Article:

Moscholios, Ioannis, Vasilakis, Vasileios orcid.org/0000-0003-4902-8226, Logothetis, Michael et al. (1 more author) (2017) State-Dependent Bandwidth Sharing Policies for Wireless Multirate Loss Networks. IEEE Transactions on Wireless Communications. 7942058. pp. 5481-5497. ISSN 1536-1276

https://doi.org/10.1109/TWC.2017.2712153

State-Dependent Bandwidth Sharing Policies for Wireless Multirate Loss Networks

Ioannis D. Moscholios, Vassilios G. Vassilakis, Michael D. Logothetis, *Senior Member, IEEE,* and Anthony C. Boucouvalas, *Fellow, IEEE*

***Abstract*—We consider a reference cell of fixed capacity in a wireless cellular network while concentrating on next-generation network architectures. The cell accommodates new and handover calls from different service-classes. Arriving calls follow a random or quasi-random process and compete for service in the cell under two bandwidth sharing policies: i) a probabilistic threshold (PrTH) policy, or ii) the multiple fractional channel reservation (MFCR) policy. In the PrTH policy, if the number of in-service calls (new or handover) of a service-class exceeds a threshold (different between new and handover calls), then an arriving call of the same service-class is accepted in the cell with a predefined state-dependent probability. In the MFCR policy, a real number of channels is reserved to benefit calls of certain service-classes; thus a service priority is introduced. The cell is modeled as a multirate loss system. Under the PrTH policy, call-level performance measures are determined via accurate convolution algorithms, while under the MFCR policy, via approximate but efficient models. Furthermore, we discuss the applicability of the proposed models in 4G/5G networks. The accuracy of the proposed models is verified through simulation. Comparison against other models reveals the necessity of the new models and policies.**

***Index Terms*—Call admission, threshold, reservation, convolution, wireless networks.**

I. Introduction

THE fast proliferation of the mobile Internet, the widespread use of social media and an increasing popularity of bandwidth-hungry applications, such as mobile cloud computing/storage and mobile video streaming, necessitate huge bandwidth capacity in network links. Fortunately, broadband wireless networks become economically viable in supporting nowadays traffic, but the need for quality-of-service (QoS) assurance is even more important than before [1]. Hence, QoS mechanisms are essential to give access to the necessary bandwidth needed by the services of the mobile users (MUs) and at the same time to ensure fairness among different "competing" mobile services/applications. On the other hand, the incorporation of the emerging technologies of software-defined networking (SDN) and network function virtualization (NFV) in next-generation wireless networks [2], provides new opportunities for fairer QoS assignment among service-classes. Despite this fact, there is lack of QoS mechanisms with efficient and fast QoS assessment in the complicated traffic environment of contemporary wireless networks.

Considering call-level traffic in a single cell of a wireless cellular network which accommodates different service-classes with different QoS requirements, such a mechanism is a bandwidth sharing policy, since it affects call-level performance measures, like call blocking probabilities (CBP). The QoS assessment of service systems under a bandwidth sharing policy is accomplished through teletraffic loss or queueing models which may incorporate the notion of equivalent bandwidth [3] and can be described via efficient, recursive, or convolutional formulas. The latter reduce computational complexity and therefore can be invoked in network planning and dimensioning procedures. In this paper, we aim at proposing teletraffic models which are suitable not only for conventional wireless networks but also for 4G and 5G networks.

The simplest bandwidth sharing policy is the complete sharing (CS) policy, where a new call is accepted in a cell if the call's bandwidth is available [3]. The CS policy leads to recursive or convolutional formulas for the CBP calculation but: i) it is unfair to calls with higher bandwidth requirements since it leads to higher CBP [4] and ii) it does not provide different treatment to handover calls, i.e., calls transferred from one cell to another while they are still in progress. These reasons motivate research on policies, such as the bandwidth reservation (BR) policy (also known as guard channel policy), the multiple fractional channel reservation (MFCR) policy, and the threshold (TH) policy (e.g., [4]–[18]). The BR policy introduces a service priority to benefit high-speed calls and can achieve CBP equalization among calls of different service-classes at the expense of substantially increasing the CBP of lower-speed calls. In addition, the system analysis under the BR policy leads to non-reversible continuous time Markov chains and consequently to approximate formulas for the CBP calculation (e.g., [8], [9], [11], [16], [17]). The MFCR policy generalizes the BR policy for a refined QoS assessment by allowing the reservation of real (not integer) number of channels (e.g., [4]- [6]). In this paper, a channel does not refer to an actual physical or logical communication channel but to a bandwidth (data rate) unit.

