Erlang-Engset Multirate Retry Loss Models for Elastic and Adaptive Traffic under the Bandwidth Reservation Policy

Ioannis D. Moscholios*, Vassilios G. Vassilakis†, Michael D. Logothetis‡, and John S. Vardakas§

*Dept. of Informatics & Telecommunications, University of Peloponnese, 221 00 Tripolis, Greece. Email: idm@uop.gr

†Dept. of Electronic Engineering, University of Surrey, GU2 7XH Guildford, U.K. Email: v.vasilakis@surrey.ac.uk

‡Dept. of Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece. Email: mlogo@upatras.gr

§Iquadrat, Barcelona, Spain. Email: jvardakas@iquadrat.com

Abstract*—In this paper, we consider a single-link multirate loss system, which accommodates different service-classes with different traffic and peak-bandwidth requirements. Calls of each service-class arrive in the system according to a random (Poisson) or a quasi-random process, and have an exponentially distributed service time. Poisson or quasi-random arriving calls belong to service-classes of infinite or finite number of traffic sources, respectively. The service-classes are also distinguished, according to the behaviour of calls under service, in elastic and adaptive service-classes. Elastic calls can compress their bandwidth by simultaneously increasing their service time, while, adaptive calls do not affect their service time. A new call (either elastic or adaptive) is accepted in the system with its peak-bandwidth requirement, if there is available link bandwidth. If not, the call retries one or more times (single and multi-retry loss model, respectively) with a reduced bandwidth. If the available link bandwidth is lower than the call's last bandwidth requirement, the call can still compress its last bandwidth requirement (down to a certain bandwidth), together with the bandwidth of all in-service calls. Call blocking occurs, if, after compression, the call's bandwidth still exceeds the available link bandwidth. The system incorporates the Bandwidth Reservation (BR) policy, whereby we can achieve certain Quality of Service (QoS) for each service-class, through a proper bandwidth allocation defined by the BR parameters. To calculate in an approximate but efficient way, time and call congestion probabilities, as well as link utilization, we propose recurrent formulas for the determination of the link occupancy distribution. The accuracy of the proposed formulas is verified by simulation, and is found to be very satisfactory. We show the consistency and the necessity of the proposed models.

***Keywords*—Poisson process, quasi-random, time-call congestion probability, elastic/adaptive traffic, reservation, Markov chains, retrials, recurrent formula.**

I. INTRODUCTION

Elastic and adaptive traffic of multirate service-classes grows rapidly in modern networks, a fact that necessitates the development of efficient analytical tools for the call-level network performance analysis [1]. The term "elastic traffic" refers to in-service calls that have the ability to compress/expand their bandwidth and simultaneously increase/decrease their service time, during their lifetime in a system. On the other hand, the term "adaptive traffic" refers to in-service calls that tolerate bandwidth compression without altering their service time. Examples of elastic traffic are generally TCP-based applications (FTP, HTTP, STMP), while examples of adaptive traffic are mostly real-time applications, like audio and video streaming, which can be transmitted with an acceptable QoS after bandwidth compression.

Assuming that the call arrival process is Poisson, the calculation of various performance measures, such as call blocking probabilities and system's utilization, can be based on the classical Erlang Multirate Loss Model (EMLM) [2] - [3], which has been extensively used for the call-level performance evaluation of wired (e.g., [4] - [14]), wireless (e.g., [15] - [21]) and optical networks (e.g., [22] - [26]). If the call arrival process is quasi-random, i.e., calls come from a finite number of users, then the Engset Multirate Loss Model (EnMLM) arises [27].

In both the EMLM and the EnMLM, calls compete for the available link bandwidth according to the complete sharing policy (i.e., calls compete for all bandwidth resources) and have fixed bandwidth requirements. The latter means that in-service calls do not compress their bandwidth during their lifetime in the system. A new call is blocked and lost, if its required bandwidth is not available. In both models, the steady state probabilities have a Product Form Solution (PFS), which leads to an accurate calculation of call blocking probabilities (see e.g., [2], [3] and [27]).

In [28] and [29], the EMLM and the EnMLM, respectively, have been extended to include retrials. Blocked calls retry one or more times (Single-Retry Model (SRM) or Multi-Retry Model (MRM), respectively) to be accepted in the link by requiring less bandwidth. A retry call is blocked and lost, if the available link bandwidth is lower than the call's last bandwidth requirement. In [30], an approximate method has been proposed for both single and multi retries in the EnMLM that simplifies the calculation of call blocking probabilities.

In [31], the authors have extended [28] by incorporating the notion of elastic traffic. Instead of rejecting immediately a retry call, the link may accept this call by compressing its bandwidth, together with the bandwidth of all in-service calls of all service-classes. Elastic calls increase their service time so that the product *bandwidth* by *service time* remains constant. After compression, the retry call is accepted in the system, if the resultant bandwidth is not higher than the available link bandwidth; otherwise the retry call is blocked and lost. When a call with compressed bandwidth leaves the system, then the remaining in-service calls expand their bandwidth.

In [32], [33], the authors have extended [31] to include adaptive traffic, as well as the Bandwidth Reservation (BR) policy. Adaptive calls compress or expand their bandwidth without altering their service time. On the other hand, the BR policy can achieve equalization of blocking probabilities among service-classes (either elastic or adaptive), or guarantee a certain QoS for each service-class, by a proper selection of the BR parameters so that each service-class meets a certain link bandwidth capacity. Note that the aforementioned models are consistent; that is, if blocked calls of all service-classes are not allowed to retry, then the model of [33] results in the model proposed in [1].

The consideration of the BR policy is of paramount importance in multirate communication networks, given that the absence of the BR policy leads to an unfair service (the less required bandwidth, the better call blocking probability). The system under the BR policy becomes non-PFS, because the Markov chain that describes the system, loses its reversibility.

In this paper, we extend [29], [30] to include elastic and adaptive traffic with retrials under the BR policy. Due to the existence of retrials, the BR policy and bandwidth compression, the proposed elastic/adaptive single-retry and multi-retry loss models for quasi-random input do not have a PFS. However, we propose approximate but recursive formulas for the calculation of the link occupancy distribution and, consequently, time and call congestion probabilities, as well as link utilization. Note that the proposed models are also consistent: if calls are generated by an infinite number of users and blocked calls are not allowed to retry, then the proposed models result in the model of [1].

The remainder of this paper is as follows: In Section II, we present application areas for teletraffic multirate loss models that support elastic traffic. In Section III, for the integrity of the paper, we review the model of [1], named herein Extended EMLM/BR (E-EMLM/BR). In Section IV, we review the models of [33], named herein Extended SRM/BR and Extended MRM/BR (E-SRM/BR and E-MRM/BR, respectively). In Section IV.A, we review the E-SRM/BR, while in Section IV.B, we consider the E-MRM/BR. In Section V, we assume that calls arrive according to a quasi-random process and propose the Extended Finite SRM/BR and the Extended Finite MRM/BR (EF-SRM/BR and EF-MRM/BR, respectively). We prove the recursive formulas for the link occupancy distribution and provide formulas for the calculation of time and call congestion probabilities, as well as link utilization. Section VI is the evaluation section. We present analytical and simulation results of the various performance measures for the proposed models. We also provide analytical results of existing models for comparison. We conclude in Section VII. Finally, we tabulate as Appendix A and B, all the symbols and acronyms, respectively, used in this paper.

II. Applications of Multirate Elastic Loss Models

Application areas for multirate loss models that include the case of elastic traffic and the notion of bandwidth compression are numerous (see e.g., [34] - [41] and the references therein). These areas can also be considered relevant to our proposed models, which include the notion of retrials, the BR policy and the case of adaptive traffic. However, the proposed models are mostly applicable in wireless networks, where calls may come from finite sources (it is justified by the limited coverage of a cell) and their bandwidth can be compressed, while the BR policy can protect handover calls.

In [34] and [35], an EMLM based model for the recursive calculation of flow throughput and packet loss rate in IP networks is proposed. A link of certain capacity accommodates elastic calls of different service-classes. Arriving calls follow a Poisson process. The link capacity is shared among calls according to a balanced fairness criterion: if the occupied link bandwidth does not exceed the capacity of the link, then all calls use their peak-bandwidth requirement; otherwise, all calls share the capacity in proportion to their peak-bandwidth requirement and the link operates at its full capacity. The main difference of the compression mechanism between [34], [35] and the E-EMLM/BR lies on the fact that in [34] and [35], there is no parameter for admission control. The application of balanced fairness in multirate tree networks and its comparison with other classical bandwidth allocation policies (e.g., max-min fairness) are examined in [36], [37].

In [38], a Code Division Multiple Access cell is considered, which accommodates multirate elastic service-classes. Elastic calls arrive in the link according to a Poisson process and have an exponentially distributed service time. The main target of this paper is the calculation of upper and lower bounds for call blocking probabilities based on an extension of [9] (named herein E-EMLM). In [39], the co-existence of stream traffic (calls cannot compress their assigned bandwidth) and elastic traffic in IEEE802.16e mobile WiMAX subject to adaptive modulation and coding is considered. Calls of both stream and elastic service-classes arrive in the system according to a Poisson process and have an exponentially distributed service time. Stream calls have priority over elastic calls and, in that sense, elastic calls share the left-over capacity of the system. This means that the blocking probability of stream service-classes does not depend on the amount of traffic generated by elastic service-classes. The co-existence of stream and elastic traffic results in an analytical model which can not be described by recursive formulas (see also [40] for a more general multirate loss model) and, therefore, the calculation of blocking probabilities is based on the solution of the steady-state probabilities equations. Such a solution is inefficient for systems with large capacity and many service-classes, due to the extremely large number of equations that arise. Another extension of the EMLM that studies the co-existence of stream and elastic traffic (with similar problems with those described for [39]) in the downlink of Orthogonal Frequency-Division Multiple Access wireless cellular networks is proposed in [41].

III. Review of The E-EMLM/BR

Consider a single link of capacity C bandwidth units (b.u.) that accommodates calls of K service-classes. Let K_e and K_a be the set of elastic and adaptive service-classes ($K_e + K_a = K$), respectively. A call of service-class k ($k = 1, \ldots, K$) follows a Poisson process with arrival rate $\lambda_{k,inf}$ and has a peak-bandwidth requirement of b_k b.u. (integer value), as well as a BR parameter of $t(k)$ b.u. The latter refers to the number of b.u. reserved so that service-class k meets a link bandwidth capacity of $C - t(k)$ b.u. By assigning a bigger BR parameter $t(k)$ to a service-class k requiring less bandwidth per call than another service-class, this, benefits all service-class calls of a higher bandwidth per call. Let j be the occupied link bandwidth when a new service-class k call arrives in the link. Bandwidth compression is introduced in the model by assuming that j may exceed C up to a value of T b.u.; T is called virtual capacity of the link. If $j + b_k \leq C$, the call is accepted in the system with its b_k b.u. and remains in the system for an exponentially distributed service time with mean μ_k^{-1}. The new service-class k call is blocked and lost if $j + b_k > T - t(k)$. If $T - t(k) \geq j + b_k > C$, the new call is accepted in the system. However, the assigned bandwidth of all in-service calls, together with the peak-bandwidth requirement of the new call is compressed. After the bandwidth compression of all calls (new and in-service) the system state becomes $j = C$. The compressed bandwidth of the new service-class k call is calculated by:

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

where $r \equiv r(\mathbf{n}) = C/j$ is the compression factor (common to all service-classes), $j^{'} = j + b_k = \mathbf{nb} + b_k$, $\mathbf{n} = (n_1, \ldots, n_k, \ldots, n_K)$, n_k is the number of in-service calls of service-class k, $\mathbf{b} = (b_1, \ldots, b_K)$ and $j = \sum_{k=1}^{K} n_k b_k = \mathbf{nb}$.

