM/BR, are mean values of 6 runs. The resultant reliability ranges of the simulation measurements (confidence intervals of 95%) are very small (less than two orders of magnitude) and, therefore, they are not presented. Simulation is based on Simscript III simulation language [40].

We consider a single link of capacity $C = 90$ bandwidth units (b.u.) that accommodates calls of three service-classes. The first two service-classes are elastic, while the third service-class is adaptive. The traffic characteristics of each service-class of the EF-EMLM are:

- 1^{st} service-class: $N_1 = 200,\ v_1 = 0.10,\ b_1 = 1$ b.u.
- 2^{nd} service-class: $N_2 = 200,\ v_2 = 0.04,\ b_2 = 4$ b.u.
- 3^{rd} service-class: $N_3 = 200,\ v_3 = 0.01,\ b_3 = 6$ b.u.

In the case of the E-EMLM, the corresponding Poisson traffic loads are: $\alpha_{1,\inf} = 20$ erl, $\alpha_{2,\inf} = 8$ erl and $\alpha_{3,\inf} = 2$ erl. We also consider two values of T:

a) $T = 90$ b.u., where no bandwidth compression takes place, and
b) $T = 100$ b.u., where bandwidth compression takes place and $r_{\min} = \frac{C}{T} = 0.9$.

In the case of the EF-EMLM/BR model, we choose $t(1)=5$, $t(2)=2$ and $t(3)=0$ in order to achieve blocking equalization between calls of all service-classes since: $b_1+t(1)=b_2+t(2)=b_3+t(3)$. In the x-axis of all figures, v_1 and v_2 increase in steps of 0.01 and 0.005 erl, respectively, while v_3 remains constant. So in Point 1 we have $(v_1, v_2, v_3) = (0.10, 0.04, 0.01)$, while in Point 6 $(v_1, v_2, v_3) = (0.15, 0.065, 0.01)$. In the case of the E-EMLM, the corresponding Poisson traffic loads in Point 1 and Point 8 are $(\alpha_{1,\inf}, \alpha_{2,\inf}, \alpha_{3,\inf}) = (20, 8, 2)$ and $(\alpha_{1,\inf}, \alpha_{2,\inf}, \alpha_{3,\inf}) = (30, 13, 2)$, respectively.

In Figs. 3-4, we present the analytical and the simulation TC probabilities of the 1^{st} service-class for $T = 90$ b.u. and $T = 100$ b.u., respectively. To better compare the corresponding TC probabilities results (while having numerical values), we present in Table II and Table III, only for Points 1 and 6, an excerpt of the results of Fig. 3 and Fig. 4, respectively. Similarly, in Figs. 5-6 and 7-8, we present the corresponding results of the 2^{nd} and 3^{rd} service-class. In the legend of all figures, the term $N = \inf.$ refers to the E-EMLM where the number of traffic sources is infinite for each service-class. Likewise, the term BR in all figures refers to the EF-EMLM/BR. Note that the call congestion probabilities of service-class k $(k = 1, 2, 3)$ are quite close to the corresponding TC probabilities of service-class k (since they are obtained for a system with $N_k - 1 = 199$ traffic sources) and, therefore, are not presented herein. The interested reader may resort to [1], where call congestion probabilities are presented for the EF-EMLM.

All figures show that:

i) analytical and simulation results of TC probabilities are very close to each other,
ii) the application of the compression/expansion mechanism reduces congestion probabilities compared to those obtained when $C = T = 90$ b.u. (compare, e.g., Figs. 3-4, Figs. 5-6 and Figs. 7-8),
iii) the co-existence of the BR policy and the compression/expansion mechanism reduces congestion probabilities compared to those obtained when $C = T = 90$ b.u.,

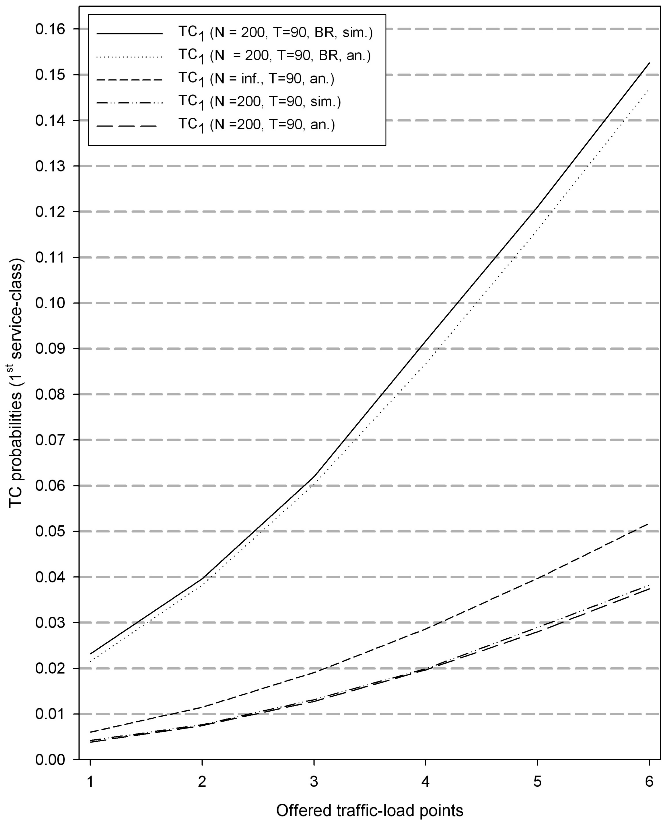

Figure 3. Time congestion probabilities of the 1^{st} service-class (T=90 b.u.).

