PAPER

An ON-OFF Multi-Rate Loss Model of Finite Sources

Ioannis D. MOSCHOLIOS[†a)], *Nonmember*, Michael D. LOGOTHETIS[†b)], *Member*, *and* Michael N. KOUKIAS[†], *Nonmember*

SUMMARY Bursty traffic is dominant in modern communication networks and keeps the call-level QoS assessment an open issue. ON-OFF traffic models are commonly used to describe bursty traffic. We propose an ON-OFF traffic model of a single link which accommodates service-classes of finite population (f-ON-OFF). Calls compete for the available link bandwidth under the complete sharing policy. Accepted calls enter the system via state ON and then may alternate between ON-OFF states. When a call is transferred to state OFF it releases the bandwidth held in state ON, while when a call tries to return to state ON, it re-requests its bandwidth. If it is available a new ON-period (burst) begins; otherwise the call remains in state OFF (burst blocking). We prove that the proposed f-ON-OFF model has a product form solution, and we provide an accurate recursive formula for the call blocking probabilities calculation. For the burst blocking probabilities calculation we propose an approximate but robust formula. In addition, we show the relation between the f-ON-OFF model and other call-level loss models. Furthermore, we generalize the f-ON-OFF model to include service-classes of both finite and infinite population. Simulation results validate our analytical methodology.
key words: *call blocking, burst blocking, ON-OFF model, finite population, recursive formula*

1. Introduction

In the multi-service environment of IPv6 based networks with bandwidth reservation capabilities, but also in the emerging and future all-optical core networks, such as MPλS/GMPLS [1], [2] or 3G wireless networks (e.g. UMTS) [3], a large variety of applications exist, like the mission-critical Internet services of on-line banking and stock trading, e-commerce, e-voting and many video communications (including video-on-demand). All these applications have as a common feature, the bursty nature of traffic, while bandwidth reservation is critical to their performance [4]. In such networking environment, the call-level QoS assessment is important for the bandwidth allocation among service-classes, the avoidance of too costly over-dimensioning of the network and the prevention, through traffic engineering mechanisms, of excessive throughput degradation [5].

Contrary to the constant/peaked bandwidth allocated service-classes (stream traffic), where calls continuously hold their assigned bandwidth, calls of a bursty service-class continuously seize and release bandwidth during their lifetime. The released bandwidth may be seized by other calls and thus bandwidth utilization is improved. The easiest way to describe the behaviour of bursty traffic is by the ON-OFF traffic model where Poisson arriving calls, when accepted in the system, alternate between transmission (ON) periods (bursts at peak bit rate) and no transmission (OFF) periods.

Despite the fundamental difference between the ON-OFF and stream traffic, the springboard of call-level modelling in both cases is a stream traffic model. This is the Erlang Multirate Loss Model (EMLM) [6], where calls of different service-classes arrive to a single link of certain capacity according to a Poisson process, and compete for the available link bandwidth under the complete sharing (CS) policy, which is the most general case of resource allocation. That is, a call is admitted if bandwidth is available at the call-arrival time; the only constraint is the total link capacity — no contiguous bandwidth requirements [7], or guaranteeing QoS [4]. The calls'holding time can be arbitrarily distributed [8]. As far as the equilibrium state probabilities are concerned, the EMLM has a product form solution (PFS). This fact leads to an accurate calculation of the link occupancy distribution and, consequently, of Call Blocking Probabilities (CBP) [8], [9]. Moreover, these calculations are recursive (the most desirable feature) and lead to efficient computer implementation, broadening the EMLM's applicability range to links of large capacities. With the aid of such a teletraffic model, we can readily quantify the trade-off between CBP and throughput, and thereby provide a computational model for traffic engineering [10].

A remarkable extension of the EMLM is a PFS model with finite population of traffic sources [11]. We prefer to name it "Engset Multirate Loss Model (EnMLM)," since for a single service-class, the proposed accurate and recursive CBP formula gives the same results with the Engset formula for the time congestion probability [12].

The EMLM and the EnMLM can cope with the bursty traffic in an approximate way through the notion of equivalent bandwidth [13]. However, more sophisticated models ([14]–[16]) exist for bursty traffic and especially for the ON-OFF traffic model, in which the notion of not only call blocking (CB) but also burst blocking (BB) is introduced. When an in-service call passes from state ON to state OFF, it releases the bandwidth held in state ON. To return to state ON, it re-requests its bandwidth and, if available, a new burst begins; otherwise, the burst is blocked (burst block-

Manuscript received June 5, 2006.
Manuscript revised December 1, 2006.
[†]The authors are with the ECE department of the University of Patras, Greece.
a) E-mail: moscholios@wcl.ee.upatras.gr
b) E-mail: m-logo@wcl.ee.upatras.gr
DOI: 10.1093/ietcom/e90–b.7.1608

ing probability - BBP) and the call remains in state OFF. In [14] a single ATM link that accommodates CBR service-classes together with VBR service-classes of ON-OFF-type sources is considered. CBR calls are always in state ON, while VBR calls enter either via state ON or OFF, under a CAC criterion that takes into account not only the required bandwidth in state ON but also the mean bandwidth of the VBR calls; this fact destroys the PFS of the model. Therefore both CBP and BBP calculations are based on approximate formulas which are non-recursive and contain exponentials and factorials; the latter restricts the applicability range of the model to small systems. In [15] a similar model is proposed that leads to a recursive BBP calculation, while the CBP calculation is done approximately, via the Erlang B-formula. In [16], Poisson arriving calls enter the system via state ON only, under the CS policy, and then alternate between ON-OFF states, or remain always in state ON. This infinite source ON-OFF (inf-ON-OFF) model has a PFS; the CBP calculation is based on an accurate recursive formula. On the other hand, the BBP calculation is based on approximate formulas. The inf-ON-OFF model is an extension of the EMLM (see section III-B). This justifies the EMLM as the basis of call-level modelling of both stream and ON-OFF traffic.

In this paper, we introduce the finite source ON-OFF model (f-ON-OFF), in which a single link accommodates several service-classes with finite population of sources, under the CS policy, whereas calls enter the system via state ON only. The new model coincides with the inf-ON-OFF model when the number of sources approaches infinity, and with the EnMLM when calls of all service-classes remain always in state ON; whereas, when the number of sources approaches infinity and all calls remain always ON, the model results in the EMLM. We prove that the f-ON-OFF model has a PFS, and we provide an accurate recursive formula for calculating CBP, while the BBP calculation is done by an approximate formula. For the completeness of this study, we present a generalized model that describes a mixture of service-classes with finite and infinite population of sources. The accuracy of the proposed CBP and BBP formulas has been evaluated by simulation and found to be highly satisfactory. Furthermore, we compare the proposed f-ON-OFF model with the inf-ON-OFF model in terms of CBP, in order to show the consistency and the necessity of the proposed model.

The remainder of this paper is as follows: In Sect. 2 we present the f-ON-OFF model; firstly we describe the service model, secondly we analyze the system in order to derive the formulas for the link occupancy distribution and the CBP calculation, and thirdly we proceed to the derivation of the BBP formula. In Sect. 3, we show the relation of the proposed f-ON-OFF model with other models and discuss possible applications. In Sect. 4 we generalize the f-ON-OFF model to include a mixture of finite and infinite sources. In Sect. 5, we present numerical examples whereby we evaluate the accuracy of the analytical formulas. We conclude in Sect. 6. Finally, in Appendix we describe a heuristic procedure whereby we determine the state space of a system.

2. The Proposed f-ON-OFF Model

2.1 The Service Model

We consider a transmission link of capacity C that accommodates K independent service-classes of ON-OFF-type calls. Each call of a service-class k ($k = 1, \dots, K$) comes from a finite source population N_k, requires b_k bandwidth units (b.u.) and competes for the available link bandwidth under the CS policy. If b_k b.u. are available, then, the call enters the system via state ON (otherwise the call is blocked and lost), and the occupied link bandwidth is characterized as real. The mean arrival rate of the idle sources is $\lambda_k = (N_k - n_k^1 - n_k^2)v_k$, where n_k^i is the number of in-service calls of service-class k in the *ith* state ($i = 1 \Rightarrow$ state ON, $i = 2 \Rightarrow$ state OFF), while v_k is the arrival rate per idle source (constant). This call arrival process is characterized as quasi-random [12] (if $N_k \to \infty$ for $k = 1, \dots, K$, then the arrival process is Poisson). The call holding time in state ON or OFF is exponentially distributed.

At the end of an ON-period a service-class k call releases b_k b.u. and begins an OFF-period with probability σ_k, or departs from the system with probability $1 - \sigma_k$. While the call is in state OFF, it is assumed that it seizes fictitious bandwidth (b_k b.u.) of a fictitious link of capacity C^*. At the end of an OFF-period the call returns to state ON with probability 1 (i.e. the call can not leave the system from state OFF), while re-requesting b_k b.u. When $C = C^*$, b_k b.u. are always available for that call in state ON, i.e. no BB occurs. When $C < C^*$, and there is available bandwidth in the link, i.e. if $j_1 + b_k \le C$ (where j_1 is the occupied real link bandwidth), the call returns to state ON and a new burst begins; otherwise BB occurs, and the call remains in state OFF for another period (Fig. 1). A new service-class k call is accepted in the system with b_k b.u., if it meets the following constraints:

$$j_1 + b_k \le C \tag{1}$$

$$j_1 + j_2 + b_k \le C^* \tag{2}$$

where j_2 is the occupied bandwidth of the fictitious link.