Similarly, the compressed bandwidth of all in-service calls is equal to $b_i^{'} = \frac{C}{j^{'}} b_i$ for $i = 1, \ldots, K$. The minimum bandwidth of a service-class k call is given by:

$$b_{k,\min}^{'} = r_{\min} b_k = \frac{C}{T} b_k \tag{2}$$

After the bandwidth compression, all elastic calls increase their service time so that the product (*service time*) by (*bandwidth*) remains constant. A simple tutorial example that describes in detail the bandwidth compression mechanism can be found in [13]. The mechanism of bandwidth compression/expansion and the existence of the BR policy destroy reversibility in the E-EMLM/BR and therefore no PFS exists. However, in [1] an approximate recursive formula is proposed, which determines the link occupancy distribution, $G(j)$, (unnormalized values):

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{\min(j,C)} \sum\limits_{k \in K_e} \alpha_{k,\inf} D_k(j - b_k) G(j - b_k) + \\ \frac{1}{j} \sum\limits_{k \in K_a} \alpha_{k,\inf} D_k(j - b_k) G(j - b_k) & \\ & \text{for } j = 1, \ldots, T \\ 0 & \text{for } j < 0 \end{cases} \tag{3}$$

$$D_k(j - b_k) = \begin{cases} b_k & for \;\; j \leq T - t(k) \\ 0 & for \;\; j > T - t(k) \end{cases} \tag{4}$$

where $\alpha_{k,\inf} = \lambda_{k,\inf}/\mu_k$ is the offered traffic-load (in erl) of service-class k.

As far as the computational complexity of (3) is concerned is in the order of O(KT).

The BR policy ensures equalization of blocking probabilities among different service-classes by a proper selection of the BR parameters. If, for example, blocking equalization is required between calls of three service-classes with b_1=1, b_2=7 and b_3=10 b.u., respectively, then $t(1) = 9$ b.u, $t(2) = 3$ and $t(3) = 0$ b.u., so that $b_1 + t(1) = b_2 + t(2) = b_3 + t(3)$.

The application of the BR policy in the E-EMLM/BR is based on the assumption that the number of service-class k calls is negligible in states $j > T - t(k)$ and is incorporated in (3) by the variable $D_k(j - b_k)$ given in (4). The states $j > T - t(k)$ belong to the so-called reservation space. Note that the population of calls of service-class k in the reservation space may not be negligible. In [8] and [42], a complex procedure is implemented in order to take into account this population and increase the accuracy of the resultant blocking probability in the EMLM and Engset multirate state-dependent loss models, respectively. However, according to [42], this procedure may not always increase the accuracy of blocking probability results compared to simulation.

Based on (3), (4), we can calculate time and call congestion probabilities, and the link utilization, as follows:

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

$$P_{b_k} = \sum_{j=T-b_k-t(k)+1}^{T} G^{-1} G(j) \tag{5}$$

where: $G = \sum_{j=0}^{T} G(j)$ 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}, is the probability that a new service-class k call is blocked and lost:

$$C_{b_k} = \sum_{j=T-b_k-t(k)+1}^{T} G^{-1} G(j) \tag{6}$$

Call congestion probabilities are determined by the proportion of arriving calls that find the system congested. Time and call congestion probabilities coincide in the case of Poisson arrivals (due to the Poisson Arrivals See Time Averages (PASTA) property [43]), but not in the case of quasi-random arrivals.

3) The link utilization, denoted as U:

$$U = \sum_{j=1}^{C} j G^{-1} G(j) + \sum_{j=C+1}^{T} C G^{-1} G(j) \tag{7}$$

Note that if the BR policy is not applied in the system, i.e., $t(k)$ = 0 for all k $(k = 1, \ldots, K)$, then the link occupancy distribution is given by the E-EMLM [9]:

$$G(j)=\begin{cases}1 & \text{for } j=0\\ \frac{1}{\min(j,C)}\sum\limits_{k\in K_e}\alpha_{k,\inf}b_kG(j-b_k)+ & \\ \frac{1}{j}\sum\limits_{k\in K_a}\alpha_{k,\inf}b_kG(j-b_k) & \text{for } j=1,\ldots,T\\ 0 & \text{for } j<0\end{cases} \quad (8)$$

In that case, the calculation of time and call congestion probabilities is given by (5), (6), respectively, where $t(k) = 0$ for all k $(k = 1, \ldots, K)$. Furthermore, if $T = C$, then the link accommodates only stream traffic (i.e., calls of all service-classes cannot compress their bandwidth) and the EMLM results. In the EMLM, the link occupancy distribution is given by the classical Kaufman-Roberts recursion [2], [3]:

$$G(j)=\begin{cases}1 & \text{for } j=0\\ \frac{1}{j}\sum\limits_{k\in K}\alpha_{k,\inf}b_kG(j-b_k) & \text{for } j=1,\ldots,C\\ 0 & \text{for } j<0\end{cases} \quad (9)$$

IV. Review of The E-SRM/BR And The E-MRM/BR

A. The E-SRM/BR

Consider again the link of capacity C b.u. that accommodates K_e and K_a elastic and adaptive service-classes, respectively. Service-class k calls $(k = 1, \ldots, K)$ follow a Poisson process with rate $\lambda_{k,\inf}$, request b_k b.u. (peak-bandwidth requirement), have a BR parameter of $t(k)$ b.u. and an exponentially distributed service time with mean μ_k^{-1}.

Let j be the occupied link bandwidth, $j = 0, 1, \ldots, T$, when a service-class k call arrives in the link. Now, we consider the following cases:

a) If $j + b_k \leq C$, the call is accepted in the link with b_k b.u.

b) If $j + b_k > C$, then the call is blocked with b_k and retries immediately to be connected in the link with $b_{kr} < b_k$. Now if:

  b1) $j + b_{kr} \leq C$ the retry call is accepted in the system with b_{kr} and $\mu_{kr}^{-1} > \mu_k^{-1}$, so that $b_{kr}\mu_{kr}^{-1} = b_k\mu_k^{-1}$,

  b2) $j + b_{kr} > T - t(k)$ the retry call is blocked and lost, and

c) $C < j + b_{kr} \leq T - t(k)$ the retry call is accepted in the system by compressing its bandwidth requirement b_{kr} together with the bandwidth of all in-service calls of all service-classes. In that case, the compressed bandwidth of the retry call becomes $b'_{kr} = rb_{kr} = \frac{C}{j+b_{kr}}b_{kr}$ where r is the compression factor, common to all service-classes. Similarly, all in-service calls, which have been accepted in the link with b_k (or b_{kr}), compress their bandwidth to $b'_k = rb_k$ (or $b'_{kr} = rb_{kr}$) for $k = 1, \ldots, K$. After the compression of all calls the link state is $j = C$. The minimum value of the compression factor is $r_{\min} = C/T$.

Similar to the E-EMLM/BR, when a service-class k call, with bandwidth b'_k (or b'_{kr}), departs from the system, the remaining in-service calls of each service-class i $(i = 1, \ldots, K)$, expand their bandwidth in proportion to their initially assigned bandwidth b_i (or b_{ir}). After bandwidth compression/expansion, only elastic service-class calls increase/decrease their service time so that the product *service time* by *bandwidth* remains constant.

The existence of retrials, the BR policy and the bandwidth compression mechanism destroy reversibility in the model and therefore no PFS exists. However, in [33] an approximate recursive formula is proposed for the calculation of the unnormalized values of the link occupancy distribution, $G(j)$:

$$G(j)=\begin{cases}1 & \text{for } j=0\\ \frac{1}{j}\sum\limits_{k\in K_a}\alpha_{k,\inf}D_k(j-b_k)\gamma_k(j)G(j-b_k)+ & \\ \frac{1}{j}\sum\limits_{k\in K_a}\alpha_{kr,\inf}D_{kr}(j-b_{kr})\gamma_{kr}(j)G(j-b_{kr})+ & \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e}\alpha_{k,\inf}D_k(j-b_k)\gamma_k(j)G(j-b_k)+ & \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e}\alpha_{kr,\inf}D_{kr}(j-b_{kr})\gamma_{kr}(j)G(j-b_{kr}) & \\ & \text{for } j=1,\ldots,T\\ 0 & \text{otherwise}\end{cases} \quad (10)$$

where: $\alpha_{k,\inf} = \lambda_{k,\inf}\,\mu_k^{-1}$ is the offered traffic-load (in erl) of service-class k calls,
$\alpha_{kr,\inf} = \lambda_{k,\inf}\,\mu_{kr}^{-1}$,

$$\gamma_k(j) = \begin{cases}1 & \text{for } 1 \leq j \leq C \text{ and } b_{kr} > 0\\ 1 & \text{for } 1 \leq j \leq T \text{ and } b_{kr} = 0\\ 0 & \text{otherwise}\end{cases},$$

$$\gamma_{kr}(j) = \begin{cases}1 & \text{for } C - b_k + b_{kr} < j \leq T\\ 0 & \text{otherwise}\end{cases},$$

$$D_k(j - b_k) = \begin{cases}b_k & \text{for } j \leq T - t(k)\\ 0 & \text{for } j > T - t(k)\end{cases},$$

$$D_{kr}(j - b_{kr}) = \begin{cases}b_{kr} & \text{for } j \leq T - t(k)\\ 0 & \text{for } j > T - t(k)\end{cases},$$

and $t(k)$ is the reserved bandwidth in favor of calls other than service-class k calls.

The proof of (10) is based on:

i) the application of local balance between adjacent states, which exists only in PFS models,

ii) an approximation in (10), expressed by $\gamma_{kr}(j)$, which assumes that the occupied link bandwidth from retry calls of service-class k is negligible, when $j \leq C - (b_k - b_{kr})$,

iii) an approximation in (10), expressed by $\gamma_k(j)$ that refers only to those service-class k calls whose $b_{kr} > 0$; this approximation assumes that the occupied link bandwidth from service-class k calls accepted in the system with b_k b.u. is negligible when $j > C$.

