Retry Loss Models Supporting Elastic Traffic

Ioannis D. Moscholios, Vassilios G. Vasilakis, John S. Vardakas and Michael D. Logothetis

***Abstract*—We consider a single-link loss system of fixed capacity, which accommodates *K* service-classes of Poisson traffic with elastic bandwidth-per-call requirements. When a new call cannot be accepted in the system with its peak-bandwidth requirement, it can retry one or more times (single and multi-retry loss model, respectively) to be connected in the system with reduced bandwidth requirement and increased service time, exponentially distributed. Furthermore, if its last bandwidth requirement is still higher than the available link bandwidth, it can be accepted in the system by compressing not only the bandwidth of all in-service calls (of all service-classes) but also its last bandwidth requirement. The proposed model does not have a product form solution and therefore we propose an approximate recursive formula for the calculation of the link occupancy distribution and consequently call blocking probabilities. The accuracy of the proposed formula is verified by simulation and is found to be quite satisfactory.**

***Index Terms*—Markov chain, call blocking, recursive formula, retry, elastic services**

I. INTRODUCTION

MULTI-RATE loss models are extensively used in the literature for the call-level QoS assessment of modern telecom networks. This assessment is critical not only for the bandwidth allocation among calls of different service-classes but also for the avoidance of over-dimensioning of a network. Despite of its importance, the call-level QoS assessment remains an open issue, due to the existence of elastic traffic in modern telecom networks. By the term "elastic traffic" we mean calls whose assigned bandwidth can be compressed or expanded during their lifetime in the system. Modeling elastic traffic at call-level can be based on the classical Erlang Multirate Loss Model (EMLM) ([1], [2]) which has been widely used in wired (e.g. [3], [4], [5], [6]), wireless (e.g. [7], [8], [9], [10]) and optical networks (e.g. [11], [12], [13], [14], [15]) to model systems that accommodate calls of different service-classes with different traffic and bandwidth requirements.

In the EMLM, calls of different service-classes arrive at a link of capacity C, following a Poisson process, and compete for the available link bandwidth under the complete sharing policy (all calls compete for all bandwidth resources). If upon arrival a call's bandwidth requirement is not available, the call is blocked and lost. Otherwise, it remains in the system for a generally distributed service time [1]. The analysis of the EMLM shows that the steady state distribution of in-service calls has a product form solution (PFS) [16]. Exploiting this fact, an accurate recursive formula (known as Kaufman-Roberts formula, KR formula) has been separately proposed by Kaufman [1] and Roberts [2] which determines the link occupancy distribution and simplifies the determination of call blocking probabilities (CBP). In [17], [18], the EMLM is extended to the retry models, in which blocked calls can immediately reattempt (one or more times – single-retry loss model (SRM) and multi-retry loss model (MRM), respectively) to be connected by requiring less bandwidth units (b.u.), while increasing their service time which is exponentially distributed, so that the product (service time) by (bandwidth per call) remains constant. A retry call is blocked and lost from the system when its last bandwidth requirement is higher than the available link bandwidth.

In this paper, we extend the models of [17], [18], by incorporating elastic traffic. We name the proposed single-retry loss model, Extended SRM (E-SRM) and the multi-retry loss model, Extended MRM (E-MRM). In the proposed models, when a retry call attempts to be connected in the system and its last bandwidth requirement is higher than the available link bandwidth, the system accepts this call (contrary to [17], [18], where this call is lost) by compressing not only the bandwidth of all in-service calls (of all service-classes) but also the last bandwidth requirement of the retry call. The corresponding service times are increased so that the product (service time) by (bandwidth per call) remains constant. On the other hand when an in-service call, whose bandwidth is compressed, departs from the system then the remaining in-service calls (of all service-classes) expand their bandwidth. A retry call is blocked and lost from the system when the compressed bandwidth should be less than a minimum proportion (r_{min}) of its required last-bandwidth. Note that r_{min} is common for all service-classes. The compression/expansion mechanism together with the existence of retrials destroys reversibility in the proposed models and therefore no PFS exists. However, we propose approximate recursive formula for the calculation of the link occupancy distribution that simplifies the CBP determination. Simulation results validate the accuracy of the proposed formulas. In the case of no retrials for calls of all service-classes, the proposed models coincide with the model of [19] which has incorporated elastic traffic in the EMLM. We name this model, Extended EMLM (E-EMLM).

The remainder of this paper is as follows: In Section II we review the SRM, MRM and E-EMLM. In Section III, we present the proposed E-SRM and E-MRM and provide formulas for the approximate calculation of the link occupancy distribution and CBP. In Section IV, we present numerical and

Ioannis D. Moscholios is with Dept. of Telecommunications Science and Technology, University of Peloponnese, 221 00 Tripolis, Greece. Email: idm@uop.gr

Vassilios G. Vasilakis is with WCL, Dept. of Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece. Email: vasilak@wcl.ee.upatras.gr

John S. Vardakas is with WCL, Dept. of Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece. Email: jvardakas@wcl.ee.upatras.gr

Michael D. Logothetis is with WCL, Dept. of Electrical and Computer Engineering, University of Patras, 265 04 Patras, Greece. Email: m-logo@wcl.ee.upatras.gr

simulation results in order to validate the models' accuracy. We conclude in Section V.