We concentrate on the MFCR policy and the TH policy. The latter: i) is applicable in wired (e.g., [7], [16], [17]), wireless (e.g., [18]–[20]) and satellite networks (e.g., [21], [22]), ii) is attractive in access tree networks and iii) does not destroy reversibility in continuous time Markov chains, unlike the BR or MFCR policies. In the TH policy, the number of in-service

Manuscript received October 31, 2016; revised March 18, 2017; accepted May 24, 2017.

I. D. Moscholios and A. C. Boucouvalas are with the Dept. of Informatics & Telecommunications, University of Peloponnese, 221 00 Tripolis, Greece.

V. G. Vassilakis is with the Dept. of Computer Science, University of York, YO10 5GH York, UK.

M. D. Logothetis is with the Dept. of Electrical & Computer Engineering, University of Patras, 265 04 Patras, Greece.

calls of a service-class must not exceed a threshold (dedicated to the service-class), after the acceptance of a new call of this service-class. Otherwise, call blocking occurs even if available bandwidth exists in the system. For a refined QoS assessment, we proposed a Probabilistic TH policy (PrTH), where call acceptance is permitted above a threshold, with a probability. This probability depends on the service-class, the type of call (new or handover) and the system state [18].

In this paper, we consider a reference cell of fixed capacity that accommodates different service-classes, under the PrTH or MFCR policies. New calls arrive in the cell according to a random (Poisson) or quasi-random process; the latter is realistic in the case of small cells, while the former in the case of large cells. Handover calls may follow a quasi-random process, because it is more realistic to assume that they are generated by a finite number of MUs (especially in small size cells which serve a finite number of MUs), and therefore their arrival process is smoother than the Poisson process (e.g., [23]–[25]). Assuming that the bandwidth of in-service calls cannot be altered, we model the cell as a multirate loss system. To distinguish between the two arrival processes, we propose: a) the PrTH quasi-random (PrTH-Q) loss model where both new and handover calls follow a quasi-random process, b) the PrTH random/quasi-random (PrTH-RQ) loss model, where new calls follow a random process while handover calls follow a quasi-random process. The case of a random process for both types of calls (PrTH-R model) has been proposed in [18]. c) the MFCR-Q model whereby new or handover calls follow a quasi-random process. d) the MFCR-RQ model whereby new calls follow a random process and handover calls follow a quasi-random process. Note that in [4]- [6] only the case of random process for both new and handover calls is considered (MFCR-R model).

The proposed PrTH loss models can be described by continuous time Markov chains which are reversible and have a product form solution (PFS) for the steady state distribution. Based on the PFS, we determine accurately time congestion (TC) and call congestion (CC) probabilities as well as the system's utilization via convolution algorithms [3]. TC probabilities refer to the proportion of time the system is congested, while CC probabilities refer to CBP. Assuming a Poisson arrival process for all calls then TC and CC probabilities coincide (PASTA property [3]). The proposed MFCR-Q and MFCR-RQ models do not have a PFS. However, we prove approximate but recursive formulas for the various performance measures. The accuracy of the proposed formulas is verified through simulation and found to be highly satisfactory.

The proposed models fit well to: a) orthogonal frequency division multiplexing (OFDM) wireless networks and b) heterogeneous radio access networks (RAN) based on the emerging cloud RAN (C-RAN) architecture. To show it, we propose a framework for the applicability of the proposed models in: a) 4G OFDM wireless networks and b) 5G wireless networks, based on the technologies of SDN and NFV.