Having determined $G(j)$'s, we can calculate time and call congestion probabilities, as well as link utilization. The final time congestion probability of a retry service-class k call, P_{bk_r}, is given by [33]:

$$P_{bk_r} = \sum_{j=T-b_{kr}-t(k)+1}^{T} G^{-1}G(j) \quad (11)$$

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

Note that time and call congestion probabilities coincide in the case of Poisson arrivals. As far as the link utilization is concerned, it is calculated according to (7), where the values of $G(j)$'s are given by (10).

B. The E-MRM/BR

In the E-MRM/BR, a service-class k call that is not accepted in the system with its peak-bandwidth requirement, b_k, may have many retry parameters $(b_{kr_l}, \mu_{kr_l}^{-1})$ for $l = 1, \ldots, s(k)$, with $b_{kr_{s(k)}} < \ldots < b_k$ and $\mu_{kr_{s(k)}}^{-1} > \ldots > \mu_k^{-1}$. Similar to the E-SRM/BR, the E-MRM/BR does not have a PFS and therefore the calculation of $G(j)$'s is based on an approximate but recursive formula [33]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k\in K_a} a_{k,\inf} D_k(j-b_k)\gamma_k(j)G(j-b_k) & \\ +\frac{1}{j}\sum\limits_{k\in K_a}\sum\limits_{s=1}^{s(k)} a_{kr_s,\inf} D_{kr_s}(j-b_{kr_s})\cdot & \\ \gamma_{kr_s}(j)G(j-b_{kr_s})+ & \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e} a_{k,\inf} D_k(j-b_k)\gamma_k(j)G(j-b_k) & \\ +\frac{1}{\min(C,j)}\sum\limits_{k\in K_e}\sum\limits_{s=1}^{s(k)} a_{kr_s,\inf} D_{kr_s}(j-b_{kr_s})\cdot & \\ \gamma_{kr_s}(j)G(j-b_{kr_s}) & \text{for } j=1,\ldots,T \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where: $\alpha_{kr_s,\inf} = \lambda_{k,\inf}\mu_{kr_s}^{-1}$

$$\gamma_k(j) = \begin{cases} 1 \text{ for } 1 \le j \le C \text{ and } b_{kr_s} > 0 \\ 1 \text{ for } 1 \le j \le T \text{ and } b_{kr_s} = 0 \\ 0 \text{ otherwise} \end{cases}$$

$$\gamma_{kr_s}(j) = \begin{cases} 1 \text{ for } C - b_{kr_{s-1}} + b_{kr_s} < j \le C \text{ and } s \ne s(k) \\ 1 \text{ for } C - b_{kr_{s-1}} + b_{kr_s} < j \le T \text{ and } s = s(k) \\ 0 \text{ otherwise} \end{cases}$$

$$D_k(j-b_k) = \begin{cases} b_k & \text{for } j \le T - t(k) \\ 0 & \text{for } j > T - t(k) \end{cases},$$

$$D_{kr_s}(j-b_{kr_s}) = \begin{cases} b_{kr_s} & \text{for } j \le T - t(k) \\ 0 & \text{for } j > T - t(k) \end{cases}$$

The computational complexity of (12) is $O(KT + \sum_{k=1}^{K}(s(k)(b_k - b_{kr_1}))$, assuming that the difference $b_{kr_{s-1}} - b_{kr_s}$ is constant.

If the BR policy is not applied, then we have the E-MRM and (12) takes the form [32]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k\in K_a} \alpha_{k,\inf} b_k \gamma_k(j)G(j-b_k)+ & \\ \frac{1}{j}\sum\limits_{k\in K_a}\sum_{s=1}^{s(k)} \alpha_{kr_s,\inf} b_{kr_s}\gamma_{kr_s}(j)G(j-b_{kr_s})+ & \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e} \alpha_{k,\inf} b_k\gamma_k(j)G(j-b_k)+ & \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e}\sum\limits_{s=1}^{s(k)} \alpha_{kr_s,\inf} b_{kr_s}\gamma_{kr_s}(j)G(j-b_{kr_s}) & \\ \quad \text{for } j=1,\ldots,T & \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

If only elastic service-classes are accommodated by the link, then (12) becomes [33]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(C,j)}\sum\limits_{k\in K_e} \alpha_{k,\inf} D_k(j-b_k)\gamma_k(j)G(j-b_k) & \\ +\frac{1}{\min(C,j)}\sum\limits_{k\in K_e}\sum\limits_{s=1}^{s(k)} \alpha_{kr_s,\inf} D_{kr_s}(j-b_{kr_s})\cdot & \\ \gamma_{kr_s}(j)G(j-b_{kr_s}) & \text{for } j=1,\ldots,T \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

If the link accommodates elastic and adaptive service-classes whose blocked calls are not allowed to retry, then (12) takes the form of (3) and the E-EMLM/BR results [1].

If calls of all service-classes may retry but are not allowed to compress their bandwidth during their service time, then the MRM under the BR policy results and (12) takes the form [44]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k\in K} \alpha_{k,\inf} D_k(j-b_k)G(j-b_k)+ & \\ \frac{1}{j}\sum\limits_{k\in K}\sum\limits_{s=1}^{s(k)} \alpha_{kr_s,\inf} D_{kr_s}(j-b_{kr_s})\gamma_{kr_s}(j)G(j-b_{kr_s}) & \\ \quad \text{for } j=1,\ldots,C & \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

Furthermore, if blocked calls of all service-classes are not allowed to retry, then the EMLM under the BR policy results and (15) takes the form [45]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k\in K} \alpha_{k,\inf} D_k(j-b_k)G(j-b_k) & \\ \quad \text{for } j=1,\ldots,C & \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Having determined $G(j)$'s in the E-MRM/BR according to (12), we can calculate the time (and call) congestion probabilities of a retry service-class k call with its last bandwidth requirement, $P_{bkr_{s(k)}}$, according to the formula [33]:

$$P_{bkr_{s(k)}} = \sum_{j=T-b_{kr_{s(k)}}-t(k)+1}^{T} G^{-1}G(j) \quad (17)$$

The calculation of the link utilization in the E-MRM/BR is based on (7) where the values of $G(j)$'s are given by (12).

V. The Proposed EF-SRM/BR And EF-MRM/BR

In this section, we extend the retry multirate loss models of [29], [30] (which do not examine elastic and adaptive traffic) to include elastic and adaptive traffic under the BR policy. Blocked calls of quasi-random arrivals have the ability to retry one or more times (EF-SRM/BR or EF-MRM/BR, respectively) to be connected in the system with reduced bandwidth. If the available link bandwidth is still higher than the last bandwidth requirement of a retry call, then the call can still try to be connected in the system by compressing its requirement together with the bandwidth of all in-service calls.

A. The EF-SRM/BR

The proposed EF-SRM/BR is a non-PFS model. In order to prove an approximate but recursive formula for the determination of $G(j)$'s we present the following example.

Consider a link of capacity C b.u. that accommodates calls of two service-classes. The 1^{st} service-class is adaptive and the 2^{nd} is elastic. Calls of both service-classes are generated by a finite source population N_k $(k = 1, 2)$. The mean call arrival rate of service-class k idle sources is $\lambda_k = (N_k - n_k)v_k$, where v_k is the arrival rate per idle source and n_k is the number of in-service calls. This call arrival process is a quasi-random process [43]. A Poisson process arises from a quasi-random process if $N_k \to \infty$ for $k = 1, \ldots, K$, while the total offered traffic-load remains constant.

Assuming that only calls of the 2^{nd} service-class can retry, the traffic parameters of both service-classes are: $(N_1, v_1, \mu_1^{-1}, b_1)$ for the 1^{st} service-class and $(N_2, v_2, \mu_2^{-1}, \mu_{2r}^{-1}, b_2, b_{2r})$ for the 2^{nd} service-class, with $b_{2r} < b_2$ and $\mu_{2r}^{-1} > \mu_2^{-1}$. Initially, let assume that the BR parameters: $t(1) = t(2) = 0$. Bandwidth compression is permitted for calls of both service-classes up to a limit T.

The description of call admission is based on a new service-class k call $(k = 1, 2)$ that arrives in the system, when the occupied link bandwidth is j b.u. Then:

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

ii) If $j + b_k > C$ we consider the following sub-cases:

a) If $T \ge j + b_1 > C$, a 1^{st} service-class call is accepted in the system by compressing b_1, as well as the assigned bandwidth of all in-service calls. The compressed bandwidth of the 1^{st} service-class call is given by $b_1^{'} = rb_1 = (C/j^{'})b_1$, where $r = C/j^{'}$, $j^{'} = j + b_1 = \mathbf{nb} + b_1$. Similarly, the bandwidth of all in-service calls will be compressed (by the same factor r) and become $b_k^{'} = (C/j^{'})b_k$ for $k = 1, 2$. After compression has taken place, all calls share the C b.u. in proportion to their bandwidth requirement, while the link operates at its full capacity C. The minimum bandwidth that a 1^{st} service-class call can tolerate is $b_{1,\min}^{'} = r_{\min} b_1 = (C/T)b_1$.

b) If $j + b_1 > T$, the 1^{st} service-class call is blocked and lost.

c) If $j + b_2 > C$, a 2^{nd} service-class call is blocked and retries with $b_{2r} < b_2$. Now, we consider three cases:

c1) If $j + b_{2r} \le C$, the retry call is accepted in the system with b_{2r}.

c2) If $j + b_{2r} > T$, the call is blocked and lost.

c3) If $C < j + b_{2r} \le T$ the call is accepted in the system by compressing b_{2r} together with the bandwidth of all in-service calls. The compressed bandwidth of the call is $b_{2r}^{'} = rb_{2r} = (C/j^{'})b_{2r}$ where $j^{'} = j + b_{2r}$. Similarly, the bandwidth of all in-service calls are compressed (by the same factor r) and become $b_k^{'} = (C/j^{'})b_k$ for $k = 1, 2$. The minimum bandwidth that a 2^{nd} service-class call tolerates is $b_{2r,\min}^{'} = (C/T)b_{2r}$.

Although the steady state probabilities in the proposed model do not have a PFS, we assume that local balance exists between the adjacent states of the 1^{st} service-class:

$$(N_1 - n_1 + 1)v_1 P(\mathbf{n}_1^-) = n_1 \mu_1 \phi_1(\mathbf{n}) P(\mathbf{n}), \quad 1 \le \mathbf{nb} \le T \quad (18)$$

where: $\mathbf{n}_1^- = (n_1 - 1, n_2, n_{2r})$, $\mathbf{n} = (n_1, n_2, n_{2r})$, $\mathbf{b} = (b_1, b_2, b_{2r})$, $n_1 \ge 1$, $P(\mathbf{n})$ is the probability distribution of state $\mathbf{n}$, and

$$\phi_1(\mathbf{n}) = \begin{cases} 1 & , \text{ when } \mathbf{nb} \le C \\ x(\mathbf{n}_1^-)/x(\mathbf{n}), & \text{when } C < \mathbf{nb} \le T \\ 0 & , \text{ otherwise} \end{cases} \quad (19)$$

where: $\mathbf{nb} = j = n_1 b_1 + n_2 b_2 + n_{2r} b_{2r}$ and n_{2r} is the number of in-service retry calls of the 2^{nd} service-class.