II. Review of the Retry Loss Models and the E-EMLM

A. Review of the single and multi-retry loss models

Consider a link of capacity C b.u. that accommodates calls of K service-classes. Let j be the occupied link bandwidth, $j = 0, 1, \ldots, C$. Calls of each service-class k $(k = 1, \ldots, K)$ arrive in the link according to a Poisson process with rate λ_k and request b_k b.u. If b_k b.u. are available, a call of service-class k remains in the system for an exponentially distributed service-time with mean μ_k^{-1}. Otherwise, the call is blocked and retries immediately to be connected in the system with "retry parameters" (b_{kr}, μ_{kr}^{-1}) where $b_{kr} < b_k$ and $\mu_{kr}^{-1} > \mu_k^{-1}$. The SRM does not have a PFS and therefore the calculation of the link occupancy distribution, $G(j)$, is based on an approximate recursive formula, [17], [18]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k=1}^{K}\alpha_k b_k G(j-b_k)+ \\ +\frac{1}{j}\sum\limits_{k=1}^{K}\alpha_{kr} b_{kr}\gamma_{kr}(j) G(j-b_{kr}) & \text{for } j=1,\ldots,C \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $\alpha_k = \lambda_k\mu_k^{-1}$, $\alpha_{kr} = \lambda_k\mu_{kr}^{-1}$, $\gamma_{kr}(j) = 1$ when $j > C-(b_k - b_{kr})$, otherwise $\gamma_{kr}(j) = 0$.

Equation (1) is based on two assumptions: 1) the application of *Local Balance* (LB), which exists only in PFS models and 2) the application of *Migration Approximation* (MA) which assumes that the occupied link bandwidth from retry calls is negligible when the link occupancy is below or equal to the retry boundary, i.e. when $j \leq C - (b_k - b_{kr})$. The existence of the MA in eq. (1) is expressed by the variable $\gamma_{kr}(j)$.

The final CBP of a service-class k, denoted as B_{kr}, is the probability of a call to be blocked with its retry bandwidth requirement and is given by:

$$B_{kr} = \sum_{j=C-b_{kr}+1}^{C} G^{-1}G(j), \quad (2)$$

where $G = \sum_{j=0}^{C} G(j)$ is the normalization constant and $b_{kr} > 0$.

In the MRM, a blocked call of service-class $k$can have multiple retrials with "retry parameters" $(b_{kr_l}, \mu_{kr_l}^{-1})$ for $l = 1, \ldots, s(k)$, 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 MRM does not have a PFS and therefore the calculation of the link occupancy distribution, $G(j)$, is based on an approximate recursive formula [18]:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{j}\sum\limits_{k=1}^{K}a_k b_k G(j-b_k)+ \\ +\frac{1}{j}\sum\limits_{k=1}^{K}\sum\limits_{l=1}^{s(k)}a_{kr_l} b_{kr_l}\gamma_{kr_l}(j) G(j-b_{kr_l}) & \text{for } j=1,\ldots,C \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where: $a_{kr_l} = \lambda_k\mu_{kr_l}^{-1}$, $\gamma_{kr_l}(j) = 1$, if $C \geq j > C-(b_{kr_{l-1}} - b_{kr_l})$, otherwise $\gamma_{kr_l}(j) = 0$.

The final CBP of a service-class k, denoted as $B_{kr_{s(k)}}$, is the probability of a call to be blocked with its last retry bandwidth requirement and is given by:

$$B_{kr_{s(k)}} = \sum_{j=C-b_{kr_{s(k)}}+1}^{C} G^{-1}G(j), \quad (4)$$

where $G = \sum_{j=0}^{C} G(j)$ and $b_{kr_l} > 0$ for $l = 1, \ldots, s(k)$.

B. Review of the E-EMLM

Consider again a link of capacity C b.u. that accommodates Poisson arriving calls of K service-classes. A call of service-class k $(k = 1, \ldots, K)$ arrives in the system with rate λ_k and requests b_k b.u. (peak-bandwidth requirement). If $j + b_k \leq C$, the call is accepted in the system with its peak-bandwidth requirement and remains in the system for an exponentially distributed service time with mean μ_k^{-1}. If $T \geq j + b_k > C$ the call is accepted in the system by compressing not only its bandwidth requirement but also the bandwidth of all in-service calls. The compressed bandwidth of the new service-class k call is:

$$b'_k = rb_k = \frac{C}{j'}b_k, \quad (5)$$

where $r \equiv r(\boldsymbol{n}) = C/j'$, $j' = j + b_k = \boldsymbol{nb} + b_k$ and T is the limit (in b.u.) up to which bandwidth compression is permitted.

Similarly, the bandwidth of all in-service calls will be compressed and become equal to $b'_i = \frac{C}{j'}b_i$ for $i = 1, \ldots, K$. After the compression of both the new call and the in-service calls the state of the system is $j = C$. The minimum bandwidth that a call of service-class k (either new or in-service) can tolerate is given by the expression:

$$b'_{k,\min} = r_{\min}b_k = \frac{C}{T}b_k, \quad (6)$$

where $r_{min} = C/T$ is the minimum proportion of the required peak-bandwidth and is common for all service-classes.

This means that if upon arrival of a service-class k call, with peak-bandwidth requirement b_k b.u., we have $j = j + b_k > T$ (or equivalently, $j' > T$ or $C/j' < r_{min}$) then the call is blocked and lost without further affecting the system.