The remainder of this paper is as follows: In Section II, we present the related work for teletraffic loss models that consider the co-existence of new and handover calls. In Section III, we present the proposed PrTH-Q model, show the PFS and provide a convolution algorithm for the calculation of the various performance measures. In Section IV, we present the proposed PrTH-RQ model. In Sections V and VI, we present the proposed MFCR-Q and MFCR-RQ models, respectively, and provide recursive formulas for the calculation of the system's occupancy distribution as well as various performance measures. In Sections VII, VIII we discuss the applicability of our models in 4G OFDM and 5G networks, respectively. In Section IX, we present analytical and simulation TC probabilities results both for the proposed models and other existing models, for evaluation. In Section X, we present the conclusion and future work.

II. Related Work

In the literature, there are many teletraffic loss models that describe the co-existence of new and handover calls in a cell (see e.g., [26]–[32], [20], [35]–[39]). Their classification can be done in various ways, e.g., in terms of the bandwidth sharing policy, the call arrival process, the amount of bandwidth requested by calls etc. To briefly describe the aforementioned models, we classify them in single-rate ([26]–[28]) and multirate ([29]–[32], [20], [35]–[39]) models. In [26]–[28], new or handover calls request only one channel in order to be connected in the system. This is a drawback when studying contemporary networks which accommodate calls of different bandwidth requirements. In [26], new or handover calls follow a Bernoulli-Poisson-Pascal process and compete for the available bandwidth under the CS policy. In [27], all calls follow a Poisson process and compete for the available bandwidth under a variation of the BR policy, named fractional guard channel policy. The models of [26], [27] do have a PFS but no efficient algorithm is proposed for the CBP calculation. In [28], blocked calls retry to be connected in the system under the BR policy. The model does not have a PFS, due to the existence of retrials, and therefore the CBP calculation is based on solving the corresponding global balance (GB) equations, an extremely complex and time consuming task for large systems. In [29], [30], [32], multirate loss systems are examined under the CS policy ([29], [31], [32]) and the processor sharing discipline [30]. This discipline is used when the bandwidth allocated to in-service calls is elastic (not fixed) during their lifetime in the system [40]–[42]. In [29]–[32], handover traffic is considered as one component of the offered traffic-load. This consideration is acceptable if the amount of handover traffic is insignificant compared to new traffic (e.g., in very large-sized cells). In this paper, we explicitly distinguish the traffic offered to the cell by new and handover calls. A basic characteristic of [29]–[32] is that the CBP calculation is based on modifications of the classical Kaufman-Roberts (K-R) recursion [33], [34]. The latter is used for the accurate and efficient calculation of the link occupancy distribution in a multirate loss system which accommodates Poisson arriving calls under the CS policy. Other teletraffic models that explicitly distinguish new and handover traffic and can be described by extensions of the K-R formula are [20], [35]–[39]. In the PrTH policy, the calculation of the system's occupancy distribution is based on a convolution algorithm. This is because the probabilities in the PrTH policy depend on the number of in-service calls in a cell, a micro-state information that cannot be retrieved by the macro-state K-R formula. On the other hand, the system's occupancy distribution in the proposed MFCR-Q and MFCR-RQ models is based on extensions of the K-R formula.

A number of recent works investigate the problem of efficient radio resource management (RRM) and packet scheduling in wireless networks using various virtualization and resource optimization schemes [43]–[46]. In [43], a novel method for dynamic optimization of long term evolution (LTE) handover parameters is proposed. This approach considers the movement of users and adjusts the handover range to optimize the mobility load balancing function. Other approaches, rather than relying on simulations, tackle this problem by modelling the system as a multi-user multi-server discrete-time queuing system [44]. In [45], the problem of resource sharing among multiple virtual wireless networks is investigated. An SDN-based solution is proposed for joint optimization of network bandwidth and power. It is demonstrated that the incorporation of SDN concepts in wireless networks can greatly facilitate the RRM and allocation in virtualized wireless environments. In [46], the problem of cross-cell coordinated radio resource allocation in next-generation cellular networks is studied. In particular, this work considers a heterogeneous C-RAN and aims at improving the network-wide resource management when multiple radio technologies coexist.