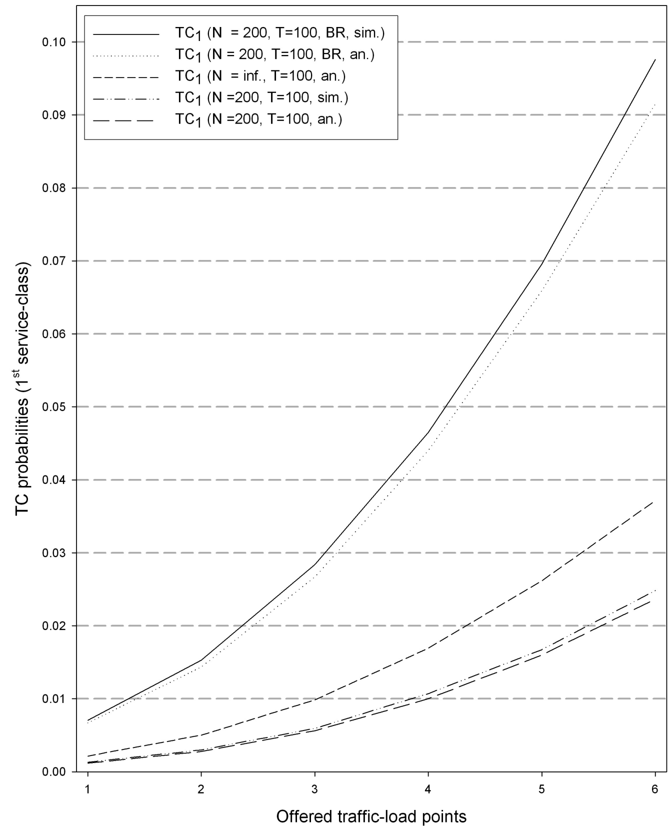

Figure 4. Time congestion probabilities of the 1^{st} service-class (T=100 b.u.).

iv) the congestion probabilities obtained by the EF-EMLM/BR and the EF-EMLM show that the BR policy favors calls of the 3^{rd} service-class, as expected, and
v) the results obtained by the E-EMLM fail to approximate the corresponding results obtained by the EF-EMLM.

TABLE II. EXCERPT OF THE RESULTS OF FIG. 3

$T=90$	TC_1 $N=200$ BR, sim.	TC_1 $N=200$ BR, anal.	TC_1 $N=\mathrm{inf.}$ anal.	TC_1 $N=200$ sim.	TC_1 $N=200$ anal.
Point 1	0.02320	0.02153	0.00604	0.00420	0.00385
Point 6	0.15260	0.14692	0.05178	0.03822	0.03743

TABLE III. EXCERPT OF THE RESULTS OF FIG. 4

$T=$ 100	TC_1 $N=200$ BR, sim.	TC_1 $N=200$ BR, anal.	TC_1 $N=\mathrm{inf.}$ anal.	TC_1 $N=200$ sim.	TC_1 $N=200$ anal.
Point 1	0.00704	0.00668	0.00215	0.00125	0.00117
Point 6	0.09760	0.09151	0.03716	0.02480	0.02367

Finally, in Figs. 9-10, we present the analytical and simulation results of the link utilization (in b.u.) for $T = 90$ and $T = 100$, respectively. It is clear, that the application of the bandwidth compression/expansion mechanism increases link utilization, since it decreases call congestion probabilities.

VI. GENERALIZATION OF THE EF-EMLM

Consider a link of capacity *C* b.u. that accommodates calls of finite and infinite number of sources. Let $K_{e,\mathrm{inf}}$ and $K_{a,\mathrm{inf}}$ be the set of elastic and adaptive service-classes ($K_{e,\mathrm{inf}} + K_{a,\mathrm{inf}} = K_{\mathrm{inf}}$) whose calls arrive in the link according to a Poisson process. Similarly, let $K_{e,\mathrm{fin}}$ and $K_{a,\mathrm{fin}}$ be the set of elastic and adaptive service-classes ($K_{e,\mathrm{fin}} + K_{a,\mathrm{fin}} = K_{\mathrm{fin}}$) whose calls arrive in the link according to a quasi-random process. Then, the calculation of the link occupancy distribution, *G*(*j*), can be done by the following formula:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(j,C)}\sum\limits_{k\in K_{e,\mathrm{fin}}}(N_k-n_k+1)\alpha_k b_k G(j-b_k) & \\ +\frac{1}{j}\sum\limits_{k\in K_{a,\mathrm{fin}}}(N_k-n_k+1)\alpha_k b_k G(j-b_k) & \\ +\frac{1}{\min(j,C)}\sum\limits_{k\in K_{e,\mathrm{inf}}}\alpha_{k,\mathrm{inf}} b_k G(j-b_k) & \\ +\frac{1}{j}\sum\limits_{k\in K_{a,\mathrm{inf}}}\alpha_{k,\mathrm{inf}} b_k G(j-b_k) & \\ & \text{for } j=1,\ldots,\mathrm{T} \\ 0 & \text{for } j<0 \end{cases} \tag{30}$$