The first constraint ensures that the occupied bandwidth of all existing ON calls together with the new call does not exceed the real link capacity. The second constraint prevents the system from accepting new calls when most of the system calls are in state OFF (and keeps BB low). Note that, when $C = C^*$, a service-class k call is accepted in the system if it meets the second constraint only.

2.2 The Analytical Call-Level Model — CBP Determination

The following assumptions and notations are necessary in our analysis:

$n^i = (n_1^i, \dots, n_k^i, \dots, n_K^i)$ – the vector of the number of

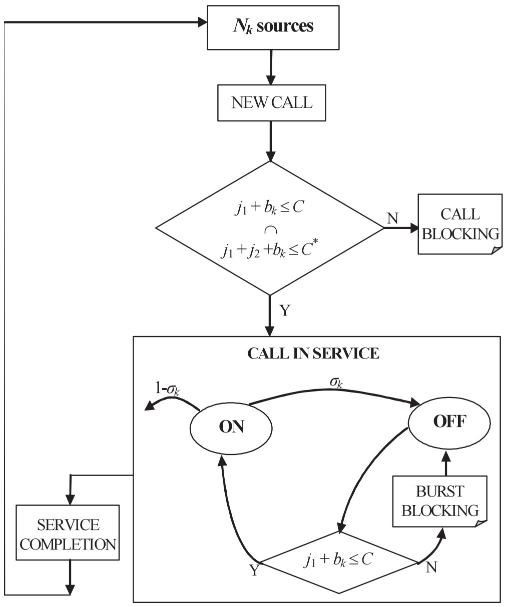


Fig. 1 The mechanisms of call and burst blocking in the f-ON-OFF model.

service-class k calls in state i ($i = 1 \Rightarrow$ state ON, $i = 2 \Rightarrow$ state OFF).

$\vec{n} = (n^1, n^2)$, $n^i_{k+} = (n^i_1, \ldots, n^i_k + 1, \ldots, n^i_K)$, $n^i_{k-} = (n^i_1, \ldots, n^i_k - 1, \ldots, n^i_K)$, $\vec{n}^1_{k+} = (n^1_{k+}, n^2)$, $\vec{n}^2_{k+} = (n^1, n^2_{k+})$, $\vec{n}^1_{k-} = (n^1_{k-}, n^2)$, $\vec{n}^2_{k-} = (n^1, n^2_{k-})$.

$\vec{n}_{k+-} = (n^1_{k+}, n^2_{k-})$, vector that shows a service-class k call transition from state OFF to ON.

$\vec{n}_{k-+} = (n^1_{k-}, n^2_{k+})$, vector that shows a service-class k call transition from state ON to OFF.

$\upsilon_{ik,fin}$ is the external arrival rate of service-class k calls to state i. Since external arrivals enter the system only in state ON, the values of $\upsilon_{ik,fin}$† are given by:

$$\upsilon_{ik,fin} = \begin{cases} \upsilon_k & \text{for } i = 1 \\ 0 & \text{for } i = 2 \end{cases}$$

$Pr_{ij}(k)$ is the Prob. of (next state is j | current state is i and the service-class is k). The values of $Pr_{ij}(k)$ are:

$$Pr_{1j}(k) = \begin{cases} 0 & \text{for } j = 1 \\ \sigma_k & \text{for } j = 2 \end{cases}, \quad Pr_{2j}(k) = \begin{cases} 1 & \text{for } j = 1 \\ 0 & \text{for } j = 2 \end{cases}$$

$e_{ik,fin}$ is the total arrival rate of service-class k call to state i. The values of $e_{ik,fin}$ are:

$$e_{ik,fin} = \upsilon_{ik,fin} + \sum_{j=1}^{2} e_{jk,fin} Pr_{ji}(k) = \begin{cases} \frac{\upsilon_k}{(1-\sigma_k)} & \text{for } i = 1 \\ \frac{\upsilon_k \sigma_k}{(1-\sigma_k)} & \text{for } i = 2 \end{cases}$$

$p_{ik,fin}$ is the offered traffic-load to state i by a service-class k call. The values of $p_{ik,fin}$ are:

$$p_{ik,fin} = \frac{e_{ik,fin}}{\mu_{ik}} = \begin{cases} \frac{\upsilon_k}{(1-\sigma_k)\mu_{1k}} & \text{for } i = 1 \\ \frac{\upsilon_k \sigma_k}{(1-\sigma_k)\mu_{2k}} & \text{for } i = 2 \end{cases}$$

where: μ^{-1}_{ik} is the mean holding time of a service-class k call to state i.

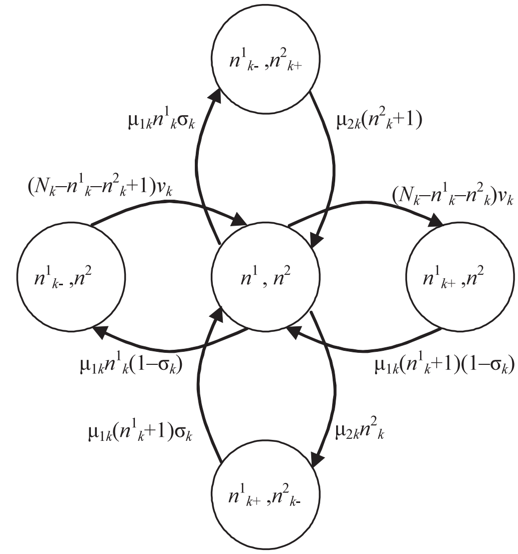


Fig. 2 The state transition diagram of the f-ON-OFF model.

Figure 2 shows the state transition rates of the f-ON-OFF model (in equilibrium). Assuming the existence of local balance (LB) between adjacent states (i.e. states that can be reached from one another with a single transition), the following LB equations (*rate up = rate down*) are extracted from the state transition diagram of Fig. 2:

$$\mu_{1k} n^1_k \sigma_k P(\vec{n}) = \mu_{2k}(n^2_k + 1) P(\vec{n}_{k-+}) \quad (3a)$$

$$\mu_{1k} n^1_k (1-\sigma_k) P(\vec{n}) = (N_k - n^1_k - n^2_k + 1)\upsilon_k P(\vec{n}^1_{k-}) \quad (3b)$$

$$(N_k - n^1_k - n^2_k)\upsilon_k P(\vec{n}) = \mu_{1k}(n^1_k + 1)(1-\sigma_k) P(\vec{n}^1_{k+}) \quad (3c)$$

$$\mu_{2k} n^2_k P(\vec{n}) = \mu_{1k}(n^1_k + 1)\sigma_k P(\vec{n}_{k+-}) \quad (3d)$$

where: $P(\vec{n})$, $P(\vec{n}_{k-+})$, $P(\vec{n}^1_{k-})$, $P(\vec{n}^1_{k+})$, $P(\vec{n}_{k+-})$ are the probability distributions of the corresponding states: $\vec{n} = (n^1, n^2)$, $\vec{n}_{k-+} = (n^1_{k-}, n^2_{k+})$, $\vec{n}^1_{k-} = (n^1_{k-}, n^2)$, $\vec{n}^1_{k+} = (n^1_{k+}, n^2)$, $\vec{n}_{k+-} = (n^1_{k+}, n^2_{k-})$, respectively. Note that the number of service class k calls in states ON and OFF should satisfy the constraints: $n^1_k + n^2_k \le N_k$, $n^1_k \le \left\lfloor \frac{C}{b_k} \right\rfloor$, $n^2_k \le \left\lfloor \frac{C^*}{b_k} \right\rfloor$, where $\lfloor x \rfloor$ is the largest integer not exceeding x.

Equations (3b) and (3c) describe the balance between the rates of a new call arrival of service-class k and the corresponding departure from the system, while Eqs. (3a) and (3d) refer to the ON-OFF alternations of a service-class k call. Based on the LB assumption, we find that the probability distribution of state $\vec{n}$, $P(\vec{n})$, can be given by a PFS†† formula which satisfies Eqs. (3a)–(3d):

†The notation *fin* in all symbols refers to the finite source model.

††The term PFS also appears to the queuing network models known to have product form state distributions (BCMP queuing networks: "Open, closed and mixed networks of queues with different classes of customers," by F. Baskett. K.M. Chandy, R.R. Muntz, and F.G. Palacios, in *Journal of ACM*, vol.22., issue 2, pp.248–260, 1975).

$$P(\vec{n})=\frac{1}{G}\prod_{k=1}^{K}\left(\binom{N_k}{n_k^1+n_k^2}\prod_{l=1}^{l=n_k^1}\frac{l+n_k^2}{l}\prod_{i=1}^{2}p_{ik,fin}^{n_k^i}\right) \quad (4)$$

where: $G \equiv G(\Omega)$ is a normalization constant chosen so that the sum of the probabilities of all states is unity, while $\vec{j}=(j_1,j_2)\in\Omega \Leftrightarrow \left\{j_1 \le C \cap \left(\sum_{s=1}^{2} j_s \le C^*\right)\right\}$,

where: $j_s=\sum_{i=1}^{2}\sum_{k=1}^{K} n_k^i b_{i,k,s}$ with $b_{i,k,s}=\begin{cases} b_k & \text{if } s=i \\ 0 & \text{if } s\neq i\end{cases}$ and $s = 1,2$ ($s=1 \Rightarrow$ state ON, $s=2 \Rightarrow$ state OFF).