Note that $\phi_k(\mathbf{n})$ is a state dependent factor which describes: i) bandwidth compression and ii) the increase factor of service time of service-class k calls in state $\mathbf{n}$. In other words, $\phi_k(\mathbf{n})$ has the same role with r, but it may be different for each service-class.

By multiplying both sides of (18) with b_1 and $r(\mathbf{n})$, and based on (19), we have:

$$(N_1 - n_1 + 1)\alpha_1 b_1 x(\mathbf{n}) r(\mathbf{n}) P(\mathbf{n}_1^-) = n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n}) P(\mathbf{n}) \quad (20)$$

where $\alpha_1 = v_1 \mu_1^{-1}$is the offered traffic-load per idle source of 1^{st} service-class, $r(n) = \min(1, C/j)$ and $1 \le \mathbf{nb} \le T$.

Based on the call admission control mechanism described for 2^{nd} service-class calls, the following local balance equations can be derived:

a) For $1 \le \mathbf{nb} \le C$, $\mathbf{n}_2^- = (n_1, n_2 - 1, n_{2r})$, and $n_2 \ge 1$:

$$(N_2 - n_2 + 1)v_2 P(\mathbf{n}_2^-) = n_2 \mu_2 \phi_2(\mathbf{n}) P(\mathbf{n}) \quad (21)$$

where:

$$\phi_2(\mathbf{n}) = \begin{cases} 1 & , \text{ when } \mathbf{nb} \le C \\ x(\mathbf{n}_2^-)/x(\mathbf{n}), & \text{when } C < \mathbf{nb} \le T \\ 0 & , \text{ otherwise} \end{cases} \quad (22)$$

By multiplying both sides of (21) with b_2, and based on (22), we obtain:

$$(N_2 - n_2 + 1)\alpha_2 b_2 x(\mathbf{n}) P(\mathbf{n}_2^-) = n_2 b_2 x(\mathbf{n}_2^-) P(\mathbf{n}) \quad (23)$$

where: $1 \le \mathbf{nb} \le C$, and $\alpha_2 = v_2 \mu_2^{-1}$.

b) If $P(\mathbf{n}_{2r}^-)$ is the probability distribution of state $\mathbf{n}_{2r}^- = (n_1, n_2, n_{2r} - 1)$,

$$(N_2 - n_2 - n_{2r} + 1)v_2 P(\mathbf{n}_{2r}^-) = n_{2r} \mu_{2r} \phi_{2r}(\mathbf{n}) P(\mathbf{n}) \quad (24)$$

where: $C - b_2 + b_{2r} < \mathbf{nb} \le T$, and

$$\phi_{2r}(\mathbf{n}) = \begin{cases} 1 & , \text{ when } \mathbf{nb} \le C \\ x(\mathbf{n}_{2r}^-)/x(\mathbf{n}), & \text{when } C < \mathbf{nb} \le T \\ 0 & , \text{ otherwise} \end{cases} \quad (25)$$

By multiplying both sides of (24) with b_{2r}, and based on (25), we obtain for $C - b_2 + b_{2r} < \mathbf{nb} \le T$:

$$(N_2 - n_2 - n_{2r} + 1)\alpha_{2r} b_{2r} x(\mathbf{n}) P(\mathbf{n}_{2r}^-) = n_{2r} b_{2r} x(\mathbf{n}_{2r}^-) P(\mathbf{n}) \quad (26)$$

where: $\alpha_{2r} = v_2\mu_{2r}^{-1}$.

Equations (20), (23) and (26) lead to a system of equations:

$$\begin{aligned}&(N_1-n_1+1)\alpha_1 b_1 x(\mathbf{n}) r(\mathbf{n}) P(\mathbf{n}_1^-)+\\&(N_2-n_2+1)\alpha_2 b_2 x(\mathbf{n}) P(\mathbf{n}_2^-)\\&=(n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n})+n_2 b_2 x(\mathbf{n}_2^-))P(\mathbf{n})\\&\text{for } 1 \le \mathbf{nb} \le C-b_2+b_{2r}\end{aligned} \tag{27}$$

$$\begin{aligned}&(N_1-n_1+1)\alpha_1 b_1 x(\mathbf{n}) r(\mathbf{n}) P(\mathbf{n}_1^-)+\\&(N_2-n_2+1)\alpha_2 b_2 x(\mathbf{n}) P(\mathbf{n}_2^-)+\\&(N_2-n_2-n_{2r}+1)\alpha_{2r} b_{2r} x(\mathbf{n}) P(\mathbf{n}_{2r}^-)=\\&(n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n})+n_2 b_2 x(\mathbf{n}_2^-)+n_{2r} b_{2r} x(\mathbf{n}_{2r}^-))P(\mathbf{n})\\&\text{for } C-b_2+b_{2r}< \mathbf{nb} \le C\end{aligned} \tag{28}$$

$$\begin{aligned}&(N_1-n_1+1)\alpha_1 b_1 x(\mathbf{n}) r(\mathbf{n}) P(\mathbf{n}_1^-)+\\&(N_2-n_2-n_{2r}+1)\alpha_{2r} b_{2r} x(\mathbf{n}) P(\mathbf{n}_{2r}^-)=\\&(n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n})+n_{2r} b_{2r} x(\mathbf{n}_{2r}^-))P(\mathbf{n})\\&\text{for } C< \mathbf{nb} \le T\end{aligned} \tag{29}$$

By assuming that retry calls with b_{2r} are negligible when $1 \le nb \le C - b_2 + b_{2r}$ and that the population of calls with b_2 is negligible when $C < \mathbf{nb} \le T$, we can combine (27), (28) and (29) into the following equation:

$$\begin{aligned}&(N_1-n_1+1)\alpha_1 b_1 x(\mathbf{n}) r(\mathbf{n}) P(\mathbf{n}_1^-)+\\&(N_2-n_2+1)\gamma_2(\mathbf{nb})\,\alpha_2 b_2 x(\mathbf{n}) P(\mathbf{n}_2^-)+\\&(N_2-n_2-n_{2r}+1)\gamma_{2r}(\mathbf{nb})\alpha_{2r} b_{2r} x(\mathbf{n}) P(\mathbf{n}_{2r}^-)=\\&(n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n})+n_2 b_2 x(\mathbf{n}_2^-)+n_{2r} b_{2r} x(\mathbf{n}_{2r}^-))P(\mathbf{n})\\&\text{for } 1 \le nb \le T\end{aligned} \tag{30}$$

where: $\gamma_2(\mathbf{nb}) = 1$ for $1 \le \mathbf{nb} \le C$, otherwise $\gamma_2(\mathbf{nb}) = 0$, and: $\gamma_{2r}(\mathbf{nb}) = 1$ for $C-b_2+b_{2r}< \mathbf{nb} \le T$, otherwise $\gamma_{2r}(\mathbf{nb}) = 0$.

In order to derive a formula for $x(\mathbf{n})$, we make the following assumptions:

1) When $C < \mathbf{nb} \le T$, the bandwidth of all in-service calls are compressed by $\phi_k(\mathbf{n})$, $k = 1, 2$, so that:

$$n_1 b_1^{'} + n_2 b_2^{'} + n_{2r} b_{2r}^{'} = C \tag{31}$$

2) 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:

$$\begin{aligned}&\tfrac{b_1 r(\mathbf{n})}{\mu_1} = \tfrac{b_1^{'}}{\mu_1\phi_1(\mathbf{n})} \ \text{ or } b_1^{'} = b_1\phi_1(\mathbf{n})r(\mathbf{n})\\&\tfrac{b_2 r(\mathbf{n})}{\mu_2 r(n)} = \tfrac{b_2^{'}}{\mu_2\phi_2(\mathbf{n})} \ \textit{ or } b_2^{'} = b_2\phi_2(\mathbf{n})\\&\tfrac{b_{2r} r(\mathbf{n})}{\mu_{2r} r(\mathbf{n})} = \tfrac{b_{2r}^{'}}{\mu_{2r}\phi_{2r}(\mathbf{n})} \ \textit{ or } b_{2r}^{'} = b_{2r}\phi_{2r}(\mathbf{n})\end{aligned} \tag{32}$$

By substituting (32) in (31), we have:

$$n_1 b_1 \phi_1(\mathbf{n}) r(\mathbf{n}) + n_2 b_2 \phi_2(\mathbf{n}) + n_{2r} b_{2r} \phi_{2r}(\mathbf{n}) = C \tag{33}$$

where $\phi_1(\mathbf{n})$, $\phi_2(\mathbf{n})$ and $\phi_{2r}(\mathbf{n})$ are given by (19), (22) and (25), respectively.

Equation (33), due to (19), (22) and (25) is written as:

$$x(\mathbf{n})=\begin{cases}1 & \text{for } \ \mathbf{nb} \le C, \ \ \mathbf{n}\in\mathbf{\Omega}\\ \frac{1}{C}n_1 b_1 x(\mathbf{n}_1^-) r(\mathbf{n})+\frac{1}{C}n_2 b_2 x(\mathbf{n}_2^-)+ & \\ \frac{1}{C}n_{2r} b_{2r} x(\mathbf{n}_{2r}^-) & \text{for } \ C < \mathbf{nb} \le T\\ 0 & \text{otherwise}\end{cases} \tag{34}$$

Based on (34), we consider again (30). Since $x(\boldsymbol{n}) = 1$, when $0 \le j \le C$, it is proved in [7] that:

$$\begin{aligned}&(N_1-n_1+1)\alpha_1 b_1 G(j-b_1)+\\&(N_2-n_2+1)\alpha_2 b_2 G(j-b_2)+\\&(N_2-n_2-n_{2r}+1)\alpha_{2r} b_{2r}\gamma_{2r}(j) G(j-b_{2r}) = jG(j)\\&\text{for } \ 1 \le j \le C\end{aligned} \tag{35}$$

where: $G(j)$ is the link occupancy distribution, $\gamma_{2r}(j) = 1$ for $C-b_2+b_{2r} < j$, otherwise $\gamma_{2r}(j) = 0$.

When $C < j \le T$, we have $\gamma_2(j)$=0, and due to (34), we may write (30) as follows:

$$\begin{aligned}&\tfrac{C}{j}(N_1-n_1+1)\alpha_1 b_1 P(\mathbf{n}_1^-)+\\&(N_2-n_2-n_{2r}+1)\gamma_{2r}(\mathbf{nb})\alpha_{2r} b_{2r} P(\mathbf{n}_{2r}^-) = CP(\mathbf{n})\end{aligned} \tag{36}$$

since $r(\mathbf{n}) = C/j$, when $C < j \le T$.