The proof of (30) is based on the combination of the proofs proposed for the corresponding finite model in [1] and the infinite model of [3], and therefore is omitted. Such a mixture of service-classes does not destroy the accuracy of the model,

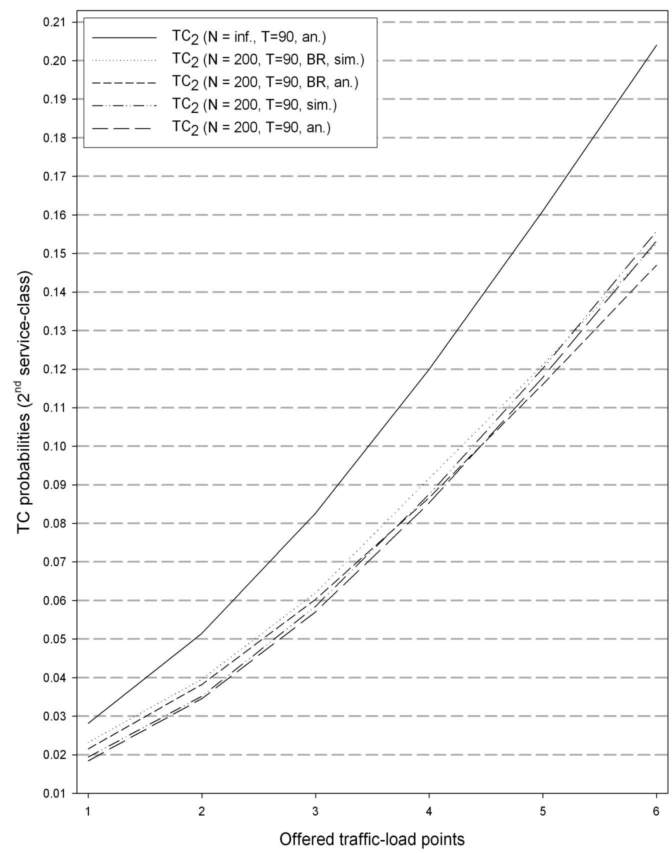

Figure 5. Time congestion probabilities of the 2^{nd} service-class (T=90 b.u.).

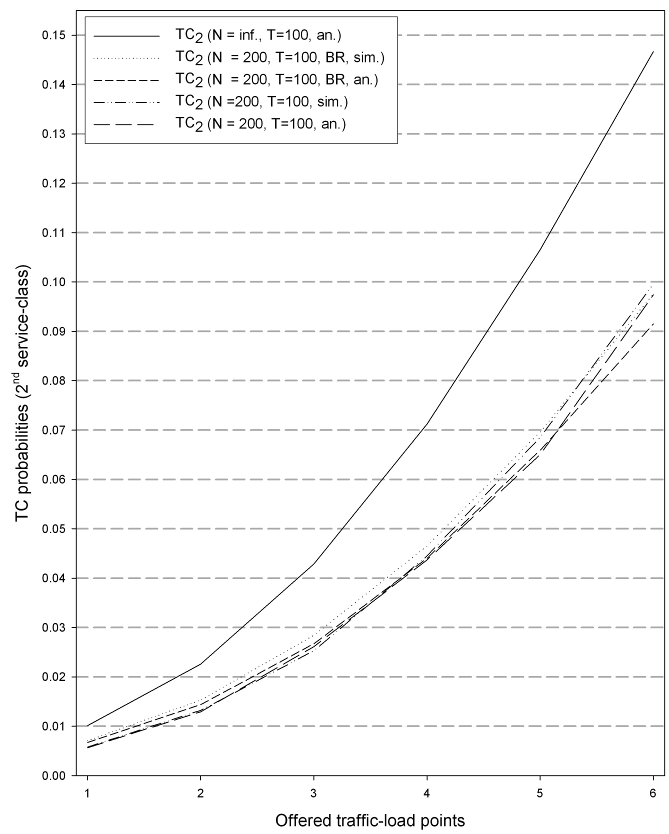

Figure 6. Time congestion probabilities of the 2^{nd} service-class (T=100 b.u.).

since both the E-EMLM and EF-EMLM give quite satisfactory results compared to simulation. The TC probabilities, CC probabilities and the link utilization in the generalized model can be determined by (18), (19) and (20), respectively.