After the bandwidth compression, calls increase their service time so that the product (service time) by (bandwidth per call) remains constant. Thus, due to bandwidth compression calls of service-class k may remain in the system more than μ_k^{-1} time units. Increasing the value of T, decreases r_{min} and increases the delay of calls of service-class k (compared to the initial service time μ_k^{-1}). Therefore the value of T can be chosen so that this delay remains within acceptable levels.

The compression/expansion of bandwidth destroys reversibility in the E-EMLM and therefore no PFS exists. However, in [19] an approximate recursive formula is proposed which determines $G(j)$'s:

$$G(j)=\begin{cases} 1 & \text{for } j=0 \\ \frac{1}{\min(j,C)}\sum\limits_{k=1}^{K} a_k b_k G(j-b_k) & \text{for } j=1,\cdots,T \\ 0 & \text{otherwise} \end{cases}, \tag{7}$$

where $\alpha_k=\lambda_k\mu_k^{-1}$.

Equation (7) is based on a reversible Markov chain which approximates the bandwidth compression/expansion mechanism of the E-EMLM, described above. The LB equations of this Markov chain are of the form [19]:

$$\lambda_k P(\boldsymbol{n}_k^-)=n_k\mu_k\phi_k(\boldsymbol{n})P(\boldsymbol{n}), \tag{8}$$

where $P(\boldsymbol{n})$ is the probability distribution of state $\boldsymbol{n}=n_1,n_2,\ldots,n_k,\ldots,n_K)$, $P(\boldsymbol{n}_k^-)$ is the probability distribution of state $\boldsymbol{n}_k^-=(n_1,n_2,\ldots,n_{k-1},n_k-1,n_{k+1},\ldots,n_K)$ and $\phi_k(\boldsymbol{n})$ is a state dependent factor which describes: i) the compression factor of bandwidth and ii) the increase factor of service time of service-class k calls in state $\boldsymbol{n}$, so that (service time) by (bandwidth per call) remains constant. In other words, $\phi_k(\boldsymbol{n})$ has the same role with $r(\boldsymbol{n})$ in eq. (5) or r_{min} in eq. (6) but it may be different for each service-class. It is apparent now why the model of eq. (7) approximates the E-EMLM. The values of $\phi_k(\boldsymbol{n})$ are given by:

$$\phi_k(\boldsymbol{n})=\begin{cases} 1 & \text{, for } \boldsymbol{nb}\le C,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ \frac{x(\boldsymbol{n}_k^-)}{x(\boldsymbol{n})} & \text{, for } C<\boldsymbol{nb}\le T,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases}, \tag{9}$$

where $\boldsymbol{\Omega}=\{\boldsymbol{n}:0\le\boldsymbol{nb}\le T$ and $\boldsymbol{nb}=\sum_{k=1}^{K}n_k b_k$.

In eq. (9), $x(\boldsymbol{n})$ is a state multiplier, associated with state $\boldsymbol{n}$, whose values, are chosen so that eq. (8) holds, [19]:

$$x(\boldsymbol{n})=\begin{cases} 1 & \text{, for } \boldsymbol{nb}\le C,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ \frac{1}{C}\sum\limits_{k=1}^{K} n_k b_k x(\boldsymbol{n}_k^-) & \text{, for } C<\boldsymbol{nb}\le T,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases}. \tag{10}$$

Having determined the values of $G(j)$'s we can calculate CBP according to the following formula:

$$B_k=\sum_{j=T-b_k+1}^{T} G^{-1}G(j), \tag{11}$$

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

III. Retry loss models Supporting Elastic Traffic

A. The extended single-retry loss model

The proposed E-SRM is a non-PFS model that combines the characteristics of the SRM and the E-EMLM. In order to provide an approximate but recursive formula for the calculation of the link occupancy distribution we present the following simple example.

Consider a link of capacity C b.u. that accommodates Poisson arriving calls of two service-classes with traffic parameters: $(\lambda_1,\mu_1^{-1},b_1)$ f or the 1st service-class and $(\lambda_2,\mu_2^{-1},\mu_{2r}^{-1},b_2,b_{2r})$ for the 2nd service-class. Calls of the 2nd service-class have "retry parameters" with $b_{2r}<b_2$ and $\mu_{2r}^{-1}>\mu_2^{-1}$. Let T be the limit up to which bandwidth compression is permitted for calls of both service-classes

Although the E-SRM is a non-PFS model we will use the LB eq. (8), initially for calls of the 1st service-class:

$$\lambda_1 P(\boldsymbol{n}_1^-)=n_1\mu_1\phi_1(\boldsymbol{n})P(\boldsymbol{n}), \tag{12}$$

for $1\le\boldsymbol{nb}\le T$, where $\boldsymbol{n}=(n_1,n_2,n_{2r})$, $\boldsymbol{n}_1^-=(n_1-1,n_2,n_{2r})$ with $n_1\ge1$ and

$$\phi_1(\boldsymbol{n})=\begin{cases} 1 & \text{, for } \boldsymbol{nb}\le C,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ \frac{x(\boldsymbol{n}_1^-)}{x(\boldsymbol{n})} & \text{, for } C<\boldsymbol{nb}\le T,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases} \tag{13}$$

with $\boldsymbol{nb}=j=\sum_{k=1}^{2}(n_k b_k+n_{kr}b_{kr})=n_1b_1+n_2b_2+n_{2r}b_{2r}$.