The fact that the system of LB Eqs. (3a)–(3d) can be solved is very important for the proposed model, because it validates the LB assumption and proves that the model has a PFS. However, Eq. (4) is restricted to small capacity systems, because it includes factorials and exponentials; consequently, the CBP calculation is hard. Therefore, our target is to base the CBP calculation on an accurate recursive formula for the link occupancy distribution, $G(\vec{j})$, $\vec{j}=(j_1,j_2)$. This will extend the applicability range of the model to large systems, and determine the CBP fast and efficiently. To this end, we proceed as follows:

Having found an expression for $P(\vec{n})$ and since the CS policy is a coordinate convex policy†, the probability $P(\vec{n}_{k-}^i)$ can be expressed by:

$$(N_k - n_k^i - n_k^t + 1)p_{ik,fin}\gamma_{ik}(\vec{n})P(\vec{n}_{k-}^i) = n_k^i P(\vec{n}) \quad (5)$$

where $i = 1,2$, $t = 1,2$ with $t \neq i$ and $\gamma_{ik}(\vec{n}) = \begin{cases} 1 & \text{if } n_k^i \ge 1 \\ 0 & \text{if } n_k^i = 0\end{cases}$.

Equation (5) indicates that in the CS policy the transfer from state $\vec{n}$ to state $\vec{n}_{k-}^i$ is always possible by the departure of a service-class k call from the system.

Consider now the set of states, $\Omega_{\vec{j}}$, with $\vec{j}=(j_1,j_2)$, whereby the occupied real and fictitious link bandwidth is exactly j_1 and j_2, respectively. That is,

$$\Omega_{\vec{j}}=\left\{\vec{n}\in\Omega:\ \vec{n}B=\vec{j},\ n_k^i\ge 0,\ k=1,\dots,K,\ i=1,2\right\}$$

where: B is a $(2K\times 2)$ matrix with entries $b_{i,k,s}$. The $(i,k)th$ row of B is denoted by $B_{i,k}=(b_{i,k,1}\ b_{i,k,2})$; e.g. for $K=2$ service-classes with bandwidth requirement per call b_1 and b_2,

$$B=\begin{bmatrix} b_{1,1,1} & b_{1,1,2}\\ b_{2,1,1} & b_{2,1,2}\\ b_{1,2,1} & b_{1,2,2}\\ b_{2,2,1} & b_{2,2,2}\end{bmatrix}=\begin{bmatrix} b_{1,1,1} & 0\\ 0 & b_{2,1,2}\\ b_{1,2,1} & 0\\ 0 & b_{2,2,2}\end{bmatrix}.$$

In order to take into account $\vec{j}=(j_1,j_2)$ in our analysis, we sum Eq. (5) over $\Omega_{\vec{j}}$:

$$N_k p_{ik,fin}\sum_{\vec{n}\in\Omega_{\vec{j}}}\gamma_{ik}(\vec{n})P(\vec{n}_{k-}^i) - p_{ik,fin}\sum_{\vec{n}\in\Omega_{\vec{j}}}(n_k^i+n_k^t-1)\gamma_{ik}(\vec{n})P(\vec{n}_{k-}^i)=\sum_{\vec{n}\in\Omega_{\vec{j}}}n_k^i P(\vec{n}) \quad (6)$$

The left hand side (LHS) of Eq. (6) can be written as:

$$\mathbf{LHS} = N_k p_{ik,fin}\sum_{\vec{n}\in\Omega_{\vec{j}}\cap\{n_k^i\ge 1\}}P(\vec{n}_{k-}^i) - p_{ik,fin}\sum_{\vec{n}\in\Omega_{\vec{j}}\cap\{n_k^i\ge 1\}}(n_k^i+n_k^t-1)P(\vec{n}_{k-}^i)$$

In the set $\Omega_{\vec{j}}\cap\left\{n_k^i\ge 1\right\}=\left\{\vec{n}:\ \vec{n}_{k-}^iB=\vec{j}-B_{i,k}\right\}$ we represent $n_k^i\ge 1$ (or n_k^i - 1 ≥ 0) by introducing a new variable:

$$n_m^{*i}=\begin{cases} n_m^i & \text{for } m\neq k\\ n_m^i-1 & \text{for } m=k\end{cases} \quad \text{where } m=1,\dots,K$$

Thus the LHS is written as:

$$\mathbf{LHS} = N_k p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}P(\vec{n}^{*i}) - p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}(n_k^{*i}+n_k^t)P(\vec{n}^{*i}) \quad (7)$$

where: $\vec{n}^{*1}=(n^{*1},n^2)$, $\vec{n}^{*2}=(n^1,n^{*2})$, $n^{*1}=\left(n_1^{*1},\dots,n_k^{*1},\dots,n_K^{*1}\right)$ and $n^{*2}=\left(n_1^{*2},\dots,n_k^{*2},\dots,n_K^{*2}\right)$

The first term of Eq. (7) is equal to:

$$N_k p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}P(\vec{n}^{*i}) = N_k p_{ik,fin}G(\vec{j}-B_{i,k})$$

where $G(\vec{j})$ is the un-normalized link occupancy distribution of $\vec{j}=(j_1,j_2)$.

The second term of Eq. (7) is equal to:

$$\begin{aligned} &p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}(n_k^{*i}+n_k^t)P(\vec{n}^{*i})\\ &= p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}(n_k^{*i}+n_k^t)\frac{P(\vec{n}^{*i})}{G(\vec{j}-B_{i,k})}G(\vec{j}-B_{i,k})\\ &= p_{ik,fin}\sum_{\vec{n}^{*i}\in\Omega_{\vec{j}-B_{i,k}}}(n_k^{*i}+n_k^t)P(\vec{n}^{*i}\,|\,\vec{j}-B_{i,k})G(\vec{j}-B_{i,k})\\ &= p_{ik,fin}E((n_k^{*i}+n_k^t)\,|\,\vec{j}-B_{i,k})G(\vec{j}-B_{i,k})\end{aligned}$$

where $E((n_k^{*i}+n_k^t)\,|\,\vec{j}-B_{i,k})$ is the expected value of $n_k^{*i}+n_k^t$ given $\vec{j}$.

By substituting the "new" first and second terms in Eq. (7), we have:

$$\mathbf{LHS}=(N_k-E((n_k^{*i}+n_k^t)\,|\,\vec{j}-B_{i,k}))p_{ik,fin}G(\vec{j}-B_{i,k}) \quad (8)$$

The right hand side (RHS) of Eq. (6) can be written as:

$$\mathbf{RHS} = \sum_{\vec{n}\in\Omega_{\vec{j}}}n_k^iP(\vec{n})=\sum_{\vec{n}\in\Omega_{\vec{j}}}n_k^i\frac{P(\vec{n})}{G(\vec{j})}G(\vec{j}) = \sum_{\vec{n}\in\Omega_{\vec{j}}}n_k^iP(\vec{n}\,|\,\vec{j})\,G(\vec{j}) = E(n_k^i\,|\,\vec{j})G(\vec{j}) \quad (9)$$

†According to [17] for $\mathbf{n}=(n_1,n_2,\dots,n_K)$, the state space Ω is coordinate convex if: a) $\mathbf{n}\in\Omega \Rightarrow n_k\ge 0$ for $k=1,\dots,K$ and b) $\mathbf{n}\in\Omega$ and $n_k>0 \Rightarrow n_k^-\in\Omega$ where $n_k^-=(n_1,n_2,\dots,n_{k-1},n_k-1,n_{k+1},\dots,n_K)$. The second constraint implies that a departure of a service-class k call is never blocked. In non-coordinate convex policies a departure of a call can be delayed until another call (of a service-class other than k) arrives or departs.

for $\vec{j} \in \Omega$ and $k = 1, \ldots, K$.

Combining Eq. (8) with Eq. (9) it results in:

$$(N_k - E((n_k^{*i} + n_k^t) \left| \vec{j} - B_{i,k} \right.))p_{ik,fin} G(\vec{j} - B_{i,k}) = E(n_k^i \left| \vec{j} \right.) G(\vec{j}) \quad (10)$$

Multiplying both sides of Eq. (10) by $b_{i,k,s}$ and summing over $k = 1, \ldots, K$ and $i = 1, 2$ we have:

$$\sum_{i=1}^{2} \sum_{k=1}^{K} (N_k - E((n_k^{*i} + n_k^t) \left| \vec{j} - B_{i,k} \right.)) b_{i,k,s} p_{ik,fin} G(\vec{j} - B_{i,k}) = \sum_{i=1}^{2} \sum_{k=1}^{K} E(n_k^i b_{i,k,s} \left| \vec{j} \right.) G(\vec{j}) = j_s G(\vec{j}) \quad (11)$$

with $\vec{j} \in \Omega \Leftrightarrow \left\{ j_1 \leq C \bigcap \left(\sum_{s=1}^{2} j_s \leq C^* \right) \right\}$.

The expected value $E((n_k^{*i} + n_k^t) \left| \vec{j} - B_{i,k} \right.)$ in Eq. (11) is not known. To determine it, we use a lemma proposed in [11] for the determination of a similar expected value in the EnMLM. According to the lemma, two stochastic systems are equivalent and result in the same CBP, if they have:

(a) the same traffic description parameters (K, N_k, $p_{ik,fin}$, where $k = 1, \ldots, K$ and $i = 1, 2$)

(b) exactly the same set of states.

The lemma can be justified as follows. Due to the condition (a) and since no other parameters (e.g. bandwidth per call requirements) are involved in the calculation of the probability distribution of state $\vec{n}$, $P(\vec{n})$, $\forall \vec{n} \in \Omega$ (Eq. (4)), the values of $P(\vec{n})$ are the same for both stochastic systems. Besides, because of the condition (b), the two stochastic systems have the same set of blocking states. Therefore, the CBP of service-class k calls is the same for both stochastic systems.