In order to introduce the link occupancy distribution ($G(j)$) in (36), we sum both sides of (36) over the set of states $\{\mathbf{n}\in\mathbf{\Omega}\,|\mathbf{nb} = j\}$, where $\mathbf{\Omega} = \{\mathbf{n} : 0 \le \mathbf{nb} \le T\}$:

$$\begin{aligned}&\tfrac{C}{j}(N_1-n_1+1)\alpha_1 b_1 \sum_{\{\mathbf{n}|\mathbf{nb}=j\}} P(\mathbf{n}_1^-)+\\&(N_2-n_2-n_{2r}+1)\gamma_{2r}(\mathbf{nb})\alpha_{2r} b_{2r} \sum_{\{\mathbf{n}|\mathbf{nb}=j\}} P(\mathbf{n}_{2r}^-)\\&= C\textstyle\sum_{\{\mathbf{n}|\mathbf{nb}=j\}} P(\mathbf{n})\end{aligned} \tag{37}$$

Since by definition $G(j) = \sum_{n\in\Omega_j} P(n)$, we may write (37) as follows:

$$\begin{aligned}&\tfrac{C}{j}(N_1-n_1+1)\alpha_1 b_1 G(j-b_1)+\\&(N_2-n_2-n_{2r}+1)\gamma_{2r}(j)\alpha_{2r} b_{2r} G(j-b_{2r}) = CG(j)\end{aligned} \tag{38}$$

where $\gamma_{2r}(j) = 1$ for $C-b_2+b_{2r} < j \le T$.

The combination of (35) and (38) gives an approximate recursive formula for the determination of $G(j)$'s, when calls of the 1^{st} service-class are adaptive while calls of the 2^{nd} service-class are elastic and have retry parameters:

$$\begin{aligned}&G(j) = \tfrac{1}{j}(N_1-n_1+1)\alpha_1 b_1 G(j-b_1)+\\&\tfrac{1}{\min(j,C)}\left[(N_2-n_2+1)\alpha_2 b_2\gamma_2(j)G(j-b_2)\right]+\\&\tfrac{1}{\min(j,C)}\left[(N_2-n_2-n_{2r}+1)\alpha_{2r} b_{2r}\gamma_{2r}(j)G(j-b_{2r})\right]\\&\text{for } \ 1 \le j \le T\end{aligned} \tag{39}$$

where: $\gamma_2(j)$=1 for $1 \le j \le C$, otherwise $\gamma_2(j)$=0 and $\gamma_{2r}(j)$=1 for $C - b_2 + b_{2r} < j \le T$, otherwise $\gamma_{2r}(j)$=0.

In the general case of K different service-classes, where all calls may retry, (39) takes the form:

$$G(j)=\begin{cases}1 \quad \text{for } j = 0\\[4pt] \frac{1}{j}\sum\limits_{k\in K_a}(N_k-n_k+1)\alpha_k b_k\gamma_k(j)G(j-b_k)+\\ \frac{1}{j}\sum\limits_{k\in K_a}(N_k-n_k-n_{kr}+1)\alpha_{kr} b_{kr}\gamma_{kr}(j)G(j-b_{kr})\\ +\frac{1}{\min(C,j)}\sum\limits_{k\in K_e}(N_k-n_k+1)\alpha_k b_k\gamma_k(j)G(j-b_k)\\ +\frac{1}{\min(C,j)}\sum\limits_{k\in K_e}(N_k-n_k-n_{kr}+1)\alpha_{kr} b_{kr}\\ \gamma_{kr}(j)G(j-b_{kr}) \qquad \text{for } j = 1, ..., T\\[4pt] 0 \quad \text{otherwise}\end{cases} \tag{40}$$

where: $\alpha_{kr} = v_k \mu_{kr}^{-1}$

$$\gamma_k(j) = \begin{cases} 1 & \text{for } 1 \leq j \leq C \text{ and } b_{kr} > 0 \\ 1 & \text{for } 1 \leq j \leq T \text{ and } b_{kr} = 0 \\ 0 & \text{otherwise} \end{cases},$$

$$\gamma_{kr}(j) = \begin{cases} 1 & \text{for } C - b_k + b_{kr} < j \leq T \text{ and } b_{kr} > 0 \\ 0 & \text{otherwise} \end{cases}$$

If the BR parameters are positive, then (40) takes the form:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{j} \sum_{k \in K_a} (N_k - n_k + 1) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) \\ + \frac{1}{j} \sum_{k \in K_a} (N_k - n_k - n_{kr} + 1) \alpha_{kr} D_{kr}(j - b_{kr}) \gamma_{kr}(j) \\ G(j - b_{kr}) + \frac{1}{\min(C,j)} \sum_{k \in K_e} (N_k - n_k + 1) \alpha_k D_k(j - b_k) \\ \gamma_k(j) G(j - b_k) + \frac{1}{\min(C,j)} \sum_{k \in K_e} (N_k - n_k - n_{kr} + 1) \\ \alpha_{kr} D_{kr}(j - b_{kr}) \gamma_{kr}(j) G(j - b_{kr}) \\ \quad \text{for } j = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases} \quad (41)$$

where: $D_k(j - b_k) = \begin{cases} b_k & \text{for } j \leq T - t(k) \\ 0 & \text{for } j > T - t(k) \end{cases}$

and $D_{kr}(j - b_{kr}) = \begin{cases} b_{kr} & \text{for } j \leq T - t(k) \\ 0 & \text{for } j > T - t(k) \end{cases}$

Furthermore, if calls of all service-classes are elastic, then (41) takes the form:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{\min(C,j)} \sum_{k \in K} (N_k - n_k + 1) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) \\ + \frac{1}{\min(C,j)} \sum_{k \in K} (N_k - n_k - n_{kr} + 1) \alpha_{kr} D_{kr}(j - b_{kr}) \cdot \\ \gamma_{kr}(j) G(j - b_{kr}) \quad \text{for } j = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases} \quad (42)$$

The calculation of $G(j)$'s in (40), (41) or (42) requires the values of n_k and n_{kr}, which are unknown. In other finite multirate loss models (e.g., [11], [27], [29]) there exist methods for the determination of these values through an equivalent stochastic system, with the same traffic description parameters and set of states. However, the state space determination of the equivalent system is complex, especially for large systems that serve many service-classes. Thus, we avoid such methods and approximate n_k and n_{kr} in state j, i.e., $n_k(j)$ and $n_{kr}(j)$, as the mean number of service-class k calls in state j, $y_k(j)$ and $y_{kr}(j)$, respectively, when Poisson arrivals are considered. Such approximations are common in the literature and induce little error (e.g., [30], [46] - [47]). In that case, we may rewrite (41) as follows:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{j} \sum_{k \in K_a} (N_k - y_k(j - b_k)) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) \\ + \frac{1}{j} \sum_{k \in K_a} (N_k - y_k(j - b_{kr}) - y_{kr}(j - b_{kr})) \cdot \\ \alpha_{kr} D_{kr}(j - b_{kr}) \gamma_{kr}(j) G(j - b_{kr}) + \\ \frac{1}{\min(C,j)} \sum_{k \in K_e} (N_k - y_k(j - b_k)) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) \\ + \frac{1}{\min(C,j)} \sum_{k \in K_e} (N_k - y_k(j - b_{kr}) - y_{kr}(j - b_{kr})) \alpha_{kr} \cdot \\ D_{kr}(j - b_{kr}) \gamma_{kr}(j) G(j - b_{kr}) \quad \text{for } j = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases} \quad (43)$$

where the values of $y_k(j)$ and $y_{kr}(j)$ are given by:

$$y_k(j) = \alpha_{k,\inf} \gamma_k(j) G_{\inf}(j - b_k) / G_{\inf}(j) \quad (44)$$

$$y_{kr}(j) = \alpha_{kr,\inf} \gamma_{kr}(j) G_{\inf}(j - b_{kr}) / G_{\inf}(j) \quad (45)$$

where: $\alpha_{k,\inf}$, $\alpha_{kr,\inf}$ and $G_{\inf}(j)$ are the offered traffic-load (in erl) of service-class *k* and the link occupancy distribution, respectively, of the corresponding infinite model (E-SRM/BR) [33], i.e., the values of $G_{\inf}(j)$ will be determined by (10).

Having determined $G(j)$'s in the EF-SRM/BR according to (43), we calculate the time congestion probabilities according to (11), and the link utilization according to (7). As far as the call congestion probabilities are concerned, we may again use (11), but the values of $G(j)$'s in (43) should be determined for a system with $N_k - 1$ traffic sources.

B. *The EF-MRM/BR*

Similar to the EF-SRM/BR, the corresponding multi-retry model does not have a PFS and therefore the $G(j)$'s calculation is based on an approximate but recursive formula. In the EF-MRM/BR, a blocked service-class *k* call retries *s*(*k*) times with parameters: $(b_{kr_s}, \mu_{kr_s}^{-1})$ for $s = 1, \ldots, s(k)$, where $b_{kr_{s(k)}} < \ldots < b_{kr_1} < b_k$ and $\mu_{kr_{s(k)}}^{-1} > \ldots > \mu_{kr_1}^{-1} > \mu_k^{-1}$. The determination of $G(j)$'s is based on (46) whose proof is similar to that of (41) and therefore is not presented:

$$G(j) = \begin{cases} 1 & \text{for } j = 0 \\ \frac{1}{j} \sum_{k \in K_a} (N_k - n_k + 1) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) + \\ \frac{1}{j} \sum_{k \in K_a} \sum_{s=1}^{s(k)} (N_k - (n_k + n_{kr_1} + \ldots + n_{kr_{s(k)}}) + 1) \cdot \\ \alpha_{kr_s} D_{kr_s}(j - b_{kr_s}) \gamma_{kr_s}(j) G(j - b_{kr_s}) + \\ \frac{1}{\min(C,j)} \sum_{k \in K_e} (N_k - n_k + 1) \alpha_k D_k(j - b_k) \gamma_k(j) G(j - b_k) + \\ \frac{1}{\min(C,j)} \sum_{k \in K_e} \sum_{s=1}^{s(k)} (N_k - (n_k + n_{kr_1} + \ldots + n_{kr_{s(k)}}) + 1) \cdot \\ \alpha_{kr_s} D_{kr_s}(j - b_{kr_s}) \gamma_{kr_s}(j) G(j - b_{kr_s}) \\ \quad \text{for } j = 1, \ldots, T \\ 0 & \text{otherwise} \end{cases} \quad (46)$$

where: $\alpha_{kr_s} = v_k \mu_{kr_s}^{-1}$,

$D_k(j - b_k) = \begin{cases} b_k & \text{for } j \leq T - t(k) \\ 0 & \text{for } j > T - t(k) \end{cases}$

$$D_{kr_s}(j-b_{kr_s}) = \begin{cases} b_{kr_s} & \text{for } j \leq T-t(k) \\ 0 & \text{for } j > T-t(k) \end{cases}$$

$$\gamma_k(j) = \begin{cases} 1 & \text{for } 1 \leq j \leq C \text{ and } b_{kr_s} > 0 \\ 1 & \text{for } 1 \leq j \leq T \text{ and } b_{kr_s} = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\gamma_{kr_s}(j) = \begin{cases} 1 & \text{for } C-b_{kr_{s-1}}+b_{kr_s} < j \leq C \text{ if } s \neq s(k) \\ 1 & \text{for } C-b_{kr_{s-1}}+b_{kr_s} < j \leq T \text{ if } s = s(k) \\ 0 & \text{otherwise} \end{cases}$$