When the BR policy is applied in the generalized model, the calculation of the link occupancy distribution is based on the following recursive formula:

$$G(j)=\begin{cases} 1 & \text{for} \quad j=0 \\ \frac{1}{\min(j,C)} \sum\limits_{k\in K_{e,\text{fin}}} (N_k-n_k+1)\alpha_k D_k(j-b_k)G(j-b_k) \\ +\frac{1}{j} \sum\limits_{k\in K_{a,\text{fin}}} (N_k-n_k+1)\alpha_k D_k(j-b_k)G(j-b_k) \\ +\frac{1}{\min(j,C)} \sum\limits_{k\in K_{e,\text{inf}}} \alpha_{k,\text{inf}} D_k(j-b_k)G(j-b_k) \\ +\frac{1}{j} \sum\limits_{k\in K_{a,\text{inf}}} \alpha_{k,\text{inf}} D_k(j-b_k)G(j-b_k) \\ & \text{for } j=1,\ldots,T \\ 0 & \text{for } \ j<0 \end{cases} \tag{31}$$

where the values of $D_k(j)$ are given by (22).

As far as the TC probabilities, CC probabilities and the link utilization are concerned, they can be determined by (28), (29) and (20), respectively.

VII. CONCLUSION

We propose an analytical model for the call-level performance assessment of telecom networks, when elastic and/or adaptive calls of different service-classes come from finite traffic-sources and compete for the available bandwidth of a single link with fixed capacity. Because of the existence of the bandwidth compression/expansion mechanism for handling the elastic/adaptive traffic, the proposed model does not have a product form solution. Therefore, we propose approximate but recursive formulas for the calculation of the most important performance measures, namely time congestion and call congestion probabilities, and link utilization. In addition, we incorporate in our model the bandwidth reservation policy (whereby a part of the link's available bandwidth is reserved to benefit calls of higher bandwidth requirements), and study its effect on the performance measures. Simulation results verify the analytical results and prove the accuracy and the consistency of the proposed model. Furthermore, we show the relation of the proposed model to other multirate loss models and generalize it to include a mixture of service-classes of finite and infinite number of traffic sources. Potential applications of the proposed model are in the environment of wireless networks that support elastic and adaptive traffic. As a future work, we would like to incorporate into the proposed model the peculiarities of wireless networks.

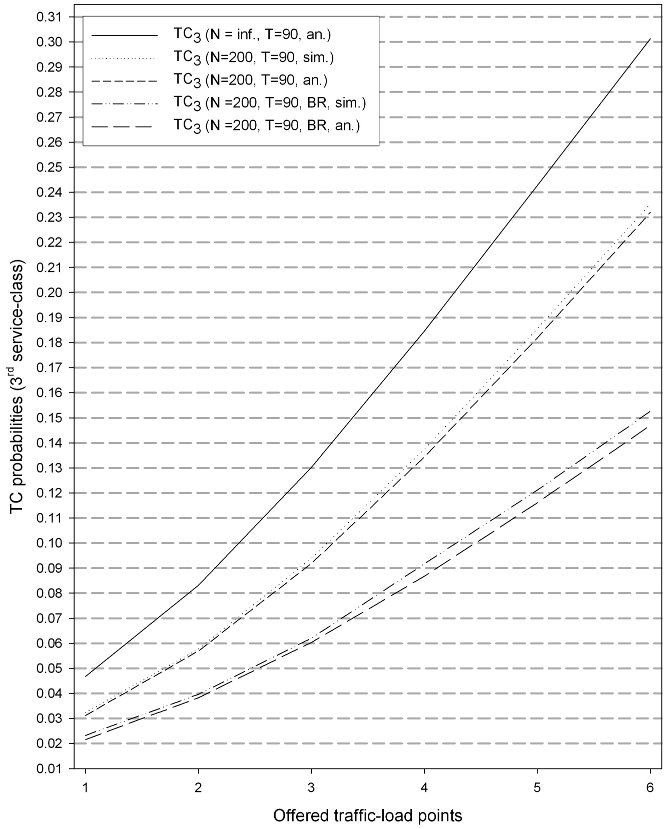

Figure 7. Time congestion probabilities of the 3^{rd} service-class (T=90 b.u.).

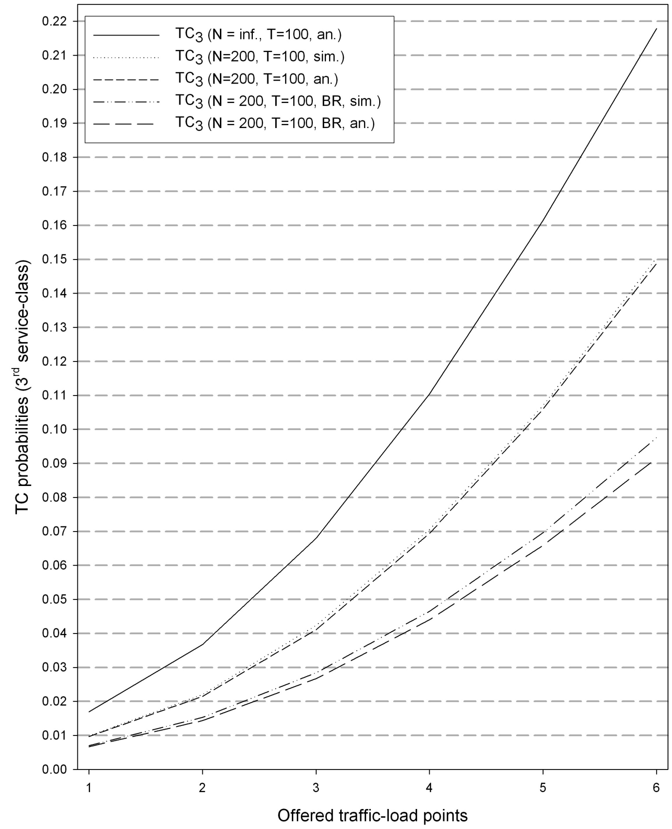

Figure 8. Time congestion probabilities of the 3^{rd} service-class (T=100 b.u.).