Our purpose is therefore to find a new stochastic system, whereby we can determine the expected value $E((n_k^{*i} + n_k^t) \left| \vec{j} - B_{i,k} \right.)$. The bandwidth requirements of the service-classes and the capacities C, C^* in the new stochastic system are chosen according to the following two criteria:

- conditions (a) and (b) are valid,
- each state has a unique occupancy $\vec{j} = (j_1, j_2)$.

Now, the state $\vec{j}$ is reached only via the previous state $\vec{j} - B_{i,k}$; thus, the expected value:

$$E((n_k^{*i} + n_k^t) \left| \vec{j} - B_{i,k} \right.) = n_k^{*i} - 1 + n_k^t \quad (12)$$

According to the above, Eq. (11) can be written as:

$$\sum_{i=1}^{2} \sum_{k=1}^{K} (N_k - n_k^i - n_k^t + 1) b_{i,k,s} p_{ik,fin} G(\vec{j} - B_{i,k}) = j_s G(\vec{j}) \quad (13)$$

Equation (13) is the two-dimensional recursive formula used for the determination of $G(\vec{j})$'s. The $G(\vec{j})$'s are calculated in terms of an arbitrary $G(\vec{0})$ under the normalization condition of $\sum_{\vec{j} \in \Omega} G(\vec{j}) = 1$. Besides, it is apparent that if $\vec{j} \notin \Omega$ then $G(\vec{j}) = 0$. Although Eq. (13) is simple, it cannot be used straightforward for the $G(\vec{j})$'s determination unless an equivalent system (mentioned above) is defined by enumerating and processing the state space of the system. The following example reveals the problems that can arise when one tries to use Eq. (13) prior to the state space enumeration and processing, and how these problems can be overcome.

We consider two service-classes which require $b_1 = 3$ and $b_2 = 1$ b.u. per call, respectively. The number of sources of each service-class is $N_1 = N_2 = 5$. The real and fictitious capacities of the link are $C = C^* = 3$ b.u. The traffic description parameters of the two service-classes are:

- 1st service-class: $\upsilon_1 = 0.002$, $\mu_{11}^{-1} = 0.5$, $\mu_{21}^{-1} = 0.6$, $\sigma_1 = 0.75$, $p_{11,fin} = 0.004$, $p_{21,fin} = 0.0036$
- 2nd service-class: $\upsilon_2 = 0.01$, $\mu_{12}^{-1} = 1.0$, $\mu_{22}^{-1} = 1.0$, $\sigma_1 = 0.75$, $p_{12,fin} = 0.04$, $p_{22,fin} = 0.03$

In Table 1 we present the system state-space of this example, consisting of 12 unique states of the form $(n_1^1, n_2^1, n_1^2, n_2^2)$, the respective real and fictitious link occupancies (j_1, j_2), the real and fictitious link occupancies of the equivalent system $(j_{1(eqv)}, j_{2(eqv)})$ and the CB states for each service-class. However, the link occupancies are not unique; e.g. $(j_1, j_2) = (0, 3)$ appears twice: for $(n_1^1, n_2^1, n_1^2, n_2^2) = (0, 0, 0, 3)$ and for $(n_1^1, n_2^1, n_1^2, n_2^2) = (0, 0, 1, 0)$. Because of this, we cannot use Eq. (13), unless we find an equivalent system whose states have unique link occupancy. In both the initial and equivalent system, the CB states of each service-class are the same. Such an equivalent system has: $b_1 = 5, b_2 = 2, C = C^* = 6$. The required bandwidth per call of the service-classes of the equivalent system is chosen by a simple heuristic algorithm. This is an iterative procedure where in each iteration, the rate of each service-class is increased unit-by-unit from its initial value, and checked whether or not the equivalent system has been found, according to the criteria mentioned above. An enchanced description of the heuristic algorithm is found in the Appendix.

Having found an equivalent system, we calculate the

Table 1 State space, respective real and fictitious occupied link bandwidth, equivalent real and fictitious link occupancies and call blocking states ($C = C^*$).

n_1^1	n_2^1	n_1^2	n_2^2	j_1	j_2	$j_{1(eqv)}$	$j_{2(eqv)}$	CB states 1st service	CB states 2nd service
0	0	0	0	0	0	0	0		
0	0	0	1	0	1	0	2	✓	
0	0	0	2	0	2	0	4	✓	
0	0	0	3	0	3	0	6	✓	✓
0	0	1	0	0	3	0	5	✓	✓
0	1	0	0	1	0	2	0	✓	
0	1	0	1	1	1	2	2	✓	
0	1	0	2	1	2	2	4	✓	✓
0	2	0	0	2	0	4	0	✓	
0	2	0	1	2	1	4	2	✓	✓
0	3	0	0	3	0	6	0	✓	✓
1	0	0	0	3	0	5	0	✓	✓

$G(\vec{j})$'s via Eq. (13), and then determine the CBP of the service-classes, P_{b_k} $(k = 1, 2)$, by using the following formula:

$$P_{b_k} = \sum_{\left\{\vec{j} \,\middle|\, [(b_k+j_1)>C] \,\cup\, [(b_k+j_1+j_2)>C^*]\right\}} G(\vec{j}) \tag{14}$$

where $\sum_{\vec{j}\in\Omega} G(\vec{j}) = 1$.

For the record, the resultant CBP in our example are: $P_{b_1} = 30.58\%$, $P_{b_2} = 2.88\%$.

2.3 The Analytical Burst-Level Model — BBP Determination

To illustrate the idea behind the proposed formula for the BBP determination, we consider again the previous example but now the fictitious capacity is increased to $C^* = 4$ b.u (while $C = 3$ b.u.). This increase results in the appearance of BB for calls of both service-classes; 1st and 2nd service-class OFF calls will face BB when $j_1 + b_1 > C$ and $j_1 + b_2 > C$, respectively. In Table 2 we present the system state-space, the respective real and fictitious link occupancies and the real and fictitious link occupancies of the equivalent system, as well as the CB and BB states for each service class. The state space now consists of 19 states. An equivalent system is the following: $b_1 = 5$, $b_2 = 2$, $C = 6$, $C^* = 8$. Both systems, initial and equivalent, have the same CB and BB states presented in Table 2.

In order to determine the BBP of calls of a service-class k, $P^*_{b_k}$, we need to find the number of service-class k calls in state OFF, n^2_k, when the system is in a BB state; e.g. according to Table 2, a BB state of the 2nd service-class, is $(n^1_1, n^1_2, n^2_1, n^2_2) = (1, 0, 0, 1)$ and the corresponding value of $\vec{j} = (j_{1(eqv)}, j_{2(eqv)}) = (5, 2)$. Multiplying n^2_k by the corresponding $G(\vec{j})$ and the service rate in state OFF μ_{2k}, we obtain the rate whereby service-class k OFF-calls would depart from the BB state if it were possible. By summing these rates over the BB state-space Ω^* of each service-class k, we obtain the following summation:

$$\sum_{(\vec{j}\in\Omega^*)} n^2_k G(\vec{j})\mu_{2k} \text{ where } \vec{j} \in \Omega^* \Leftrightarrow$$

$$\left\{C - b_k + 1 \le j_1 \le C \bigcap \left(\sum_{s=1}^{2} j_s \le C^*\right)\right\}$$

By normalizing it (taking into account the whole state-space Ω), we obtain the following formula for the BBP calculation:

$$P^*_{b_k} = \frac{\sum_{(\vec{j}\in\Omega^*)} n^2_k G(\vec{j})\mu_{2k}}{\sum_{(\vec{j}\in\Omega)} n^2_k G(\vec{j})\mu_{2k}} \tag{15}$$

where: $\vec{j} \in \Omega \Leftrightarrow \left\{j_1 \le C \bigcap \left(\sum_{s=1}^{2} j_s \le C^*\right)\right\}$ and the values of $G(\vec{j})$'s are calculated by Eq. (13).

Thus, Eq. (15) can be seen as the normalised rate of the service-class k OFF-calls by which OFF-calls would depart from the BB states if it were possible.

For the record, the BBP of this example are: $P^*_{b_1} = 14.81\%$, $P^*_{b_2} = 1.50\%$, while the corresponding simulation results (with 95% confidence interval) are: $P^*_{b_1 sim} = 14.32 \pm 0.31\%$, $P^*_{b_2 sim} = 1.48 \pm 0.05\%$.

3. The f-ON-OFF Model in Relation to Other Multi-rate Loss Models and Their Applications

3.1 To the inf-ON-OFF Model

The proposed f-ON-OFF model leads to the inf-ON-OFF model, when $N_k \to \infty$ for each service-class k $(k = 1, \ldots, K)$. In this case, the call arrival process changes from quasi-random to Poisson, while the formula for the determination of $G(\vec{j})$'s changes from Eq. (13) to the following [16]:

$$\sum_{i=1}^{2}\sum_{k=1}^{K} b_{i,k,s} p_{ik} G(\vec{j} - B_{i,k}) = j_s G(\vec{j}) \tag{16}$$

where:

$$p_{ik} = \frac{e_{ik}}{\mu_{ik}} = \begin{cases} \frac{\lambda_k}{(1-\sigma_k)\mu_{1k}} & for\ i = 1 \\ \frac{\lambda_k \sigma_k}{(1-\sigma_k)\mu_{2k}} & for\ i = 2 \end{cases} \tag{17}$$

where: e_{ik} is the total arrival rate of service-class k calls to the ith state.