Based on eq. (13) and multiplying both sides of eq. (12) with b_1 we have:

$$a_1 b_1 x(\boldsymbol{n})P(\boldsymbol{n}_1^-)=n_1 b_1 x(\boldsymbol{n}_1^-)P(\boldsymbol{n}), \tag{14}$$

where $\alpha_1=\lambda_1\mu_1^{-1}$ and the values of $x(\boldsymbol{n})$ are given by:

$$x(\boldsymbol{n})=\begin{cases} 1 & \text{, for } \boldsymbol{nb}\le C,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ \frac{1}{C}\sum\limits_{k=1}^{K} n_k b_k x(\boldsymbol{n}_k^-)+ +n_{kr}b_{kr}x(\boldsymbol{n}_{kr}^-) & \text{, for } C<\boldsymbol{nb}\le T,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases}. \tag{15}$$

To derive the corresponding LB equations of 2nd service-class calls consider that a call of the 2nd service-class arrives in the system when the occupied link bandwidth is j b.u. with $j=0,1,\ldots,T$. If $j\le C-b_2$, the call will be accepted in the system with b_2 b.u. If $j>C-b_2$, the call will be blocked with its b_2 requirement and will immediately try to be connected in the system with $b_{2r}<b_2$. We consider three cases: 1) If $j+b_{2r}\le C$ the retry call will be accepted in the system with b_{2r}. 2) If $j+b_{2r}>T$ the retry call will be blocked and lost. 3) If $C<j+b_{2r}\le T$ the retry call will be accepted in the system by compressing not only its bandwidth requirement b_{2r} but also the bandwidth of all in-service calls. The compressed bandwidth of the retry call is $b'_{2r}=rb'_{2r}=\frac{C}{j'}b'_{2r}$ where $r=C/j$, $j'=j+b_{2r}=\boldsymbol{nb}+b_{2r}$. Similarly, the bandwidth of all in-service calls will be compressed (by the same factor) and become $b'_i=\frac{C}{j'}b_i$ for $i=1,2$. After the compression of both the new call and the in-service calls the state of the system is $j=C$. The minimum bandwidth that a call of the 2nd service-class (either new or in-service) can tolerate is: $b'_{2r,\min}=r_{\min}b_{2r}=\frac{C}{T}b_{2r}$.

Based on the previous discussion we consider the following LB equations for calls of the 2nd service-class:

a)
$$\lambda_2 P(n_2^-)=n_2\mu_2\phi_2(n)P(n), \tag{16}$$

for $1\le\boldsymbol{nb}\le C$, where $\boldsymbol{n}=(n_1,n_2,n_{2r})$, $\boldsymbol{n}_2^-=(n_1,n_2-1,n_{2r})$ with $n_2\ge1$ and

$$\phi_2(\boldsymbol{n})=\begin{cases} 1 & \text{, for } \boldsymbol{nb}\le C,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ \frac{x(\boldsymbol{n}_2^-)}{x(\boldsymbol{n})} & \text{, for } C<\boldsymbol{nb}\le T,\ \boldsymbol{n}\in\boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases}. \tag{17}$$

Based on eq. (17) and multiplying both sides of eq. (16) with b_2 we have:

$$a_2 b_2 x(\boldsymbol{n}) P(\boldsymbol{n}_2^-) = n_2 b_2 x(\boldsymbol{n}_2^-) P(\boldsymbol{n}), \quad (18)$$

for $1 \leq \boldsymbol{nb} \leq C$, where $\alpha_2 = \lambda_2 \mu_2^{-1}$ and the values of $x(\boldsymbol{n})$ are given by eq. (15).

b) $$\lambda_2 P(\boldsymbol{n}_{2r}^-) = n_{2r} \mu_{2r} \phi_{2r}(\boldsymbol{n}) P(\boldsymbol{n}), \quad (19)$$

for $C - b_2 + b_{2r} < \boldsymbol{nb} \leq T$, where$P(\boldsymbol{n}_{2r}^-)$ is the probability distribution of state $\boldsymbol{n}_{2r}^- = (n_1, n_2, n_{2r} - 1)$ and

$$\phi_{2r}(\boldsymbol{n}) = \begin{cases} 1 & \text{, for } \boldsymbol{nb} \leq C,\, \boldsymbol{n} \in \boldsymbol{\Omega} \\ \frac{x(\boldsymbol{n}_{2r}^-)}{x(\boldsymbol{n})} & \text{, for } C < \boldsymbol{nb} \leq T,\, \boldsymbol{n} \in \boldsymbol{\Omega} \\ 0 & \text{, otherwise} \end{cases} . \quad (20)$$

Based on eq. (20) and multiplying both sides of eq. (19) with b_{2r} we have:

$$a_{2r} b_{2r} x(\boldsymbol{n}) P(\boldsymbol{n}_{2r}^-) = n_{2r} b_{2r} x(\boldsymbol{n}_{2r}^-) P(\boldsymbol{n}), \quad (21)$$

for $C - b_2 + b_{2r} < \boldsymbol{nb} \leq T$, where $\alpha_{2r} = \lambda_{2r} \mu_{2r}^{-1}$ and the values of $x(\boldsymbol{n})$ are given by eq. (15).