III. The Proposed PrTH-Q Model

Consider a cell of capacity C channels that accommodates calls of K service-classes under the PrTH policy, while the input traffic is quasi-random (PrTH-Q model). To separate new from handover calls of the same service-class, we assume that the system accommodates $2K$ service-classes. A service-class k call is new if $1 \leq k \leq K$ and handover if $K+1 \leq k \leq 2K$. A new service-class k call requires the same number of channels, b_k, with a handover service-class $K+k$ call. Service-class k calls ($k = 1, ..., 2K$) come from a finite source population N_k. The effective arrival rate of service-class k calls is $\lambda_{k,fin} = (N_k - n_k)v_k$ where n_k is the number of in-service calls and v_k is the arrival rate per idle source. The offered traffic-load per idle service-class k source is $\alpha_{k,fin} = v_k/\mu_k$ (in erl) where μ_k^{-1} is the mean service time (generally distributed) of an accepted service-class k call. This arrival process is named quasi-random [3] (if $N_k \to \infty$ for $k = 1, ..., 2K$ and the total offered traffic is constant, a Poisson process arises).

To describe the call admission mechanism, consider a service-class k call that requires b_k channels. If they are not available in the cell, then the call is blocked and lost; otherwise:

a) If the number n_k of in-service calls of service-class k ($k = 1, \ldots, 2K$) plus the new or handover call, does not exceed a threshold n_k^*, i.e., $n_k+1 \leq n_k^*$, then the call is accepted in the system.

b) If $n_k+1 > n_k^*$, the call is accepted with probability $p_k(n_k)$ or blocked with probability $1 - p_k(n_k)$. The set of $p_k(n_k)$ constitutes the vector

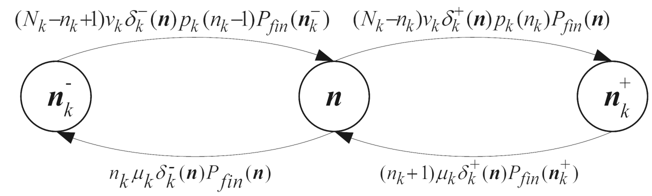

Fig. 1: State transition diagram of the PrTH-Q model for service-class k.

$$\begin{aligned}\boldsymbol{p_k} =& (p_k(0), p_k(1), \ldots, p_k(n_k^*), \ldots, p_k(\lfloor C/b_k \rfloor - 1), \\ & p_k(\lfloor C/b_k \rfloor)),\end{aligned} \tag{1}$$

where $\lfloor C/b_k \rfloor$ is the maximum number of service-class k calls that the system can service.

In (1), we assume that:

1) $p_k(0) = ... = p_k(n_k^* - 1) = 1$, i.e., a service-class k call is accepted if n_k^* is not exceeded.

2) the probabilities $p_k(n_k^*), \ldots, p_k(\lfloor C/b_k \rfloor - 1)$ may be different for new or handover calls of the same service-class k. In the TH policy these probabilities are all zero. In the PrTH policy, they can be set either all positive, or zero after a number greater than n_k^*. To prioritize handover calls over new calls of a service-class k, we can choose higher values for the thresholds of handover calls or set the corresponding probabilities of handover calls to higher values.

3) $p_k(\lfloor C/b_k \rfloor) = 0$ obviously, due to lack of available bandwidth.

Let the steady state vector be $\boldsymbol{n} = (n_1, ..., n_k, ..., n_{2K})$ and $\boldsymbol{n}_k^- = (n_1, ..., n_k - 1, ..., n_{2K})$, $\boldsymbol{n}_k^+ = (n_1, ..., n_k + 1, ..., n_{2K})$ and $P_{fin}(\boldsymbol{n}), P_{fin}(\boldsymbol{n}_k^-), P_{fin}(\boldsymbol{n}_k^+)$ are the probability distributions of states $\boldsymbol{n}, \boldsymbol{n}_k^-, \boldsymbol{n}_k^+$, respectively.