As in the EF-SRM/BR, we approximate $n_k(j)$ and $n_{kr_s}(j)$ for $s = 1, \ldots, s(k)$ with the corresponding values of the infinite model [33]. In that case, (46) takes the form:

$$G(j) = \begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum_{k\in K_a}(N_k - y_k(j-b_k))\alpha_k D_k(j-b_k)\gamma_k(j)G(j-b_k) \\ +\frac{1}{j}\sum_{k\in K_a}\sum_{s=1}^{s(k)}(N_k - Y_k(j-b_{kr_s}))\alpha_{kr_s}D_{kr_s}(j-b_{kr_s}) \\ \cdot\gamma_{kr_s}(j)G(j-b_{kr_s})+ \\ \frac{1}{\min(C,j)}\sum_{k\in K_e}(N_k - y_k(j-b_k))\alpha_k D_k(j-b_k)\cdot \\ \gamma_k(j)G(j-b_k)+\frac{1}{\min(C,j)}\sum_{k\in K_e}\sum_{s=1}^{s(k)}(N_k - Y_k(j-b_{kr_s})) \\ \cdot\alpha_{kr_s}D_{kr_s}(j-b_{kr_s})\gamma_{kr_s}(j)G(j-b_{kr_s}) \\ & \text{for } j=1,\ldots,T \\ 0 & \text{otherwise} \end{cases} \quad (47)$$

where: $Y_k(j-b_{kr_s}) = y_k(j-b_{kr_s}) + y_{kr_1}(j-b_{kr_s}) + \ldots + y_{kr_{s(k)}}(j-b_{kr_s})$ and the values of $y_k(j)$ and $y_{kr_s}(j)$ are given by:

$$y_k(j) = a_{k,\text{inf}}\gamma_k(j)G_{\text{inf}}(j-b_k)/G_{\text{inf}}(j) \quad (48)$$

$$y_{kr_s}(j) = \alpha_{kr,\text{inf}}\gamma_{kr_s}(j)G_{\text{inf}}(j-b_{kr_s})/G_{\text{inf}}(j) \quad (49)$$

where $G_{\text{inf}}(j)$ refers to the link occupancy distribution of the corresponding infinite model (E-MRM/BR) [33], i.e., the values of $G_{\text{inf}}(j)$ will be given by (12).

If the BR policy is not applied, then we have the EF-MRM and (47) takes the form:

$$G(j) = \begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum_{k\in K_a}(N_k - y_k(j-b_k))\alpha_k b_k\gamma_k(j)G(j-b_k) \\ +\frac{1}{j}\sum_{k\in K_a}\sum_{s=1}^{s(k)}(N_k - Y_k(j-b_{kr_s}))\alpha_{kr_s}b_{kr_s}\gamma_{kr_s}(j) \\ G(j-b_{kr_s})+\frac{1}{\min(C,j)}\sum_{k\in K_e}(N_k - y_k(j-b_k))\alpha_k b_k\cdot \\ \gamma_k(j)G(j-b_k)+\frac{1}{\min(C,j)}\sum_{k\in K_e}\sum_{s=1}^{s(k)}(N_k - Y_k(j-b_{kr_s})) \\ \alpha_{kr_s}b_{kr_s}\gamma_{kr_s}(j)G(j-b_{kr_s}) & \text{for } j=1,\ldots,T \\ 0 & \text{otherwise} \end{cases} \quad (50)$$

where: $Y_k(j-b_{kr_s})$ is determined again through (48) and (49), but the values of $G_{\text{inf}}(j)$ will be given by (13), because of the absence of the BR policy.

Having determined $G(j)$'s in the EF-MRM/BR according to (47), we calculate the time congestion probabilities according to (17) and the link utilization according to (7). Call congestion probabilities are determined again by (17) but the values of $G(j)$'s in (47) should be calculated for a system with $N_k - 1$ traffic sources.

VI. Application Example – Evaluation

We consider an application example in order to compare the analytical Time Congestion (TC) probabilities with those obtained by simulation, in the case of the proposed EF-MRM and EF-MRM/BR of quasi-random input. Simulation is based on SIMSCRIPT III [48]. For a better evaluation, we comparatively present the corresponding analytical results of the E-MRM and the E-MRM/BR [33], i.e., assuming Poisson arrivals. We also show the analytical and simulation results of the proposed EF-MRM for the link utilization; they are compared with the corresponding analytical results obtained by the E-MRM. The simulation results of this section are mean values of 7 runs with 95% confidence interval. The resultant reliability ranges of the simulation measurements are very small and, therefore, we present only mean values.

To facilitate the reader, in what follows, we summarize the order of calculations per model, regarding the analytical TC probabilities and link utilization:

(A) EF-MRM/BR: Determine $G(j)$'s according to (47) with the aid of (48), (49) and (12). Then, determine link utilization according to (7) and TC probabilities according to (17).

(B) EF-MRM: Determine $G(j)$'s according to (50) with the aid of (48), (49) and (13). Then, determine link utilization according to (7) and TC probabilities according to (17) by assuming that $t(k) = 0$ for all $k = 1, \ldots, K$.

(C) E-MRM/BR: Determine $G(j)$'s according to (12), link utilization according to (7) and TC probabilities according to (17).

(D) E-MRM: Determine $G(j)$'s according to (13). Then, determine link utilization according to (7) and TC probabilities according to (17) by assuming that $t(k) = 0$ for all $k = 1, \ldots, K$.

Let us consider a link of capacity $C = 80$ b.u. that accommodates three service-classes of elastic calls. All calls arrive in the system according to a quasi-random process. The traffic characteristics of each service-class are the following:

1^{st} service-class: $N_1 = 100$, $v_1 = 0.20$, $b_1 = 1$ b.u.
2^{nd} service-class: $N_2 = 100$, $v_2 = 0.06$, $b_2 = 2$ b.u.
3^{rd} service-class: $N_3 = 100$, $v_3 = 0.02$, $b_3 = 6$ b.u.

The call holding time is exponentially distributed with mean value $\mu_1^{-1} = \mu_2^{-1} = \mu_3^{-1} = 1$. Calls of the 3^{rd} service-class may retry two times with reduced bandwidth requirement: $b_{3r_1} = 5$ b.u. and $b_{3r_2} = 4$ b.u. and increased service time so that $\alpha_3 b_3 = \alpha_{3r_1}b_{3r_1} = \alpha_{3r_2}b_{3r_2}$, where $\alpha_k = v_k\mu_k^{-1}$, $k = 1, 2, 3$. The corresponding Poisson traffic-loads are: $\alpha_{1,\text{inf}} = 20$, $\alpha_{2,\text{inf}} = 6$, $\alpha_{3,\text{inf}} = 2$ erl. In the x-axis of all figures (below), we assume that v_3 remains constant while v_1, v_2 increase in steps of 0.01 and 0.005, respectively. The last value of $v_1 = 0.28$, while that of $v_2 = 0.10$. The corresponding last values of the Poisson traffic-loads are: $\alpha_{1,\text{inf}} = 28$, $\alpha_{2,\text{inf}} = 10$, $\alpha_{3,\text{inf}} = 2$ erl.

Two different values of T are considered: a) $T = C = 80$ b.u., where no bandwidth compression takes place. In that case, the proposed model gives exactly the same results with the model of [30], b) $T = 82$ b.u., where bandwidth compression takes place and $r_{\min} = C/T = 80/82$. As far as the BR parameters are concerned, we choose $t(1) = 3$ b.u., $t(2) = 2$ b.u. and $t(3) = 0$ b.u., so that $b_1 + t(1) = b_2 + t(2) = b_{3r_2} + t(3)$. The selection of these BR parameters achieves equalization of TC probabilities for calls of all service-classes.

In Figs. 1-3 we present the TC probabilities of the EF-MRM and E-MRM, i.e., we consider the case whereby the BR policy is not applied. In Fig. 1, we show the analytical and simulation TC probabilities results of the 1^{st} service-class for both values of T. Similar results are presented in Figs. 2 and 3, for the 2^{nd} and 3^{rd} service-class, respectively (TC probabilities of calls with b_{3r_2}). The results of these three figures show that:

i) The model's accuracy is absolutely satisfactory compared to simulation (because of the very small differences of the results).
ii) The TC probabilities are lower, when the compression/expansion mechanism is introduced. A small increase of T resulted in a great effect on TC probabilities.
iii) The proposed model is important (necessary), since the results obtained by the existing infinite model (E-MRM) fail to approximate the results of the proposed finite model (EF-MRM).

Successive increases of T will result in even lower TC probabilities but at the cost of increasing the service time of all calls. Such behaviour has been observed in various papers that propose multirate loss models of elastic and/or adaptive traffic (e.g., [9], [32], [33]). Similarly, the increase of retrials for a particular service-class will cause the decrease of TC probabilities for that service-class and a possible increase of TC probabilities for the rest service-classes. This behaviour has been observed in various papers that study multirate loss models with retrials (e.g., [28]-[30]).

In Fig. 4 we present the equalized TC probabilities, when the BR policy is applied. In addition to the aforementioned comments for Figs. 1-3, the results of Fig. 4 show an increase of the TC probabilities of the 1^{st} and 2^{nd} service-classes in comparison to the corresponding curves of Fig. 1 and Fig. 2, respectively, and a decrease of the TC probabilities of the 3^{rd} service-class in comparison to the corresponding curves of Fig. 3.

Finally, in Fig. 5 we show the analytical and simulation results of the link utilization in the case of the E-MRM and the EF-MRM. Similar results are obtained in the case of the BR policy, and therefore they are not presented. These results also show that: i) The model's accuracy is absolutely satisfactory compared to simulation. ii) The introduction of the compression/expansion mechanism slightly increases the link utilization; the increase of T above C results in a slight increase of the link utilization, which is anticipated due to the decrease of TC probabilities. iii) The results obtained by the E-MRM cannot approximate the results of the proposed EF-MRM.