The CBP calculation is done via Eq. (14), while the BBP is determined approximately by [16]:

$$P^*_{b_k} = \frac{\overline{n}^2_k \mu_{2k}}{\overline{n}^1_k \mu_{1k} - \lambda_k(1 - P_{b_k})} - 1 \tag{18}$$

Table 2 State space, respective real and fictitious occupied link bandwidth, equivalent real and fictitious link bandwidth, call and burst blocking states ($C = C^*$).

n^1_1	n^1_2	n^2_1	n^2_2	j_1	j_2	$j_{1(eqv)}$	$j_{2(eqv)}$	CB states 1st service	CB states 2nd service	BB states 1st service	BB states 2nd service
0	0	0	0	0	0	0	0				
0	0	0	1	0	1	0	2				
0	0	0	2	0	2	0	4	✓			
0	0	0	3	0	3	0	6	✓			
0	0	0	4	0	4	0	8	✓	✓		
0	0	1	0	0	3	0	5	✓			
0	0	1	1	0	4	0	7	✓	✓		
0	1	0	0	1	0	2	0	✓		✓	
0	1	0	1	1	1	2	2	✓		✓	
0	1	0	2	1	2	2	4	✓		✓	
0	1	0	3	1	3	2	6	✓	✓	✓	
0	1	1	0	1	3	2	5	✓	✓	✓	
0	2	0	0	2	0	4	0	✓		✓	
0	2	0	1	2	1	4	2	✓		✓	
0	2	0	2	2	2	4	4	✓	✓	✓	
0	3	0	0	3	0	6	0	✓	✓	✓	✓
0	3	0	1	3	1	6	2	✓	✓	✓	✓
1	0	0	0	3	0	5	0	✓	✓	✓	✓
1	0	0	1	3	1	5	2	✓	✓	✓	✓

where $\overline{n}_k^i$ is the average number of the k service-class calls in the i th state given by [16]:

$$\overline{n}_k^i = \sum_{(\vec{j} \in \Omega)} p_{ik} G(\vec{j} - B_{\mathrm{i},k}) \tag{19}$$

The derivation of Eq. (18) is based on the assumption that a call of service-class k will not be blocked more than once each time it is in state OFF, i.e. a call will be at most two consecutive times in state OFF. For values of practical interest ($P_{b_k}^* \leq 10^{-4}$ [6]), Eq. (18) provides quite satisfactory results compared to simulation results. However, Eq. (18) fails in cases where the BBP are extremely high, or extremely low [18]. To overcome this problem, we have proposed a robust BBP calculation [18]:

$$P_{b_k}^* = \frac{\sum_{(\vec{j} \in \Omega^*)} y_{2k}(\vec{j}) G(\vec{j}) \mu_{2k}}{\sum_{(\vec{j} \in \Omega)} y_{2k}(\vec{j}) G(\vec{j}) \mu_{2k}} \tag{20}$$

where: $y_{ik}(\vec{j})$ is the average number of service-class k calls in state i ($i = 1 \Rightarrow$ state ON, $i = 2 \Rightarrow$ state OFF) of the system state $\vec{j} = (j_1, j_2)$. The values of $y_{ik}(\vec{j})$ are calculated via the formula:

$$y_{ik}(\vec{j}) = \frac{p_{ik} G(\vec{j} - B_{i,k})}{G(\vec{j})} \tag{21}$$

while $G(\vec{j})$'s are calculated by Eq. (16). For the proof of Eq. (21) see [19].

The logic behind Eq. (20) is similar to that of Eq. (15). Besides, both equations do not need the restrictive assumption of the two consecutive times that a call may remain in state OFF, used in Eq. (18). Note that one can use Eq. (20) (instead of Eq. (15)) for the BBP calculation in the f-ON-OFF model and get the same BBP results, if the values of $y_{ik}(\vec{j})$ are determined as follows (for $i = 2$):

$$\begin{aligned} y_{ik}(\vec{j}) &\equiv E(n_k^i \big| \vec{j}\,) \\ &= \frac{(N_k - n_k^i - n_k^t + 1) p_{ik,fin} G(\vec{j} - B_{i,k})}{G(\vec{j})} \end{aligned} \tag{22}$$

while $G(\vec{j})$'s are calculated by Eq. (13).

Equation (22) results from Eq. (10), via Eq. (12) by a simple substitution.

3.2 To the EnMLM, the EMLM and the Delbrouck's Model

The f-ON-OFF model and the EnMLM are equivalent, in the sense that they provide the same CBP, when $\sigma_k = 0$ (i.e. when there is no OFF state) for each service-class k ($k = 1, \ldots, K$). In this case the calculation of the link occupancy distribution can be done either by Eq. (13) or by the EnMLM formula, [11]:

$$\sum_{k=1}^{K} (N_k - n_k + 1)\, \alpha_k\, b_k\; G(j - b_k) = jG(j) \tag{23}$$

where: $\alpha_k = \upsilon_k/\mu_k$, μ_k is the mean holding time of calls of service-class k, while n_k is the number of in-service sources of service-class k.

When $\sigma_k > 0$ and $C = C^*$, then the mean holding time μ_k^{-1} of a service-class k call of the f-ON-OFF model is exponentially distributed and can be determined by the following formula:

$$\mu_k^{-1} = \frac{\sigma_k}{(1 - \sigma_k)} (\mu_{1k}^{-1} + \mu_{2k}^{-1}) + \mu_{1k}^{-1} \tag{24}$$

Note that, if BB occurs ($C < C^*$), the total call holding time would remain exponentially distributed, given that consecutive BB events were independent from each other.

The summation $(\mu_{1k}^{-1} + \mu_{2k}^{-1})$ is possible, since the mean holding time in the i th state, μ_{ik}^{-1}, is exponentially distributed. The ratio $\frac{\sigma_k}{(1-\sigma_k)}$ shows the average number of times that a service-class k call visits state OFF during its lifetime. Therefore the f-ON-OFF model is equivalent to the EnMLM with traffic parameters υ_k, b_k and μ_k^{-1} given by Eq. (24). Exactly the same relationship (i.e. Eq. (24)) exists between the inf-ON-OFF model and the EMLM ([16]).

The EnMLM leads to the EMLM, when $N_k \to \infty$ for each service-class k. Thus, from Eq. (23) the following recursive formula is obtained [8], [9]:

$$\sum_{k=1}^{K} \alpha_k\, b_k\; G(j - b_k) = jG(j) \tag{25}$$

where $\alpha_k = \lambda_k/\mu_k$.

Another interesting relation of the proposed f-ON-OFF model is the relation with the Delbroucks model, in which the call arrival process is more peaked or variable than the Poisson process [20]. More precisely, each service-class k offers traffic load $\alpha_k = \lambda_k/\mu_k$, at a peakedness factor Z_k whose values can be equal, or not, to 1. The link occupancy distribution can be calculated via the recursive formula:

$$\sum_{k=1}^{K} \hat{\alpha}_k\, b_k \sum_{y=1}^{\lfloor j/b_k \rfloor} \beta_k^{y-1} G(j - y\, b_k) = jG(j) \tag{26}$$

where $\hat{\alpha}_k = \frac{\alpha_k}{Z_k}$, $\beta_k = 1 - \frac{1}{Z_k}$ and $\lfloor j/b_k \rfloor$ the greatest integer not exceeding j/b_k.

If $Z_k < 1$, then the Bernoulli process comes as a result, and $\beta_k < 0$ but with $-\hat{\alpha}_k/\beta_k$ a positive integer for equilibrium to exist. If $Z_k = 1$, then the Poisson process is the result, and $\beta_k = 0$ so that $\beta_k^0 = 1$ but $\beta_k^x = 0$ for $x > 0$. If $Z_k > 1$, then the resultant process is the Pascal process, and $0 < \beta_k < 1$ for equilibrium to exist [21].

In [22], Roberts shows that if the arrival process in the EMLM changes from Poisson to a state-dependent Poisson process:

$$\lambda_k(n_k) = \tilde{\alpha}_k + \tilde{\beta}_k(n_k) \tag{27}$$

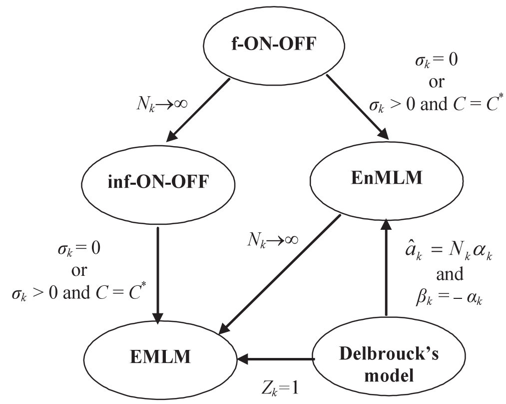


Fig. 3 Relation between the f-ON-OFF model and other multi-rate loss models.

then the Delbrouck's recursive formula results from replacing $\hat{\alpha}_k$ with $\tilde{\alpha}_k/\mu_k$ and β_k with $\tilde{\beta}_k/\mu_k$.

Equation (27) holds also for $\tilde{\beta}_k < 0$ if the resultant $\lambda_k(n_k)$ is non-negative. In that case we have the finite source model [6]. By replacing in Eq. (26) $\hat{\alpha}_k$ with $N_k\alpha_k$ and β_k with $-\alpha_k$, where $\alpha_k = v_k/\mu_k$, the EnMLM and the Delbrouck's model coincide. Since we have already shown the relation between the f-ON-OFF model and the EnMLM, it is apparent that the same relation exists between the f-ON-OFF and the Delbrouck's model. Figure 3 summarizes the relations between all models.