APPENDIX A
PROOF OF (12) FOR THE DETERMINATION OF $G(j)$' S WHEN $C < j \leq T$

When $C < j \leq T$, we multiply both sides of (4) by b_k^{comp} and sum over $k = 1, \ldots, K_e$ to have:

$$\sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k^{\mathrm{comp}} P(\mathbf{n}_k^{-1}) = P(\mathbf{n}) \sum_{k \in K_e} n_k b_k^{\mathrm{comp}} \phi_k(\mathbf{n}) \quad (32)$$

Based on (6) and (9), (32) is written as:

$$x(\mathbf{n}) \sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) = P(\mathbf{n}) \sum_{k \in K_e} x(\mathbf{n}_k^{-1}) n_k b_k \quad (33)$$

We continue by multiplying both sides of (5) by b_k^{comp} and sum over $k = 1, \ldots, K_a$ to obtain:

$$\sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k^{\mathrm{comp}} P(\mathbf{n}_k^{-1}) = P(\mathbf{n}) \sum_{k \in K_a} n_k b_k^{\mathrm{comp}} \phi_k(\mathbf{n}) \quad (34)$$

Based on (6) and (10) and since $r(\mathbf{n}) = \frac{C}{j}$, (34) is written as:

$$x(\mathbf{n}) \frac{C}{j} \sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) = P(\mathbf{n}) \frac{C}{j} \sum_{k \in K_a} x(\mathbf{n}_k^{-1}) n_k b_k \quad (35)$$

Adding (33) and (35), we have:

$$\begin{aligned} & x(\mathbf{n}) \left(\sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) \right) + \\ & + x(\mathbf{n}) \left(\frac{C}{j} \sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) \right) \\ & = P(\mathbf{n}) \left(\sum_{k \in K_e} x(\mathbf{n}_k^{-1}) n_k b_k + \frac{C}{j} \sum_{k \in K_a} x(\mathbf{n}_k^{-1}) n_k b_k \right) \end{aligned} \quad (36)$$

Due to (7), (36) can be written as:

$$\begin{aligned} & \sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) + \\ & + \frac{C}{j} \sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) = C P(\mathbf{n}) \end{aligned} \quad (37)$$

To introduce the link occupancy distribution $G(j)$ in (37), let $\mathbf{\Omega}_j = \{\mathbf{n} \in \mathbf{\Omega} : \mathbf{nb} = j\}$ be the state space where exactly j bandwidth units are occupied. Then, since $\sum_{\mathbf{n} \in \mathbf{\Omega}_j} P(\mathbf{n}) = G(j)$, summing both sides of (37) over $\mathbf{\Omega}_j$, we obtain:

$$\begin{aligned} & \sum_{\mathbf{n} \in \mathbf{\Omega}_j} \sum_{k \in K_e} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) + \\ & \frac{C}{j} \sum_{\mathbf{n} \in \mathbf{\Omega}_j} \sum_{k \in K_a} (N_k - n_k + 1)\alpha_k b_k P(\mathbf{n}_k^{-1}) = C G(j) \end{aligned} \quad (38)$$

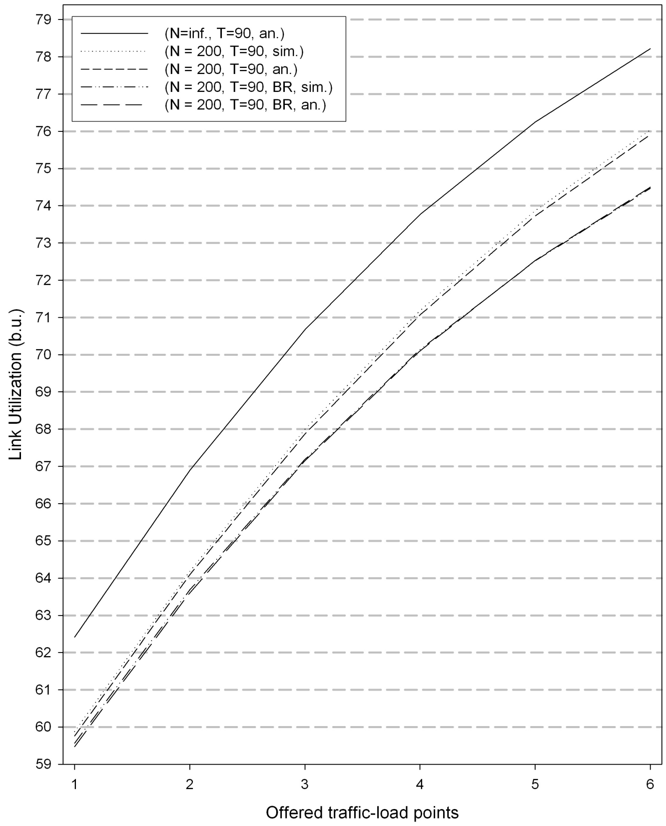

Figure 9. Link utilization (in b.u.), when T=90 b.u.