Equations (14), (18) and (21) lead to the following system of equations:

$$a_1 b_1 x(\boldsymbol{n}) P(\boldsymbol{n}_1^-) + a_2 b_2 x(\boldsymbol{n}) P(\boldsymbol{n}_2^-) = (n_1 b_1 x(\boldsymbol{n}_1^-) + n_2 b_2 x(\boldsymbol{n}_2^-)) P(\boldsymbol{n}), \quad (22)$$

for 1$\leq \boldsymbol{nb} \leq C - b_2 + b_{2r}$,

$$a_1 b_1 x(\boldsymbol{n}) P(\boldsymbol{n}_1^-) + a_2 b_2 x(\boldsymbol{n}) P(\boldsymbol{n}_2^-) + a_{2r} b_{2r} x(\boldsymbol{n}) P(\boldsymbol{n}_{2r}^-) = (n_1 b_1 x(\boldsymbol{n}_1^-) + n_2 b_2 x(\boldsymbol{n}_2^-) + n_{2r} b_{2r} x(\boldsymbol{n}_{2r}^-)) P(\boldsymbol{n}), \quad (23)$$

for $C - b_2 + b_{2r} < \boldsymbol{nb} \leq C$,

$$a_1 b_1 x(\boldsymbol{n}) P(\boldsymbol{n}_1^-) + a_{2r} b_{2r} x(\boldsymbol{n}) P(\boldsymbol{n}_{2r}^-) = (n_1 b_1 x(\boldsymbol{n}_1^-) + n_{2r} b_{2r} x(\boldsymbol{n}_{2r}^-)) P(\boldsymbol{n}) \quad (24)$$

or $C < \boldsymbol{nb} \leq T$.

Equations (22)-(24) can be combined into one equation by assuming that calls with b_{2r} are negligible when $1 \leq \boldsymbol{nb} \leq C - b_2 + b_{2r}$ (MA) and calls with b_2 are negligible when $C < \boldsymbol{nb} \leq T$:

$$a_1 b_1 x(\boldsymbol{n}) P(\boldsymbol{n}_1^-) + a_2 b_2 \gamma_2(\boldsymbol{nb}) x(\boldsymbol{n}) P(\boldsymbol{n}_2^-) + a_{2r} b_{2r} \gamma_{2r}(\boldsymbol{nb}) x(\boldsymbol{n}) P(\boldsymbol{n}_{2r}^-) = (n_1 b_1 x(\boldsymbol{n}_1^-) + n_2 b_2 x(\boldsymbol{n}_2^-) + n_{2r} b_{2r} x(\boldsymbol{n}_{2r}^-)) P(\boldsymbol{n}), \quad (25)$$

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

Note that the approximations introduced in eq. (25) are similar to those introduced in the single- threshold model of [18].

Since $x(\boldsymbol{n}) = 1$, when $0 \leq \boldsymbol{nb} \leq C$, it is proved in [18] that:

$$a_1 b_1 G(j - b_1) + a_2 b_2 G(j - b_2) + a_{2r} b_{2r} \gamma_{2r}(j) G(j - b_{2r}) = j G(j), \quad (26)$$

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

To prove eq. (26), the MA is needed, which assumes that the population of retry calls of the 2nd service-class is negligible in states $j \leq C - b_2 + b_{2r}$.

When $C < \boldsymbol{nb} \leq T$ and based on eq. (15), eq. (25) can be written as:

$$a_1 b_1 P(\boldsymbol{n}_1^-) + a_{2r} b_{2r} \gamma_{2r}(\boldsymbol{nb}) P(\boldsymbol{n}_{2r}^-) = C P(\boldsymbol{n}). \quad (27)$$

To introduce the link occupancy distribution $G(j)$ in eq. (27) we sum both sides of eq. (27) over the set of states $\boldsymbol{\Omega}_j = \{\boldsymbol{n} \in \boldsymbol{\Omega}\, | \boldsymbol{nb} = j\}$:

$$a_1 b_1 \sum_{\{\boldsymbol{n}|\boldsymbol{nb}=j\}} P(\boldsymbol{n}_1^-) + a_{2r} b_{2r} \gamma_{2r}(\boldsymbol{nb}) \sum_{\{\boldsymbol{n}|\boldsymbol{nb}=j\}} P(\boldsymbol{n}_{2r}^-) = C \sum_{\{\boldsymbol{n}|\boldsymbol{nb}=j\}} P(\boldsymbol{n}). \quad (28)$$

Since by definition $\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} P(\boldsymbol{n}) = G(j)$, eq. (28) is written as:

$$a_1 b_1 G(j - b_1) + a_{2r} b_{2r} \gamma_{2r}(j) G(j - b_{2r}) = C G(j), \quad (29)$$

where $\gamma_{2r}(j) = 1$ for $C < j \leq T$.

The combination of eq. (26) and eq. (29) gives the following approximate recursive formula for the calculation of $G(j)$'s in the case of two service-classes when only calls of the 2^{nd} service-class have "retry parameters":

$$G(j) = \frac{1}{\min(j,C)} \left[a_1 b_1 G(j - b_1) + a_2 b_2 \gamma_2(j) G(j - b_2) + a_{2r} b_{2r} \gamma_{2r}(j) G(j - b_{2r}) \right] \quad (30)$$

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

In the case of K service-classes and assuming that all service-classes may have "retry parameters", eq. (30) takes the general form:

$$G(j) = \begin{cases} 1 & \text{, for } j=0 \\ \frac{1}{\min(j,C)} \sum_{k=1}^{K} \alpha_k b_k \gamma_k(j) G(j - b_k) + \\ + \frac{1}{\min(j,C)} \sum_{k=1}^{K} \alpha_{kr} b_{kr} \gamma_{kr}(j) G(j - b_{kr}) & \text{, for } j=1, \ldots, T, \\ 0 & \text{, otherwise} \end{cases} \quad (31)$$

where $\gamma_k(j) = 1$ for $1 \leq j \leq C$, otherwise $\gamma_k(j) = 0$ and $\gamma_{kr}(j) = 1$ for $C - b_k + b_{kr} < j \leq T$, otherwise $\gamma_{kr}(j) = 0$.