3.3 To Their Applications

From the above discussion it is apparent that the proposed f-ON-OFF model is strongly related to other models: the inf-ON-OFF model, the EnMLM, the EMLM and the Delbrouck's model. These relations reveal various possible applications of the f-ON-OFF model, in the same areas that the other (four) models have already been used. The inf-ON-OFF model is applicable in ATM networks [16], by using the fast buffer reservation (FBR) scheme. In the FBR scheme, each call has two states, active (ON) or idle (OFF). An active call is assigned a number of link buffers (according to its predefined bandwidth requirements) and is guaranteed access to them until it becomes idle. When the call becomes idle the link buffers are released. The transition between an active and an idle state occurs upon receipt of a marked cell which denotes the start (start cell – active state) or the end (end cell – idle state) of a burst. All the other cells in the burst which are between the start cell and the end cell are marked as middle cells. The scheme also allows for the transmission of low priority cells with no QoS guarantee. These cells are discarded in the case of congestion. Note that the cell's marking (start cell, middle cell, end cell and low priority cell) can be realized by using two bits in the ATM cell header (the one is the CLP bit and the second can be one of the three bits of the payload-type) [23]. The EnMLM has been applied in a radio ATM local-area network [24]; this is also a potential application of the f-ON-OFF model, because it exploits both the finite and the bursty features of the model. The EMLM has been extensively used in both wired and wireless environments (e.g. [25]–[32]).

Furthermore, the f-ON-OFF model is mostly applicable to other modern networks supporting bursty traffic, e.g. IPv6 based networks, which are extended with resource allocation capabilities (like MPLS). Moreover, it can be favourably used in wireless networks, because of the finite number of sources. We can concentrate, for example, on a single base station of a cellular network (UMTS) with several multirate/multimedia service-classes, aiming at determining call-level system parameters: CBP of new calls (new connections admitted to the base station within its coverage area) and failure probabilities of handover calls (calls that attempt to be connected to the base station from the neighbourhood of the cell). The base station can be modelled as a system (transmission link) of certain capacity. Usually it is supposed that the number of users roaming in the cell or in the vicinity of the cell approaches infinity, and therefore the EMLM is used as the basis of the call-level QoS assessments [31], [32]. Recently, an approximate model for the uplink blocking probability calculation of cellular/WCDMA systems that takes into account only the finite population was presented in [33]. However, the consideration of both the finite population and the bursty nature of traffic are more realistic, and therefore the call-level assessment should be based on the f-ON-OFF model. In a small cell, however, the mobility of the users is an important factor and affects CBP; the introduction of users' mobility to the f-ON-OFF model is for further study.

4. Generalization of the f-ON-OFF Model to Include Service-Classes with a Mixture of Finite and Infinite Sources

Consider a transmission link with a pair of capacities C (real) and C^* (fictitious), accommodating K_{fin} service-classes of finite ON-OFF sources (quasi-random input) and K_{inf} service-classes of infinite ON-OFF sources (random -Poisson input). Then the calculation of the link occupancy distribution is done by the combination of Eq. (13) and Eq. (16) [34]:

$$\sum_{i=1}^{2}\sum_{k=1}^{K_{fin}}(N_k - n_k^i - n_k^t + 1)b_{i,k,s}p_{ik,fin}G(\vec{j} - B_{i,k}) + \sum_{i=1}^{2}\sum_{k=1}^{K_{inf}} b_{i,k,s}p_{ik}G(\vec{j} - B_{i,k}) = j_s G(\vec{j}) \tag{28}$$

Such mixture of service-classes does not destroy the accuracy of the model — Eq. (28). The CBP calculation can be done via Eq. (14) while the BBP calculation can be done via Eq. (15) for the service-classes of finite population and via Eq. (20) for the service-classes of infinite population.

5. Numerical Examples — Evaluation

Firstly, we present analytical CBP and BBP results of the generalized f-ON-OFF model. Because of the PFS of the model and the accurate derivation of the CBP formulas, the simulation CBP results are very close to the analytical CBP results and therefore are not presented. For the BBP we provide simulation results, to evaluate the accuracy of the proposed BBP formulas. Simulation was done by using SIMSCRIPT II.5 and the simulations results are mean values of 10 runs with 95% confidence interval. Secondly, we compare the CBP results of the proposed f-ON-OFF model with those of the inf-ON-OFF model, to show the necessity and the consistency of the proposed model.

In the first example, we consider two service-classes which require $b_1 = 12$ and $b_2 = 2$ b.u. per call, respectively, and a link with $C = C^* = 46$ b.u. Calls of the 1st service-class arrive according to a quasi-random process, while calls of the 2nd service-class arrive according to a Poisson process ($N_2 = \infty$). The number of sources of the 1st service-class is $N_1 = 5$. The traffic description parameters are:

- 1st service-class: $\upsilon_1 = 0.02$, $\mu_{11}^{-1} = 0.5$, $\mu_{21}^{-1} = 0.6$, $\sigma_1 = 0.9$, $p_{11,fin} = 0.1$, $p_{21,fin} = 0.108$
- 2nd service-class: $\lambda_2 = 0.1$, $\mu_{12}^{-1} = 1.0$, $\mu_{22}^{-1} = 1.0$, $\sigma_2 = 0.9$, $p_{12} = 1.0$, $p_{22} = 0.9$

The equivalent system used for the CBP calculation is the following: $b_1 = 25$, $b_2 = 4$ and $C = C^* = 95$. When $C = C^*$ no BB occurs. We increase the fictitious capacity of the initial system from $C^* = 46$ to $C^* = 51$, in order for BB to occur; then the equivalent system used for the CBP and BBP calculation is: $b_1 = 29$, $b_2 = 5$, $C = 115$ and $C^* = 125$. We further increase C^* to 56 b.u.; then, the equivalent system is: $b_1 = 29$, $b_2 = 5$, $C = 115$ and $C^* = 140$. The determination of the value of C^* makes a trade-off between BBP and CBP.

Figures 4 and 5 show the analytical CBP results of the 1st and 2nd service-class, respectively, for: a) $C = C^* = 46$, b) $C = 46$, $C^* = 51$ and c) $C = 46$, $C^* = 56$, whereas Fig. 6 shows the analytical and simulation BBP results for both service-classes for $C = 46$, $C^* = 51$ and for $C = 46$, $C^* = 56$. At each point in the horizontal axis entitled "arrival rate" υ_1 is constant, while λ_2 increases in steps of 0.1, i.e. point 1 is $(\upsilon_1, \lambda_2) = (0.02, 0.1)$, point 2: $(\upsilon_1, \lambda_2) = (0.02, 0.2)$, ..., point 7: $(\upsilon_1, \lambda_2) = (0.02, 0.7)$.

As Figs. 4 and 5 show, there is a significant decrease in the CBP of both service-classes, when we increase the fictitious capacity C^* from 46 to 51, or to 56. This CBP decrease was anticipated, because, when C^* increases, more calls can pass to state OFF, releasing bandwidth (in state ON) for new calls. However, the increase of C^* results in the BB increase, as Fig. 6 shows.

In the second example, we consider the system of the first example where both service-classes follow: a) the f-ON-OFF model and b) the inf-ON-OFF. In the case of the f-ON-OFF model, the traffic description parameters of the 2nd service-class are:

- $N_2 = 32$, $\upsilon_2 = 0.003125$, $\mu_{11}^{-1} = 0.5$, $\mu_{21}^{-1} = 0.6$, $\sigma_1 = 0.9$, $p_{11,fin} = 0.03125$, $p_{21,fin} = 0.028125$.

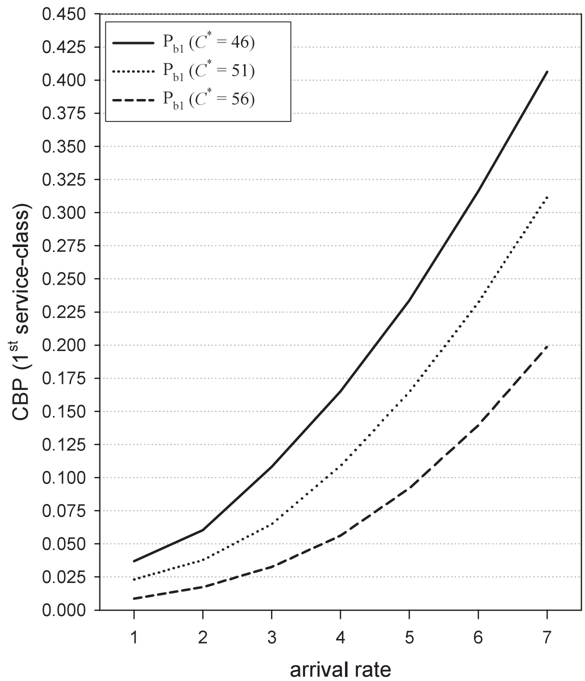


Fig. 4 CBP for the 1st serive-class when a) $C^* = 46$, b) $C^* = 51$, c) $C^* = 56$.