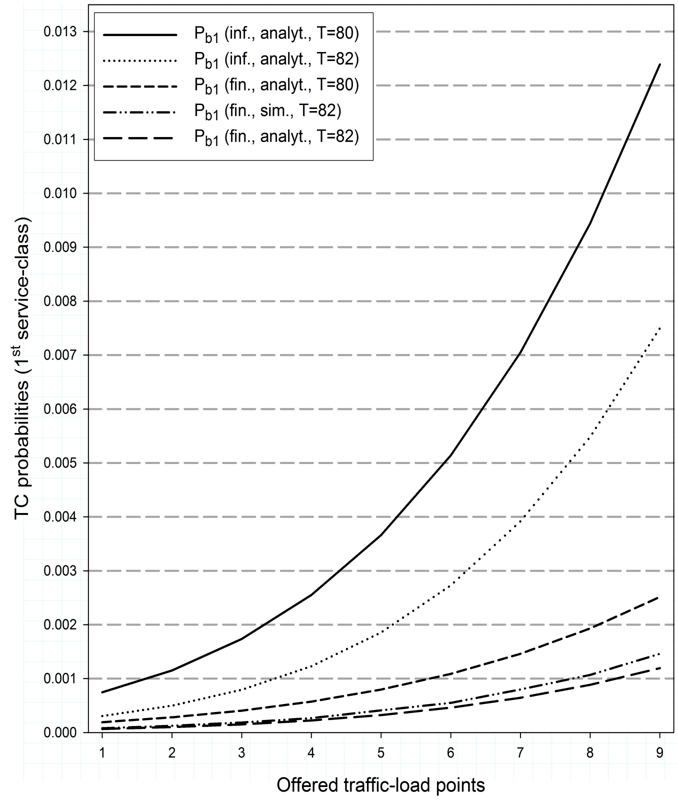

Fig. 1. TC probabilities – 1^{st} service-class (without BR).

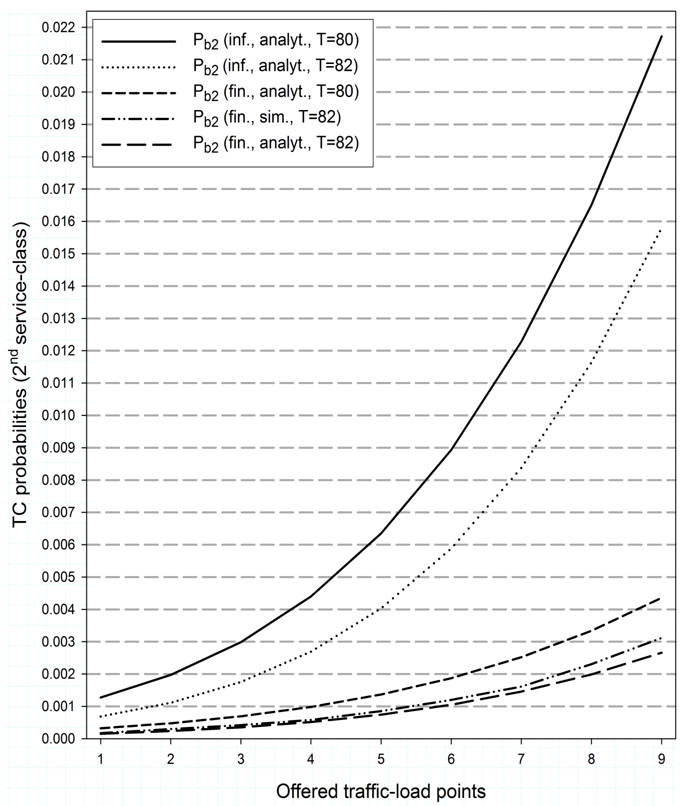

Fig. 2. TC probabilities – 2^{nd} service-class (without BR).

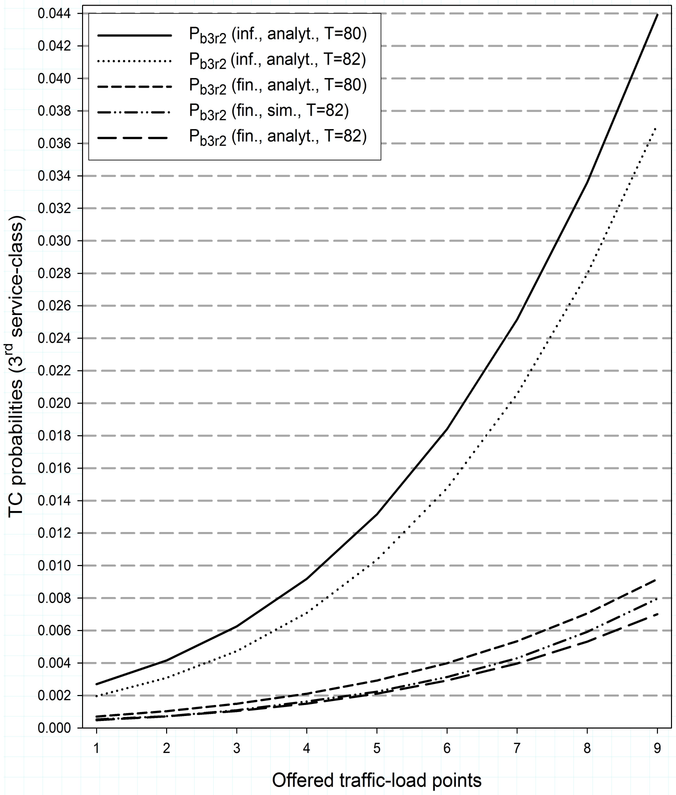

Fig. 3. TC probabilities – 3^{rd} service-class (without BR).

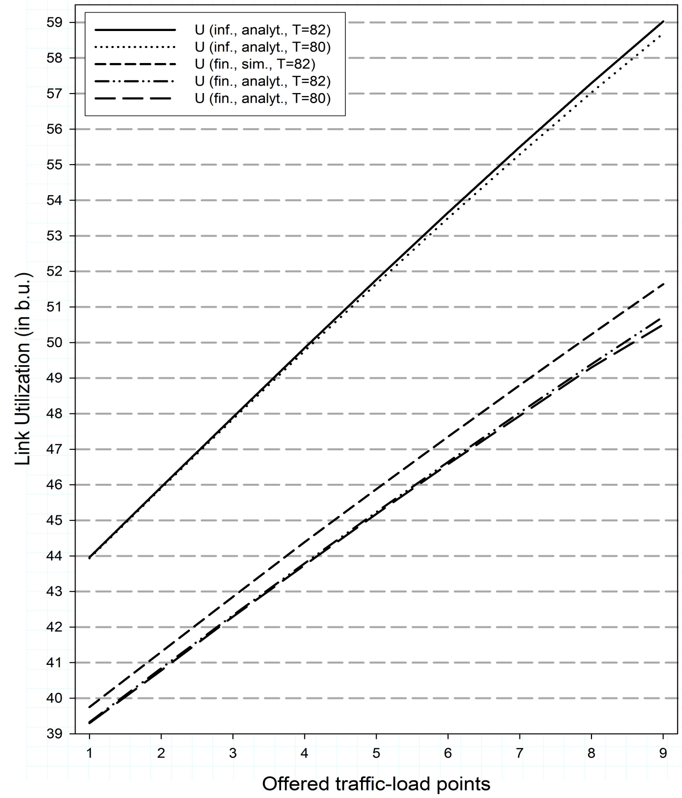

Fig. 5. Link Utilization (without BR).

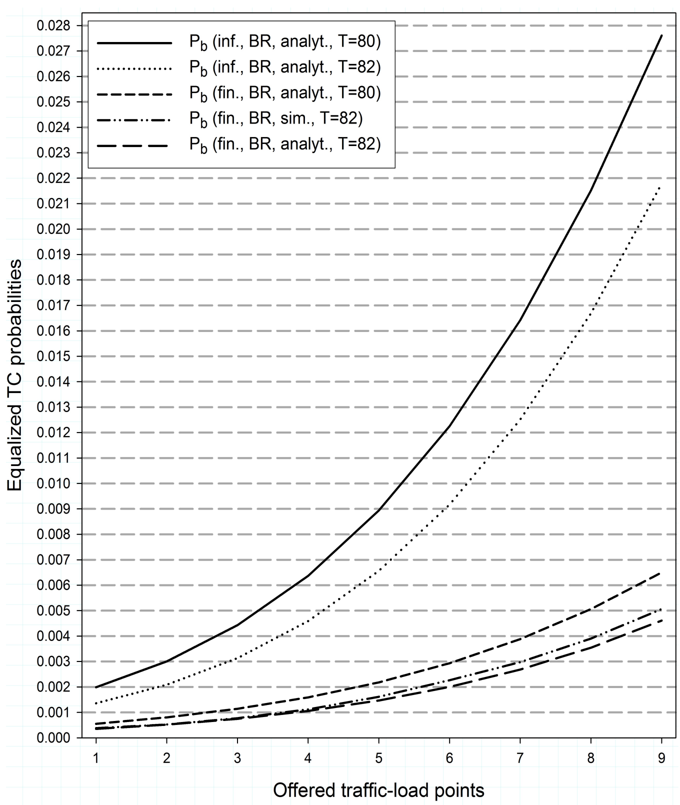

Fig. 4. Equalized TC probabilities (with BR).

VII. CONCLUSION

We propose multirate retry loss models that support elastic and adaptive traffic assuming that calls arrive to the link according to a quasi-random process and have an exponentially distributed service time. Blocked calls have the ability to retry to be connected in the system one or more times with reduced bandwidth and increased service time requirements. Furthermore, if a retry call is blocked with its last bandwidth requirement, it can still be accepted in the system by compressing its bandwidth together with the bandwidth of all in-service calls. In addition, we incorporate into our models 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 effects on the performance measures. The proposed models do not have a PFS. However, we propose approximate but recursive formulas for the calculation of the link occupancy distribution and, consequently, time and call congestion probabilities, as well as link utilization. Simulation results verify the analytical results. As a future work, we intend to study the application of these models in CDMA networks, and to consider also other bandwidth allocation policies, such as the threshold policy, whereby calls of a service-class are not allowed to enter the system (even if there is available bandwidth), if the number of in-service calls of that service-class exceeds a predefined threshold.