The final CBP of a service-class k, B_{kr}, is the probability of a call to be blocked with its retry bandwidth requirement:

$$B_{kr} = \sum_{j=T-b_{kr}+1}^{T} G^{-1} G(j), \quad (32)$$

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

B. The extended multi-retry loss model

Similar to the MRM, in the E-MRM a blocked call of service-class k can have more than one "retry parameters" $(b_{kr_l}, \mu_{kr_l}^{-1})$ for $l = 1, \ldots, s(k)$, 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 E-MRM

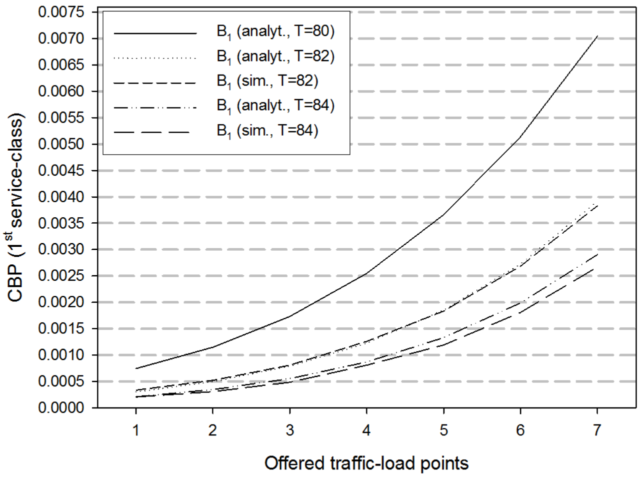

Fig. 1. CBP for the 1st service-class.

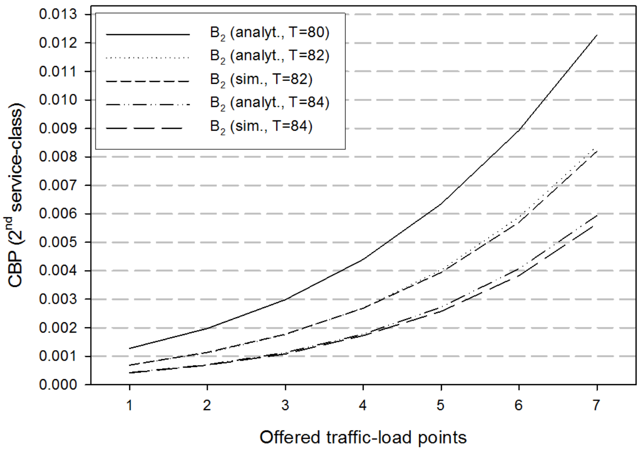

Fig. 2. CBP for the 2nd service-class.

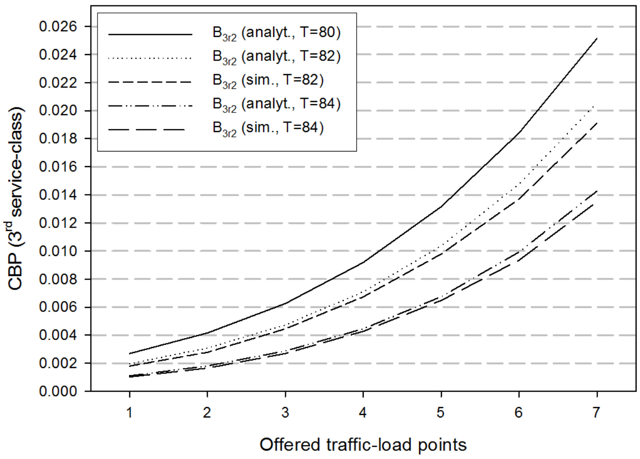

Fig. 3. CBP for the 3rd service-class (retry calls with b_{3r2}).

does not have a PFS and therefore the calculation of the occupancy distribution, $G(j)$, is based on an approximate recursive formula whose proof is similar to that of eq. (31):

$$G(j)=\begin{cases} 1 & ,\text{for } j=0 \\ \frac{1}{\min(j,C)}\left(\sum_{k=1}^{K} a_k b_k \gamma_k(j) G(j-b_k)+ \right. \\ \left. +\sum_{k=1}^{K}\sum_{l=1}^{s(k)} a_{kr_l} b_{kr_l} \gamma_{kr_l}(j) G(j-b_{kr_l})\right) & ,\text{ for } j=1,\dots,T, \\ 0 & ,\text{ otherwise} \end{cases} \quad (33)$$

where: $a_{kr_l} = \lambda_k \mu_{kr_l}^{-1}$, $\gamma_k(j) = 1$ for $1 \leq j \leq C$, otherwise $\gamma_k(j) = 0$ and $\gamma_{kr_l}(j) = 1$, if $j > C - (b_{krl-1} - b_{krl})$, otherwise $\gamma_{kr_l}(j) = 0$.