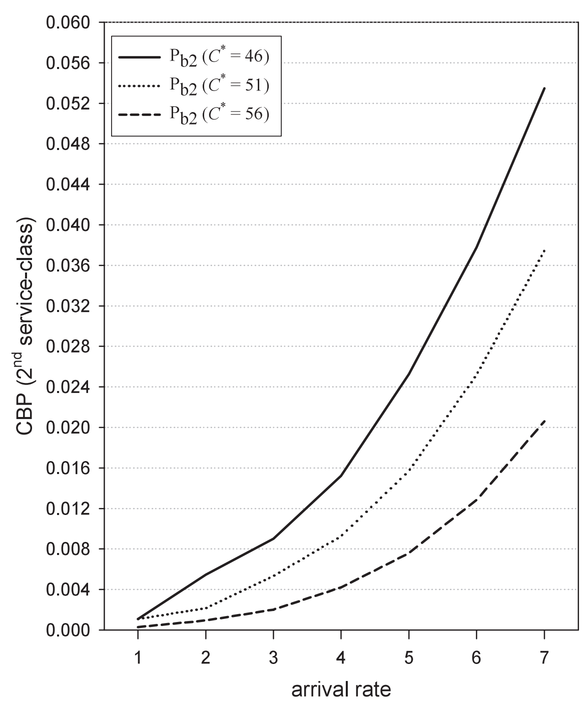


Fig. 5 CBP for the 2nd serive-class when a) $C^* = 46$, b) $C^* = 51$, c) $C^* = 56$.

In the case of the inf-ON-OFF model, the traffic description parameters of the 1st service-class are:

- $\lambda_1 = 0.1$, $\mu_{11}^{-1} = 0.5$, $\mu_{21}^{-1} = 0.6$, $\sigma_1 = 0.9$, $p_{11} = 0.5$, $p_{21} = 0.54$.

Figure 7 shows the analytical CBP results of the 1st

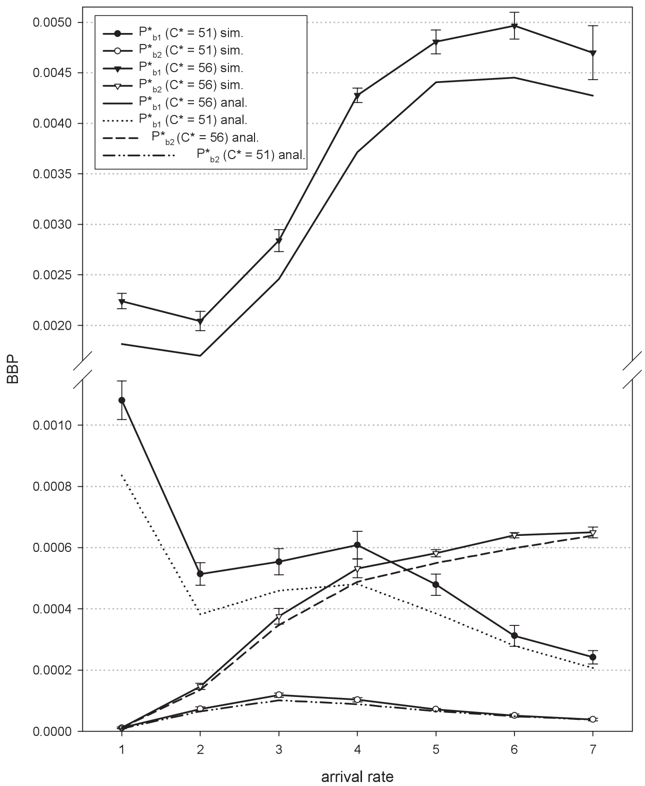


Fig. 6 CBP for both serive-class when a) $C^* = 51$ and b) $C^* = 56$.

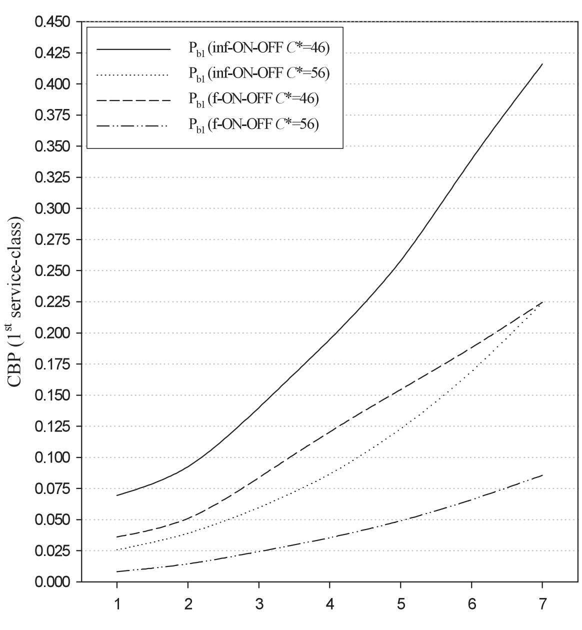


Fig. 7 CBP for the 1st serive-class when a) $C^* = 46$ and b) $C^* = 56$.

service-class, for the f-ON-OFF and the inf-ON-OFF model comparatively, when: a) $C = C^* = 46$ and b) $C = 46$, $C^* = 56$. Concentrating on the inf-ON-OFF model the arrival rate λ_1 is constant, while λ_2 increases in steps of 0.1, i.e. in Fig. 7 point 1 is $(\lambda_1, \lambda_2) = (0.1, 0.1)$, point 2: $(\lambda_1, \lambda_2) = (0.1, 0.2)$, ..., point 7: $(\lambda_1, \lambda_2) = (0.1, 0.7)$. Figure 8 shows the corresponding analytical CBP results of the 2nd service-class. As both Figs. 7 and 8 show, there is a significant difference between the two models.

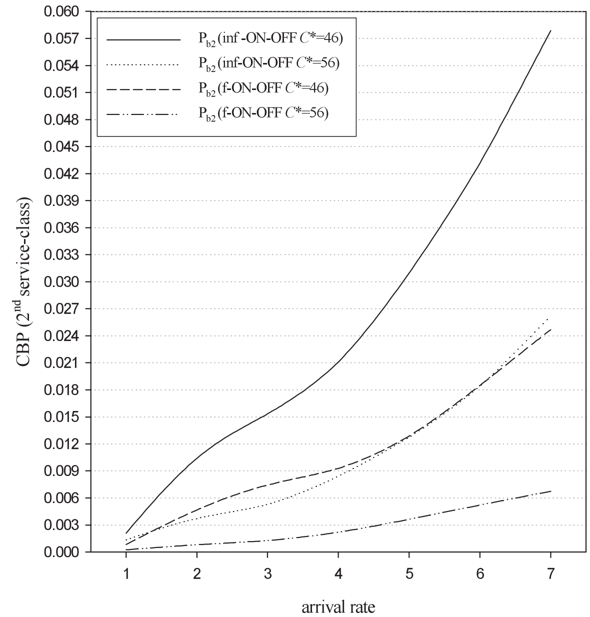


Fig. 8 CBP for the 2nd serive-class when a) $C^* = 46$ and b) $C^* = 56$.

6. Conclusion

We propose an ON-OFF traffic model for a single link, which accommodates service-classes of quasi-random input, under the CS policy. The proposed f-ON-OFF model has a PFS, which helps us derive an accurate and recursive formula for the CBP calculation. The BBP calculation is based on an approximate but robust formula. We show that the f-ON-OFF model is highly related to other call-level loss models and discuss possible applications. Furthermore, we generalize the f-ON-OFF model so as to include a mixture of service-classes with finite and infinite population of sources. We give numerical CBP and BBP results of the generalized f-ON-OFF model, in order to show the model's behavior, while providing simulation BBP results for evaluation. Furthermore, we compare the proposed f-ON-OFF model with the inf-ON-OFF model in terms of CBP, in order to show the consistency and the necessity of the proposed model.

Acknowledgment

Work supported by the research program Caratheodory of the Research Committee of the University of Patras, Greece. The authors wish to thank Prof. M. Mehmet Ali, (Concordia University, Canada) for his valuable comments on the ON-OFF model, Prof. G. Kokkinakis (University of Patras, Greece) for his permanent help and support, and Prof. T. Mathiopoulos (National Observatory of Athens, Greece) for our discussions.

References

[1] D. Awduche and Y. Rekhter, "Multiprotocol lambda switching: Combining MPLS traffic engineering control with optical cross-connects," IEEE Commun. Mag., vol.39, no.3, pp.111–116, March 2001.

[2] M.J. Morrow, M. Tatipamula, and A. Farrel, "GMPLS the promise of the next-generation optical control plane," IEEE Commun. Mag., vol.43, no.7, pp.26–27, July 2005.

[3] H. Holma, M. Kristensson, J. Salonen, and A. Toskala, UMTS Services and Applications, Harri Holma and Antti Toskala, WCDMA for UMTS, 3rd ed., Wiley, W. Sussex, England, 2004.

[4] V. Stanisic and M. Devetsikiotis, "An analysis of bandwidth allocation strategies in multiservice networks," Proc. IEEE ICC 2005, pp.1529–1533, Seoul, 2005.

[5] J.W. Roberts, "Traffic theory and the Internet," IEEE Commun. Mag., vol.39, no.1, pp.94–99, Jan. 2001.

[6] Keith W. Ross, Multiservice loss models for broadband telecommunication networks, 3rd ed., Springer, Berlin, Germany, 1995.

[7] C. Chigan, R. Nagarajan, Z. Dziong, and T. Robertazzi, "On the capacitated loss network with heterogeneous traffic and contiguous resource allocation constraints," Proc. Advanced Simulation Technologies Conference, Seattle, Washington, USA, April 2001.