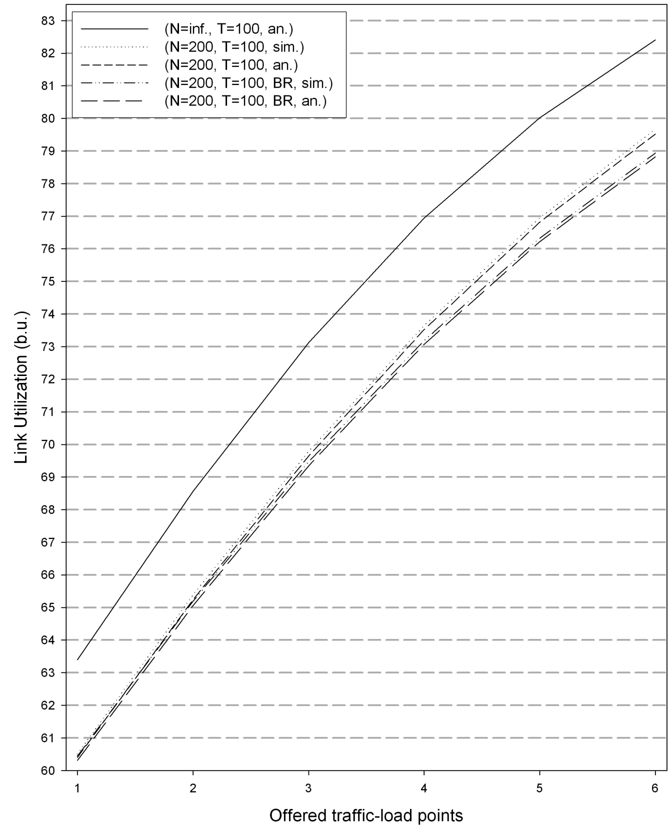

Figure 10. Link utilization (in b.u.), when T=100 b.u.

Interchanging the order of summations in (38) and assuming that each state has a unique occupancy j, we have:

$$\begin{aligned}&\sum_{k\in K_e}(N_k-n_k+1)a_kb_k\sum_{\mathrm{n}\in\Omega_j}P(\mathbf{n}_k^{-1})+\\&\frac{C}{j}\sum_{k\in K_a}(N_k-n_k+1)a_kb_k\sum_{n\in\Omega_j}P(\mathbf{n}_k^{-1})=CG(j)\end{aligned}\tag{39}$$

which can become exactly the same with (12):

$$\begin{aligned}&\sum_{k\in K_e}(N_k-n_k+1)\alpha_kb_kG(j-b_k)+\\&\frac{C}{j}\sum_{k\in K_a}(N_k-n_k+1)\alpha_kb_kG(j-b_k)=CG(j)\end{aligned}\tag{40}$$

ACKNOWLEDGMENT

The authors would like to thank Dr. V. Vassilakis (University of Surrey, U.K.) for his support. Also, the authors thank anonymous reviewers for their valuable remarks and suggestions.

REFERENCES

[1] I. Moscholios, J. Vardakas, M. Logothetis and M. Koukias, ‘A Quasi-random Multirate loss model supporting elastic and adaptive traffic’, Proc. of 4th Int. Conference on Emerging Network Intelligence, EMERG-ING 2012, Barcelona, Spain, pp. 56-61, 23-28 Sep. 2012.

[2] K. Ross, Multiservice loss models for broadband telecommunication networks, Springer, 1995, ISBN 3-540-19918-7.

[3] S. Rácz, B. Gerö, and G. Fodor, ‘Flow level performance analysis of a multi-service system supporting elastic and adaptive services’, Performance Evaluation, vol. 49, pp. 451-469, Sep. 2002.

[4] W. Bziuk, ‘Approximate state probabilities in large shared multirate loss systems with an application to trunk reservation’, European Transactions on Telecommunications, vol. 16, no. 3, pp. 205-216, May/June 2005.

[5] M. Glabowski, A. Kaliszan and M. Stasiak, ‘Asymmetric convolution algorithm for blocking probability calculation in full-availability group with bandwidth reservation’, IET Circuits, Devices & Systems, vol. 2, no. 1, pp. 87-94, Feb. 2008.

[6] I. Moscholios, V. Vassilakis, M. Logothetis and M. Koukias, ‘QoS Equalization in a Multirate Loss Model of Elastic and Adaptive Traffic with Retrials’, Proc. IARIA 5th Int. Conference on Emerging Network Intelligence, EMERGING 2013, Porto, Portugal, 29 Sep. 4 Oct. 2013.

[7] C. Oliveira, J. B. Kim, and T. Suda, ‘An adaptive bandwidth reservation scheme for high-speed multimedia wireless networks’, IEEE J. Select. Areas Commun., vol. 16, pp. 858-874, Aug. 1998.