The final CBP of a service-class k, denoted as $B_{kr_{s(k)}}$, is the probability of a call to be blocked with its last retry bandwidth requirement and is given by:

$$B_{kr_{s(k)}} = \sum_{j=T-b_{kr_{s(k)}}+1}^{T} G^{-1} G(j), \quad (34)$$

where $G = \sum_{j=0}^{T} G(j)$ and $b_{krl} > 0$ for $l = 1, \dots, s(k)$.

IV. Evaluation

In this section, we present an application example and compare the analytical CBP probabilities with those obtained by simulation. The latter is based on SIMSCRIPT II.5 [20]. Simulation results are mean values of 7 runs with 95% confidence interval. Since, the resultant reliability ranges of the measurements are small enough we present only mean values.

Consider a link of capacity $C = 80$ b.u. that accommodates three service-classes of elastic calls which require $b_1 = 1$ b.u., $b_2 = 2$ b.u. and $b_3 = 6$ b.u., respectively. All calls arrive in the system according to a Poisson process. The call holding time is exponentially distributed with mean value $\mu_1^{-1} = \mu_2^{-1} = \mu_3^{-1} = 1$. The initial values of the offered traffic-load are: $\alpha_1 = 20$ erl, $\alpha_2 = 6$ erl and $\alpha_3 = 2$ erl. Calls of the 3rd service-class may retry two times with reduced bandwidth requirement: $b_{3_{r1}} = 5$ b.u. and $b_{3_{r2}} = 4$ b.u. and increased service time so that $\alpha_3 b_3 = a_{3_{r1}} b_{3_{r1}} = a_{3_{r2}} b_{3_{r2}}$. In the x-axis of all figures, we assume that α_3 remains constant while α_1, α_2 increase in steps of 1.0 and 0.5 erl, respectively. The last value of $\alpha_1 = 26$ erl while that of $\alpha_2 = 9$ erl.

We consider three different values of T: a) $T = C = 80$ b.u., where no bandwidth compression takes place. In that case, the proposed E-MRM gives exactly the same CBP results with the MRM of [18], b) $T = 82$ b.u. where bandwidth compression takes place and $r_{min} = C/T = 80/82$ and c) $T = 84$ b.u. where bandwidth compression takes place and $r_{min} = C/T = 80/84$.

In Fig. 1, we present the analytical and simulation CBP results of the 1st service-class for all values of T. Similar results are presented in Fig. 2, for the 2nd service-class and in Fig. 3 for the 3rd service-class (CBP of calls with $b_{3_{r2}}$). All figures presented herein show that: i) the model's accuracy is absolutely satisfactory compared to simulation and ii) the increase of T above C results in a CBP decrease due to the existence of the compression mechanism.

V. Conclusion

We propose multirate loss models that support elastic traffic under the assumption that Poisson arriving calls have the ability, when blocked with their initial bandwidth requirement, to retry to be connected in the system one (E-SRM) or more times (E-MRM) 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 not only the bandwidth of all in-service calls (of all service-classes) but also its last bandwidth requirement. The proposed models do not have a PFS and therefore we propose approximate but recursive formulas for the CBP calculation. Simulation results verify the analytical results. As a future work, we will examine multirate retry loss models that support both elastic and adaptive traffic (e.g. adaptive video). Adaptive calls can tolerate bandwidth compression, but their service time cannot be altered.

References

[1] J. S. Kaufman, "Blocking in a shared resource environment," *IEEE Trans. Commun.*, vol. 29, no. 10, pp. 1474–1481, Oct. 1981.

[2] J. W. Roberts, *Performance of Data Communications systems and their applications*. North Holland, Amsterdam, 1981, ch. A service system with heterogeneous user requirements, pp. 423–431.

[3] M. Logothetis and G. Kokkinakis, "Path Bandwidth Management for Large Scale Telecom Networks," *IEICE Trans. Commun.*, vol. E83-B, no. 9, pp. 2087–2099, Sep. 2000.

[4] M. Stasiak and M. Głąbowski, "A simple approximation of the link model with reservation by a one-dimensional Markov chain," *Journal of Performance Evaluation*, vol. 41, no. 2-3, pp. 195–208, Jul. 2000.

[5] I. Moscholios, M. Logothetis, and G. Kokkinakis, "Call-burst blocking of ON-OFF traffic sources with retrials under the complete sharing policy," *Journal of Performance Evaluation*, vol. 59, no. 4, pp. 279–312, Mar. 2005.

[6] W. Bziuk, "Approximate state probabilities in large shared multirate loss systems with an application to trunk reservation," *European Trans. Telecom*, vol. 16, no. 3, pp. 205–216, May 2005.

[7] W. Choi and J. Y. Kim, "Joint Erlang Capacity of DS/CDMA Forward Link Based on Resource Sharing Algorithm," *IEICE Trans. Fundamentals*, vol. E84-A, no. 6, pp. 1406–1412, Jun. 2001.

[8] P. Fazekas, S. Imre, and M. Telek, "Modelling and Analysis of Broadband Cellular Networks with Multimedia Connections," *Telecommunication systems*, vol. 19, no. 3-4, pp. 263–288, Mar. 2002.