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

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

[10] S. Racz, B.P. Gero, and G. Fodor, "Flow level performance analysis of a multi-service system supporting elastic and adaptive services," Performance Evaluation, vol.49, pp.451–469, 2002.

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

[12] H. Akimaru and K. Kawashima, Teletraffic — Theory and Applications, Springer Verlag, Berlin, Germany, 1993.

[13] R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," IEEE J. Sel. Areas Commun., vol.9, no.7, pp.968–981, 1991.

[14] F. Hubner and M. Ritter, "Blocking in multi-service broadband systems with CBR and VBR input traffic," Proc. 7th ITG/GI, pp.212–225, Aachen, Sept. 1993.

[15] T.V.J. Ganesh Babu, T.L. Ngoc, and J.F. Hayes, "A unified approach for evaluating call blocking and burst blocking in high speed networks," Proc. Canadian Conference on Electrical and Computer Engineering (CCECE95), pp.846–849, Sept. 1995.

[16] M. Mehmet Ali, "Call-burst blocking and call admission control in a broadband network with bursty sources," Performance Evaluation, vol.38, pp.1–19, 1999.

[17] J.M. Aein, "A multi-user-class, blocked-calls-cleared demand access model," IEEE Trans. Commun., vol.26, no.3, pp.378–385, 1978.

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

[19] I. Moscholios, P. Nikolaropoulos, and M. Logothetis, "Call level blocking of ON-OFF traffic sources with retrials under the complete sharing policy," Proc. 18th International Teletraffic Congress (ITC), pp.811–820, Berlin, Aug.-Sept. 2003.

[20] L.E.N. Delbrouck, "On the steady state distribution in a service facility with different peakedness factors and capacity requirements," IEEE Trans. Commun., vol.COM-31, no.11, pp.1209–1211, 1983.

[21] G.A. Awater and H.A.B. van de Vlag, "Exact computation of time and call blocking probabilities in large, multi-traffic, multi-resource loss systems," Performance Evaluation, vol.25, no.1, pp.41–58, March 1996.

[22] J.W. Roberts, "Teletraffic models for the telecom 1 integrated services network," Proc. ITC-10, 1983.

[23] J.S. Turner, "Managing bandwidth in ATM networks with bursty traffic," IEEE Network, pp.52–58, 1992.

[24] 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, June 1997.

[25] S.P. Chung and K.W. Ross, "Reduced load approximations for multirate loss networks," IEEE Trans. Commun., vol.41, no.8, pp.1222–1231, Aug. 1993.

[26] A. Greenberg and R. Srikant, "Computational techniques for accurate performance evaluation of multirate, multihop communication networks," IEEE/ACM Trans. Netw., vol.5, no.2, pp.266–277, April 1997.

[27] M. Logothetis and S. Shioda, "Medium-term centralized virtual path bandwidth control based on traffic measurements," IEEE Trans. Commun., vol.43, no.10, pp.2630–2640, 1995.

[28] M. Logothetis and G. Kokkinakis, "Path bandwidth management for large scale telecom networks," IEICE Trans. Commun., vol.E83-B, no.9, pp.2087–2099, Sept. 2000.

[29] H. Shengye, Y. Wu, F. Suili, and S. Hui, "Coordination-based optimisation of path bandwidth allocation for large-scale telecommunication networks," Comput. Commun., vol.27, no.1, pp.70–80, Jan. 2004.

[30] P. Fazekas, S. Imre, and M. Telek, "Modelling and analysis of broadband cellular networks with multimedia connections," Telecommunication Systems, vol.19, pp.263–288, 2002.

[31] D. Staehle and A. Mader, "An analytic approximation of the uplink capacity in a UMTS network with heterogeneous traffic," Proc. 18th International Teletraffic Congress (ITC), pp.81–90, Berlin, Aug.-Sept. 2003.

[32] A. Mader and D. Staehle, "Analytic modeling of the WCDMA downlink capacity in multi-service environments," ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems, pp.217–226, Aug.-Sept. 2004.

[33] M. Glabowski, M. Stasiak, A. Wisniewski, and P. Zwierzykowski, "Uplink blocking probability calculation for cellular systems with WCDMA radio interference, finite source population and differently loaded neighboring cells," Proc. 2005 Asia Pacific conference on Communications, pp.138–142, Perth, Australia, Oct. 2005.

[34] I. Moscholios, M. Logothetis, and M. Koukias, "An on-off multirate loss model with a mixture of service-classes of finite and infinite number of sources," Proc. IEEE International Conference on Communications, ICC 2005, pp.357–362, Seoul, May 2005.

Appendix: Description of the Heuristic Algorithm for the Determination of an Equivalent Stochastic System

Input: K, C, C^*, $\vec{b}=(b_1, b_2, \ldots, b_k, \ldots, b_K), N_k, (k = 1, \ldots, K)$

Initialization: $n^1_{k,\max} = \left\lfloor \frac{C}{b_k} \right\rfloor$, $n^2_{k,\max} = \left\lfloor \frac{C^*}{b_k} \right\rfloor$

Procedure A

Step 1: For each service-class k (k=1, …,K), let $n^1_k = 0, \ldots, n^1_{k,\max}$ and $n^2_k = 0 , \ldots, n^2_{k,\max}$

Step 2: Take a permutation of n^1_k, n^2_k for all service-classes (k=1,…,K) in order to form a state vector $\vec{n} = (n^1, n^2)=(n^1_1, \ldots, n^1_k, \ldots, n^1_K, \ldots, n^2_1, \ldots, n^2_k, \ldots, n^2_K)$ so that $n^1_k + n^2_k \leq N_k$ for each service-class k.

Step 3: Compute $j_1=n^1\vec{b}$ and $j_2 = n^2\vec{b}$. If $j_1 \leq C$ and

$j_1 + j_2 \leq C^*$, the current state vector $\vec{n} = (n^1, n^2)$ is a valid state of the initial system.

Step 4: Repeat *Steps* 2 and 3 for all permutations in order to form the entire state space.

End of Procedure A

At this point we have formed the state space of the initial stochastic system together with j_1, j_2. If the value of the pair $\vec{j} = (j_1, j_2)$ appears more than once in the state space then proceed to *Procedure B*. Otherwise there is no need to define an equivalent system.

Procedure B

Step 1: Increase the elements of vector $\vec{b}$ by one b.u. (new $\vec{b}$) and compute the pair $\vec{j} = (j_1, j_2)$ for every state $\vec{n} = (n^1, n^2)$ of the initial stochastic system. If the resultant pair $\vec{j} = (j_1, j_2)$ has a unique value in the entire state space, then the new $\vec{b}$ is a candidate bandwidth vector of the equivalent stochastic system.

Step 2: The capacity of the equivalent stochastic system is determined by new $C = max(j_1)$, new $C^* = max(j_2)$.

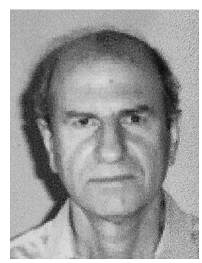
Step 3: Based on the new $\vec{b}$ and the new C, C^*, validate that the state-space of the new system coincides with the initial state-space; the validation is performed by *Procedure A*. If the new system has indeed the same state-space with the initial one, then it is an equivalent system with parameters: new $\vec{b}$, new C, new C^*. Go to *Output*. Otherwise go to *Step* 1 of *Procedure B*.

End of Procedure B.

Output: new C, new C^*, new $\vec{b}=(b_1, b_2, \ldots, b_k, \ldots, b_K)$, $(k = 1, \ldots, K)$

Ioannis D. Moscholios was born in Athens, Greece, in 1976. He received a Dipl.Eng. degree in Electrical and Computer Engineering from the University of Patras, Patras, Greece, in 1999, an M.Sc. degree in Spacecraft Technology and Satellite Communications from the University College London (UCL), London, UK, in 2000 and his PhD in Electrical and Computer Engineering from the University of Patras, Greece, in 2005. Currently, he is a Research Associate at the Wire Communications Laboratory, Department of Electrical and Computer Engineering, University of Patras, Greece. Since he was an undergraduate student, he was involved in national research and development projects. His research includes simulation and performance analysis of communication networks, while his main interest is on teletraffic engineering. He is a member of the Technical Chamber of Greece (TEE).

Michael D. Logothetis was born in Stenies, Andros, Greece, in 1959. He received his Dipl.Eng. degree and Ph.D. 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 EC research programmes, dealing with telecommunication networks, as well as with natural language processing. From 1991 to 1992 he was Research Associate in NTTs Telecommunication Networks Laboratories, Tokyo, Japan. Afterwards, he was a Lecturer in the Department of Electrical and Computer Engineering of the University of Patras, and since 2003 he is an Associate Professor. His research interests include teletraffic engineering, traffic/network control, simulation and performance optimization of telecommunications networks. He organised the international conference CSNDSP 2006. He is a member of the IEEE, FITCE and the Technical Chamber of Greece (TEE).



Michael N. Koukias was born in Chios, Greece, in 1951. He received his Diploma in Electrical Engineering from the University of Patras in 1975, the M.Sc. from the University of Manchester (UMIST) in 1980 and the Ph.D. degree from the University of Patras in 1986. Dr. Koukias is an Assistant Professor in the Department of Electrical and Computer Engineering of the University of Patras, where he lectures Telematics, Computer Programming, Information Systems and Communication Systems. Since 1981 he has been involved in various research and development projects funded from industry, national research sources, EEC etc. His current interests include telematics applications, broadband communications, multimedia communications and web applications. He is a member of FITCE and the Technical Chamber of Greece (TEE).