APPENDIX A
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), $j = 0, \ldots, T$.
$G(j)$	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 $(k = 1, \ldots, K)$
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, $\mathbf{b} = (b_1, b_2, \ldots, b_K)$.
b_{kr}	Retry bandwidth requirement of service-class k calls – single retry
b_{kr_s}	The s^{th} retry bandwidth requirement of service-class k calls – multi retrials, $s = 1, \ldots, s(k)$.
$s(k)$	Number of retrials of service-class k calls
λ_k	Mean arrival rate of service-class k idle sources
$\lambda_{k,\text{inf}}$	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
μ_{kr}^{-1}	Mean of the exponentially distributed service time of service-class k calls – single retry
$\mu_{kr_s}^{-1}$	Mean of the exponentially distributed service time of service-class k calls – multi retrials, $s = 1, \ldots, s(k)$.
α_k	Offered traffic-load (in erl) per idle source of service-class k, $\alpha_k = v_k/\mu_k$.
α_{kr}	Offered traffic-load (in erl) per idle source of service-class k – single retry, $\alpha_{kr} = v_k/\mu_{kr}$.
α_{kr_s}	Offered traffic-load (in erl) per idle source of service-class k – multi retrials, $\alpha_{kr_s} = v_k/\mu_{kr_s}$.
$\alpha_{k,\text{inf}}$	Offered traffic-load (in erl) of Poisson service-class k calls, $\alpha_{k,\text{inf}} = \lambda_{k,\text{inf}}/\mu_k$.
$\alpha_{kr,\text{inf}}$	Offered traffic-load (in erl) of Poisson service-class k calls – single retry, $\alpha_{kr,\text{inf}} = \lambda_{k,\text{inf}}/\mu_{kr}$.
$\alpha_{kr_s,\text{inf}}$	Offered traffic-load (in erl) of Poisson service-class k calls – multi retrials, $\alpha_{kr_s,\text{inf}} = \lambda_{k,\text{inf}}/\mu_{kr_s}$.
n_k	Number of in-service calls of service-class k
$\mathbf{n}$	Vector of all in service calls of all service-classes, $\mathbf{n} = (n_1, n_2, \ldots, n_K)$.
$P(\mathbf{n})$	Steady state distribution
$b_k^{'}$	Compressed bandwidth of service-class k calls
r	Compression factor
$b_{kr}^{'}$	Compressed bandwidth of service-class k calls with single retry
$t(k)$	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

APPENDIX B
LIST OF ACRONYMS

Acronym	*Meaning*
EMLM	Erlang Multirate Loss Model
EnMLM	Engset Multirate Loss Model
PFS	Product Form Solution
SRM	Single – Retry Model
MRM	Multi – Retry Model
BR	Bandwidth Reservation
QoS	Quality – of – Service
E – EMLM/BR	Extended – EMLM/BR
E – SRM/BR	Extended – SRM/BR
E – MRM/BR	Extended – MRM/BR
EF – SRM/BR	Extended Finite – SRM/BR
EF – MRM/BR	Extended Finite – MRM/BR
TC	Time Congestion
PASTA	Poisson Arrivals See Time Averages

ACKNOWLEDGMENT

The authors would like to thanks the anonymous reviewers for their constructive comments.

REFERENCES

[1] I. Moscholios, V. Vassilakis, M. Logothetis, and J. Vardakas, “Bandwidth Reservation in the Erlang Multirate Loss Model for Elastic and Adaptive Traffic,” Proc. of 9th Advanced Int. Conf. on Telecommunications, AICT 2013, Rome, Italy, 23-28 June 2013, pp. 148-153.

[2] J. Kaufman, “Blocking in a shared resource environment,” IEEE Trans. Commun. vol. 29, no. 10, pp. 1474-1481, October 1981.

[3] 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.

[4] Z. Dziong, and J. Roberts, “Congestion probabilities in a circuit switched integrated services network,” Performance Evaluation, vol. 7, issue 4, pp. 267-284, November 1987.

[5] D. Tsang, and K. Ross, “Algorithms to Determine Exact Blocking Probabilities for Multirate Tree Networks,” IEEE Trans. Commun., vol. 38, issue 8, pp. 1266-1271, August 1990.

[6] J. Kaufman, and K. Rege, “Blocking in a shared resource environment with batched Poisson arrival processes,” Performance Evaluation, vol. 24, issue 4, pp. 249-263, February 1996.

[7] G. Stamatelos, and V. Koukoulidis, “Reservation – Based Bandwidth Allocation in a Radio ATM Network,” IEEE/ACM Trans. Networking, vol. 5, pp.420-428, June 1997.

[8] 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.

[9] S. Rácz, B. Gerő, and G. Fodor, "Flow level performance analysis of a multi-service system supporting elastic and adaptive services," Performance Evaluation, vol. 49, issues 1-4, pp. 451-469, September 2002.

[10] I. Moscholios, P. Nikolaropoulos, and M. Logothetis, "Call level blocking of ON-OFF traffic sources with retrials under the complete sharing policy," Proc. of 18th ITC, Berlin, Germany, 31 August – 5 September 2003, pp. 811-820.

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

[12] Q. Huang, King-Tim Ko, and V. Iversen, "Approximation of loss calculation for hierarchical networks with multiservice overflows," IEEE Trans. Commun., vol. 56, issue 3, pp. 466-473, March 2008.

[13] I. Moscholios, J. Vardakas, M. Logothetis, and M. Koukias, "A Quasi-random Multirate Loss Model supporting Elastic and Adaptive Traffic under the Bandwidth Reservation Policy," Int. Journal on Advances in Networks and Services, vol. 6, no. 3 & 4, pp. 163-174, December 2013.

[14] M. Glabowski, and M. D. Stasiak, "Internal Blocking Probability calculation in switching networks with additional inter-stage links and mixture of Erlang and Engset Traffic," Image Processing & Communication, vol. 17, no. 1-2, pp. 67-80, January 2013.

[15] D. Staehle, and A. Mäder, "An Analytic Approximation of the Uplink Capacity in a UMTS Network with Heterogeneous Traffic," Proc. 18th ITC, Berlin, pp. 81-90, September 2003.

[16] V. Iversen, V. Benetis, N. Ha, and S. Stepanov, "Evaluation of Multi-service CDMA Networks with Soft Blocking," Proc. of ITC Specialist Seminar, Antwerp, Belgium, pp. 223-227, August/September 2004.

[17] F. Cruz-Pérez, J. Vázquez-Ávila, and L. Ortigoza-Guerrero, "Recurrent Formulas for the Multiple Fractional Channel Reservation Strategy in Multi-Service Mobile Cellular Networks," IEEE Communications Letters, vol. 8, no. 10, pp. 629-631, October 2004.

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

[19] 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, April 2009.

[20] M. Stasiak, and S. Hanczewski, "A new analytical model of UMTS cell," Image Processing & Communication, vol. 17, no. 1-2, pp. 81-90, January 2013.

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

[22] K. Kuppuswamy, and D. Lee, "An analytic approach to efficiently computing call blocking probabilities for multiclass WDM networks," IEEE/ACM Trans. Netw., vol. 17, issue 2, pp. 658-670, April 2009.

[23] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "An Analytical Approach for Dynamic Wavelength Allocation in WDM-TDMA PONs Servicing ON-OFF Traffic," IEEE/OSA Journal of Optical Commun. Netw., vol. 3, no. 4, pp. 347-358, April 2011.

[24] N. Jara, and A. Beghelli, "Blocking probability evaluation of end-to-end dynamic WDM networks," Photonic Network Communications, vol. 24, issue 1, pp. 29-38, August 2012.

[25] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "Blocking Performance of Multi-rate OCDMA PONs with QoS Guarantee," Int. Journal on Advances in Telecommunications, vol. 5, no. 3 - 4, December 2012, pp. 120-130.

[26] J. Vardakas, I. Moscholios, M. Logothetis, and V. Stylianakis, "Performance Analysis of OCDMA PONs Supporting Multi-Rate Bursty Traffic," IEEE Trans. Commun., vol. 61, no. 8, pp. 3374-3384, August 2013.

[27] G. Stamatelos, and J. Hayes, "Admission control techniques with application to broadband networks," Comput. Commun., vol. 17, no. 9, pp. 663-673, 1994.

[28] J. Kaufman, "Blocking with retrials in a completely shared resource environment," Performance Evaluation, vol. 15, issue 2, pp. 99-113, June 1992.

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

[30] I. Moscholios, M. Logothetis, and G. Kokkinakis, "A Simplified Blocking Probability Calculation in the Retry Loss Models for Finite Sources," Proc. of Communication Systems, Networks and Digital Signal Processing, Patras, Greece, 19-21 July 2006.

[31] I. Moscholios, V. Vassilakis, J. Vardakas, and M. Logothetis, "Retry loss models supporting elastic traffic," Advances in Electronics and Telecommunications, Poznan University of Technology, Poland, vol. 2, no. 3, September 2011, pp. 8-13.

[32] I. Moscholios, V. Vassilakis, J. Vardakas, and M. Logothetis, "Call Blocking Probabilities of Elastic and Adaptive Traffic with Retrials," Proc. of AICT 2012, Stuttgart, Germany, 27 May-1 June 2012.

[33] I. Moscholios, V. Vassilakis, M. Logothetis, and M. Koukias, "QoS Equalization in a Multirate Loss Model of Elastic and Adaptive Traffic with Retrials," Proc. of EMERGING 2013, Porto, Portugal, 29 September – 4 October 2013.

[34] T. Bonald, and J. Virtamo, "A recursive formula for multirate systems with elastic traffic," IEEE Commun. Letters, vol. 9, no. 8, pp. 753-755, August 2005.

[35] T. Bonald, "A Recursive Formula for Estimating the Packet Loss Rate in IP Networks," Proc. of Valuetools, Pisa, Italy, October 2009.

[36] T. Bonald, and J. Virtamo, "Calculating the flow level performance of balanced fairness in tree networks," Performance Evaluation, vol. 58, issue 1, pp. 1-14, October 2004.

[37] T. Bonald, L. Massoulie, A. Proutiere, and J. Virtamo, "A queueing analysis of max-min fairness, proportional fairness and balanced fairness," Queueing Systems, vol. 53, issues 1-2, pp. 65-84, June 2006.

[38] G. Fodor, and M. Telek, "Bounding the Blocking Probabilities in Multirate CDMA Networks Supporting Elastic Services," IEEE/ACM Trans. Netw., vol. 15, issue 4, pp. 944-956, August 2007.

[39] C. Tarhini, and T. Chahed, "QoS-oriented resource allocation for streaming flows in IEEE802.16e Mobile WiMAX," Telecommunication Systems, vol.51, issue 1, pp. 65-71, September 2012.

[40] B. Gerő, P. Pályi, and S. Rácz, "Flow-level performance analysis of a multi-rate system supporting stream and elastic services," Int. Journal of Communication Systems, vol. 26, issue 8, pp. 974-988, August 2013.

[41] M. Karrey, "Analytical Evaluation of QoS in the Downlink of OFDMA Wireless Cellular Networks Serving Streaming and Elastic Traffic," IEEE Trans. on Wireless Communications, vol. 9, no. 5, pp. 1799 1807, May 2010.

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

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

[44] I. Moscholios, M. Logothetis, and G. Kokkinakis, "Connection Dependent Threshold Model: A Generalization of the Erlang Multiple Rate Loss Model," Performance Evaluation, vol.48, issues 1-4, pp. 177-200, May 2002.

[45] J. Roberts, "Teletraffic models for the Telecom 1 Integrated Services Network," Proceedings of ITC-10, Mondreal, Canada, 1983.

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

[47] M. Glabowski, M. Stasiak, and J. Weissenberg, "Properties of recurrent equations for the full-availability group with BPP traffic," Mathematical Problems in Engineering, vol. 2012, Article ID 547909, 17 pages, 2012. doi:10.1155/2012/547909.

[48] Simscript III, http://www.simscript.com (retrieved: May 2014).