[9] D. Staehle and A. Mäder, "An Analytic Approximation of the Uplink Capacity in a UMTS Network with Heterogeneous Traffic," in *Proc. 18th International Teletraffic Congress (ITC)*, Berlin, Sep. 2003, pp. 81–90.

[10] M. Głąbowski, M. Stasiak, A. Wiśniewski, 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.

[11] A. Washington and H. Perros, "Call blocking probabilities in a traffic-groomed tandem optical network," *Computer Networks*, vol. 45, no. 3, pp. 281–294, Jun. 2004.

[12] A. Sahasrabudhe and D. Manjunath, "Performance of optical burst switched networks: A two moment analysis," *Computer Networks*, vol. 50, no. 18, pp. 3550–3563, Dec. 2006.

[13] J. Vardakas, V. Vassilakis, and M. Logothetis, "Blocking Analysis in Hybrid TDM-WDM Passive Optical Networks," in *Proc. 5th Int. Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NETs 2008)*, Karlskrona, Sweden, Feb. 2008.

[14] K. Kuppuswamy and D. Lee, "An analytic approach to efficiently computing call blocking probabilities for multiclass WDM networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 2, pp. 658–670, Apr. 2009.

[15] J. S. Vardakas, I. D. Moscholios, M. D. Logothetis, and V. G. Stylianakis, "An Analytical Approach for Dynamic Wavelength Allocation in WDM-TDMA PONs Servicing ON-OFF Traffic," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 3, no. 4, pp. 347–358, Apr. 2011.

[16] K. W. Ross, *Multiservice loss models for broadband telecommunication networks*. Springer, Berlin, 1995.

[17] J. S. Kaufman, "Blocking in a Completely Shared Resource Environment With State Dependent Resource and Residency Requirements," in *Proc. IEEE INFOCOM'92*, 1992, pp. 2224–2232.

[18] ——, "Blocking with retrials in a completely shared resource environment," *Journal of Performance Evaluation*, vol. 15, no. 2, pp. 99–113, Jun. 1992.

[19] G. Stamatelos and V. Koukoulidis, "Reservation – Based Bandwidth Allocation in a Radio ATM Network," *IEEE/ACM Trans. Netw.*, vol. 5, no. 3, pp. 420–428, Jun. 1997.

[20] "Simscript II.5," [online], http://www.simscript.com.

Ioannis D. Moscholios was born in Athens, Greece, in 1976. He received the Dipl.-Eng. degree in Electrical & Computer Engineering from the University of Patras, Patras, Greece, in 1999, the M.Sc. degree in Spacecraft Technology & Satellite Communications from the University College London, UK, in 2000 and the Ph.D. degree in Electrical & Computer Engineering from the University of Patras, in 2005. From 2005 to 2009 he was a Research Associate at the Wire Communications Laboratory, Dept. of Electrical & Computer Engineering, University of Patras. Currently, he is a Lecturer in the Dept. of Telecommunications Science and Technology, University of Peloponnese, Greece. His research interests include simulation and performance analysis of communication networks. He has published over 65 papers in international journals/ conferences and has over 120 third-part citations. He is a member of the Technical Chamber of Greece (TEE).

Vassilios G. Vassilakis was born in Sukhumi, Georgia, in 1978. He received his Dipl.-Eng. degree in Computer Engineering & Informatics and PhD in Electrical and Computer Engineering, both from the University of Patras, Greece, in 2003 and 2011, respectively. Hi is involved in national research and R&D projects. His main research interests are in the areas of teletraffic engineering and QoS in wireless networks. He has published over 20 papers in international journals/ conferences and has over 40 thirs-part citations. He is a member of the Technical Chamber of Greece (TEE).

John S.Vardakas was born in Alexandria, Greece, in 1979. He received his Dipl.-Eng. degree in Electrical & Computer Engineering from the Democritus University of Thrace, Greece, in 2004. Since 2005 he is a Ph.D student at the Wire Communications Laboratory, Department of Electrical and Computer Engineering, University of Patras, Greece. His research interests include teletraffic engineering in optical and wireless networks. He is a member of the Optical Society of America (OSA) and the Technical Chamber of Greece (TEE).

Michael D. Logothetis was born in Stenies, Andros, Greece, in 1959. He received his Dipl.-Eng. degree and Doctorate in Electrical Engineering, both from the University of Patras, Patras, Greece, in 1981 and 1990, respectively. From 1982 to 1990, he was a Teaching and Research Assistant at the Laboratory of Wire Communications, University of Patras, and participated in many national and EU research programmes, dealing with telecommunication networks, as well as with office automation. From 1991 to 1992 he was Research Associate in NTT's Telecommunication Networks Laboratories, Tokyo, Japan. Afterwards, he was a Lecturer in the Dept. of Electrical & Computer Engineering of the University of Patras, and recently he has been elected Professor in the same Department. His research interests include teletraffic theory and engineering, traffic/network control, simulation and performance optimization of telecommunications networks. He has published over 120 conference/journal papers and has over 220 third-part citations. He has published a teletraffic book in Greek. He has organised the 5th International Conference on Communications Systems, Networks and Digital Signal Processing, CSNDSP 2006. He served/is serving on the Technical Program Committee of several international conferences while he organizes and chairs several technical sessions. He has become a Guest Editor in three journals, while participates in the Editorial Board of journals. He is a member of the IARIA (Fellow), IEEE (Senior), IEICE, ETRI, FITCE and the Technical Chamber of Greece (TEE).