[8] G. Raskutti, A. Zalesky, E.W.M Wong and M. Zukerman, ‘Enhanced Blocking Probability Evaluation Method for Circuit-Switched Trunk Reservation Networks’, IEEE Communications Letters, vol.11, no. 6, pp.543-545, June 2007.

[9] M. Stasiak, P. Zwierzykowski, D. Parniewicz, ‘Modelling of the WCDMA Interface in the UMTS Network with Soft Handoff Mechanism’, IEEE Globecom 2009, pp. 1-6, Nov. 2009.

[10] K. Kuppuswamy and D. C. Lee, ‘An Analytic Approach to Efficiently Computing Call Blocking Probabilities for Multiclass WDM Networks’, IEEE/ACM Trans. on Networking, vol. 17, no. 2, pp. 658-670, Apr. 2009.

[11] E. W. M. Wong, J. Baliga, M. Zukerman, A. Zalesky, and G. Raskutti, ‘A New Method for Blocking Probability Evaluation in OBS/OPS

Appendix B – List of Symbols

Symbol	Meaning
C	Capacity of the link (in bandwidth units)
T	Virtual capacity of the link (in bandwidth units)
j	Occupied link bandwidth (in bandwidth units), $j = 0, \dots, T$
$G(j)$	Link occupancy distribution
G	Normalization constant
K_e	Set of elastic service-classes
K_a	Set of adaptive service-classes
K	Set of service-classes, $K = K_e + K_a$
k	Service-class k $(k = 1, \dots, K)$
N_k	Finite number of sources of service-class k
b_k	Peak-bandwidth requirement of service-class k calls
$\mathbf{b}$	Vector of the required peak-bandwidth per call of all service-classes, $\mathbf{b} = (b_1, b_2, \dots, b_K)$
λ_k	Mean arrival rate of service-class k idle sources
$\lambda_{k,\inf}$	Mean arrival rate of Poisson service-class k calls
v_k	Arrival rate per idle source of service-class k
μ_k^{-1}	Mean of the exponentially distributed service time of service-class k calls
α_k	Offered traffic-load (in erl) per idle source of service-class k, $\alpha_k = v_k/\mu_k$
$\alpha_{k,\inf}$	Offered traffic-load (in erl) of Poisson service-class k calls, $\alpha_{k,\inf} = \lambda_{k,\inf}/\mu_k$
n_k	Number of in-service calls of service-class k
$\mathbf{n}$	Vector of all in service calls of all service-classes, $\mathbf{n} = (n_1, n_2, \dots, n_K)$
$P(\mathbf{n})$	Steady state distribution
b_k^{comp}	Compressed bandwidth of service-class k calls
r	Compression factor
b_k^{expan}	Expanded bandwidth of service-class k calls
$t(k)$	Bandwidth reservation parameter of service-class k
$\phi_k(\mathbf{n})$	State-dependent multiplier of service-class k
$x(\mathbf{n})$	State-dependent variable
$y_k(j)$	Mean number of Poisson service-class k calls in state j
P_{b_k}	Time Congestion probabilities of service-class k
C_{b_k}	Call Congestion probabilities of service-class k
U	Link utilization

Networks With Deflection Routing', J. Lightwave Technol., vol. 27, issue 23, pp. 5335-5347, Dec. 2009.

[12] F. Kelly, Reversibility and Stochastic Networks, Wiley Series in Probability and Mathematical Statistics. Wiley: New York, 1979.

[13] J. Kaufman, 'Blocking in a shared resource environment', IEEE Trans. Commun. vol. 29, Oct. 1981, pp. 1474-1481.

[14] J. Roberts, 'A service system with heterogeneous user requirements', in: G. Pujolle (Ed.), Performance of Data Communications systems and their applications, North Holland, Amsterdam, 1981, pp. 423-431.

[15] G. Kallos, V. Vassilakis, I. Moscholios, and M. Logothetis, 'Performance Modelling of W-CDMA Networks Supporting Elastic and Adaptive Traffic', Proc. 4th IFIP HET-NETs, Ilkley, U.K., pp. 09/1-09/10, 11-13 Sep. 2006.

[16] G. Fodor and M. Telek, 'Bounding the Blocking Probabilities in Multirate CDMA Networks Supporting Elastic Services', IEEE/ACM Trans. on Networking, vol. 15, pp. 944-956, Aug. 2007.

[17] M. Logothetis, V. Vassilakis and I. Moscholios, 'Call-level Performance Modeling and QoS Assessment of W-CDMA Networks', Wireless Networks: Research, Technology and Applications, Ed. Jia Feng, Nova Science Publishers, New York, USA, 2009, pp. 57-90.

[18] M. Glabowski, M. Stasiak, A. Wisniewski, and P. Zwierzykowski, 'Blocking Probability Calculation for Cellular Systems with WCDMA Radio Interface Servicing PCT1 and PCT2 Multirate Traffic', IEICE Trans. Commun., vol. E92-B, pp. 1156-1165, Apr. 2009.

[19] G. Kallos, V. Vassilakis and M. Logothetis, 'Call-level performance analysis of a W-CDMA cell with finite population and interference cancellation', European Trans. Telecom., vol. 22, no. 1, pp. 25-30, 2011.