Based on the state transition diagram (Fig. 1) of the proposed PrTH-Q model, the GB equation for state $\boldsymbol{n}$, expressed as *rate into state* $\boldsymbol{n}$ = *rate out of state* $\boldsymbol{n}$, is

$$\begin{aligned}&\sum_{k=1}^{2K}\Big[(N_k - n_k + 1)v_k\delta_k^-(\boldsymbol{n})p_k(n_k - 1)P_{fin}(\boldsymbol{n}_k^-) + \\ &(n_k+1)\mu_k\delta_k^+(\boldsymbol{n})P_{fin}(\boldsymbol{n}_k^+)\Big] = \\ &\sum_{k=1}^{2K}\Big[(N_k - n_k)v_k\delta_k^+(\boldsymbol{n})p_k(n_k)P_{fin}(\boldsymbol{n}) + n_k\mu_k\delta_k^-(\boldsymbol{n})P_{fin}(\boldsymbol{n})\Big],\end{aligned} \tag{2}$$

where $\delta_k^+(\boldsymbol{n}) = \begin{cases} 1, & \text{if } \boldsymbol{n}_k^+ \in \boldsymbol{\Omega} \\ 0, & \text{otherwise} \end{cases}$, $\delta_k^-(\boldsymbol{n}) = \begin{cases} 1, & \text{if } \boldsymbol{n}_k^- \in \boldsymbol{\Omega} \\ 0, & \text{otherwise} \end{cases}$, $\boldsymbol{\Omega}$ is the state space of the system, $\boldsymbol{\Omega} = \{\boldsymbol{n} : 0 \leq \boldsymbol{nb} \leq C, k = 1, ..., 2K\}$ and $\boldsymbol{nb} = \sum_{k=1}^{2K} n_k b_k$, $\boldsymbol{b} = (b_1, ..., b_{2K})^T$.

Based on Fig. 1, the Markov chain of the PrTH-Q model is reversible due to Kolmogorov's criterion [47]: the circulation flow among four adjacent states equals zero: Flow clockwise = Flow counter-clockwise. Because of this, local balance (LB) exists between adjacent states and the following LB equations are extracted as *(rate up = rate down)*, for $k = 1, ..., 2K$ and $\boldsymbol{n} \in \boldsymbol{\Omega}$

$$(N_k - n_k + 1) v_k \delta_k^-(\boldsymbol{n}) p_k(n_k - 1) P_{fin}(\boldsymbol{n}_k^-) = n_k \mu_k \delta_k^-(\boldsymbol{n}) P_{fin}(\boldsymbol{n}), \quad (3)$$

$$(N_k - n_k) v_k \delta_k^+(\boldsymbol{n}) p_k(n_k) P_{fin}(\boldsymbol{n}) = (n_k + 1) \mu_k \delta_k^+(\boldsymbol{n}) P_{fin}(\boldsymbol{n}_k^+). \quad (4)$$

The system of LB equations is satisfied by the following PFS, for $k = 1, ..., 2K$ and $\boldsymbol{n} \in \boldsymbol{\Omega}$

$$P_{fin}(\boldsymbol{n}) = G^{-1} \left(\prod_{k=1}^{2K} \binom{N_k}{n_k} \prod_{x=n_k^*}^{n_k - 1} p_k(x) \alpha_{k,fin}^{n_k} \right), \quad (5)$$

where G is the normalization constant determined by

$$G \equiv G(\boldsymbol{\Omega}) = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}} \left(\prod_{k=1}^{2K} \binom{N_k}{n_k} \prod_{x=n_k^*}^{n_k - 1} p_k(x) \alpha_{k,fin}^{n_k} \right). \quad (6)$$

In a system with quasi-random input, CBP are distinguished to TC and CC probabilities. To calculate the TC probabilities of service-class k calls, B_k, let $\boldsymbol{\Omega}_k = \{\boldsymbol{n} : 0 \le \boldsymbol{nb} \le C - b_k, k = 1, ..., 2K\}$ be the state space which denotes the set of states for which a service-class k call will be definitely accepted or accepted with a state-dependent probability in the system. Thus

$$B_k = 1 - G_k / G, \quad (7)$$

where $G_k = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_k} p_k(n_k) P_{fin}(\boldsymbol{n})$.