[20] B. P. Gerö, P. L. Pályi and S. Rácz, 'Flow-level performance analysis of a multi-rate system supporting stream and elastic services', Int. J. Commun. Syst., Wiley, doi: 10.1002/dac.1383, 2012.

[21] I. Moscholios, V. Vassilakis, J. Vardakas and M. Logothetis, 'Call Blocking Probabilities of Elastic and Adaptive Traffic with Retrials', Proc. of 8th Advanced Int. Conf. on Telecommunications, AICT 2012, Stuttgart, Germany, 27 May-1 June 2012, pp. 92-97.

[22] V. Vassilakis, I. Moscholios and M. Logothetis, 'The extended connection-dependent threshold model for call-level performance analysis of multi-rate loss systems under the bandwidth reservation policy', Int. J. Commun. Syst., Wiley, vol. 25, no. 7, pp. 849-873, July 2012.

[23] I. Moscholios, J. Vardakas, M. Logothetis and A. Boucouvalas, 'QoS Guarantee in a Batched Poisson Multirate Loss Model Supporting Elastic and Adaptive Traffic', Proc. of IEEE ICC 2012, Ottawa, Canada, 10-15 June 2012, pp. 1296-1301.

[24] I. Moscholios, J. Vardakas, M. Logothetis and A. Boucouvalas, 'Congestion Probabilities in a Batched Poisson Multirate Loss Model Supporting Elastic and Adaptive Traffic', Annals of Telecommunications, vol. 68, issue 5, pp. 327-344, 2013.

[25] H. Akimaru and K. Kawashima, Teletraffic Theory and Applications, 2nd edition, Springer-Verlag, Berlin, 1999.

[26] I. Moscholios, M. Logothetis and M. Koukias, 'A State-Dependent Multi-Rate Loss Model of Finite Sources with QoS Guarantee for Wireless Networks', Mediterranean Journal of Computers and Networks, vol. 2, no. 1, pp. 10-20, Jan. 2006.

[27] M. Stasiak, M. Glabowski, A.Wisniewski, and P. Zwierzykowski, Modeling and Dimensioning of Mobile Networks, Wiley & Sons, 2011.

[28] S. Yashkov and A. Yashkova, 'Processor sharing: a survey of the mathematical theory', Automation and Remote Control, vol. 68, no. 9, pp. 1662-1731, Sep. 2007.

[29] G. Stamatelos and J. Hayes, 'Admission control techniques with application to broadband networks', Computer Communications, vol. 17, no. 9, pp. 663-673, 1994.

[30] I. Moscholios, M. Logothetis and P. Nikolaropoulos, 'Engset Multi-Rate State-Dependent Loss Models', Performance Evaluation, vol. 59, issues 2-3, pp. 247-277, Feb. 2005.

[31] I. Moscholios, M. Logothetis and M. Koukias, 'An ON-OFF Multirate Loss Model of Finite Sources', IEICE Trans. on Commun., vol. E90-B, no.7, pp. 1608-1619, July 2007.

[32] M. Glabowski and M. Stasiak, 'An approximate model of the full-availability group with multi-rate traffic and a finite source population', Proc. of 12th MMB&PGTS, Dresden, Germany, pp. 195-204, Sep. 2004.

[33] V. Vassilakis, G. Kallos, I. Moscholios and M. Logothetis, 'Call-Level Analysis of W-CDMA Networks Supporting Elastic Services of Finite Population', Proc. of IEEE ICC 2008, Beijing, China, 19-23 May 2008.

[34] I. Moscholios, M. Logothetis, V. Stylianakis and J. Vardakas, 'The Priority Wavelength Release Protocol for Dynamic Wavelength Allocation in WDM-TDMA PONs Supporting Random and Quasi-Random Bursty Traffic', Proc. of the 18th NOC 2013, Graz, Austria, 10-12 July 2013.

[35] M. Stasiak and M. Glabowski, 'A simple approximation of the link model with reservation by a one-dimensional Markov chain', Performance Evaluation, vol. 41, pp. 195-208, July 2000.

[36] I. Moscholios and M. Logothetis, 'Engset multi-rate state-dependent loss models with QoS guarantee', Int. J. Commun. Syst., vol. 19, pp. 67-93, Feb. 2006.

[37] J. Roberts, 'Teletraffic models for the Telecom 1 Integrated Services Network', Proc. of ITC-10, Mondreal, Canada, 1983.

[38] I. Moscholios, V. Vassilakis, M. Logothetis and J. Vardakas, 'Bandwidth Reservation in the Erlang Multirate Loss Model for Elastic and Adaptive Traffic', Proc. IARIA 9th Advanced Int. Conference on Telecommunications, AICT 2013, Rome, Italy, 23-28 June 2013.

[39] I. Moscholios, V. Vassilakis, M. Logothetis and A. Boucouvalas, 'Blocking Equalization in the Erlang Multirate Loss Model for Elastic Traffic', Proc. IARIA 2nd Int. Conference on Emerging Network Intelligence, EMERGING 2010, Florence, Italy, 25-30 Oct. 2010.

[40] Simscript III, http://www.simscript.com *(retrieved: December 2013).*