CC probabilities, i.e., CBP seen by an arriving call, are calculated via (7) by considering $N_k - 1$ traffic sources. For an efficient calculation of TC or CC probabilities we can exploit the PFS of the PrTH-Q model, and use the following 3-step convolution algorithm:

Define j as the occupied system's bandwidth, $j = 0, 1, ..., C$.
Step 1) Determine the occupancy distribution $q_k(j)$ of each service-class k ($k = 1, ..., 2K$), assuming that only service-class k exists in the system

$$q_k(j) = \begin{cases} q_k(0) \binom{N_k}{i} \alpha_{k,fin}^i, & \text{for } 1 \le i \le n_k^* \text{ and } j = i b_k \\ q_k(0) \binom{N_k}{i} \prod_{x=n_k^*}^{i-1} p_k(x) \alpha_{k,fin}^i, & \text{for } n_k^* < i \le \lfloor \frac{C}{b_k} \rfloor \\ & \text{and } j = i b_k \end{cases} \quad (8)$$

Step 2) Determine the aggregated occupancy distribution $Q_{(-k)}$ based on the successive convolution of all service-classes apart from service-class k

$$Q_{(-k)} = q_1 * \cdots * q_{k-1} * q_{k+1} * \cdots * q_{2K}.$$

The term "successive" means that we initially convolve q_1 and q_2 in order to obtain q_{12}. Then we convolve q_{12} with q_3 to obtain q_{123} etc. The convolution operation between two occupancy distributions of service-class k and r is defined as

$$q_k * q_r = \left\{ q_k(0) q_r(0), \sum_{x=0}^{1} q_k(x) q_r(1 - x), \ldots, \sum_{x=0}^{C} q_k(x) q_r(C - x) \right\}. \quad (9)$$

Step 3) Calculate the TC probabilities of service-class k based on $Q_{(-k)}$ (step 2) and q_k as

$$Q_{(-k)} * q_k = \left\{ Q_{(-k)}(0) q_k(0), \sum_{x=0}^{1} Q_{(-k)}(x) q_k(1 - x), \ldots, \sum_{x=0}^{C} Q_{(-k)}(x) q_k(C - x) \right\}. \quad (10)$$

Normalizing the values of (10), we obtain the occupancy distribution $q(j), j = 0, 1, \ldots, C$ via

$$q(0) = Q_{(-k)}(0) q_k(0) / G,$$
$$q(j) = \left(\sum_{x=0}^{j} Q_{(-k)}(x) q_k(j - x) \right) \Big/ G, \ j = 1, \ldots, C. \quad (11)$$

Based on $q(j)$'s, we propose the following formula for the TC probabilities of service-class k

$$B_k = \sum_{j=C-b_k+1}^{C} q(j) + \sum_{x=n_k^* b_k}^{C-b_k} (1 - p_k(x)) q_k(x) \sum_{y=x}^{C-b_k} Q_{(-k)}(C - b_k - y). \quad (12)$$

The first term of (12) refers to states j where there is no bandwidth available for service-class k calls. The second term refers to states $x = n_k^* b_k, \ldots, C - b_k$ where there is available bandwidth for service-class k calls but call blocking occurs due to the PrTH policy and the threshold n_k^*.

Based on (11), the system's utilization U (in channels) can be determined by

$$U = \sum_{j=1}^{C} j q(j). \quad (13)$$

The mean number of service-class k calls in the system, $E(n_k)$, can be determined by

$$E(n_k) = \sum_{j=1}^{C} y_k(j) q(j), \quad (14)$$

where $y_k(j)$ is the average number of service-class k calls in state j and can be calculated by

$$y_k(j) = \alpha_{k,fin} \frac{\sum_{x=0}^{j-b_k} (N_k - \lfloor x/b_k \rfloor) p_k(x) q_k(x) Q_{(-k)}(j - b_k - x)}{q(j)}. \quad (15)$$

The rationale behind (15) is similar to (6) of [33] and is based on the fact that LB does exist between the adjacent states $j - b_k$ and j in the PrTH-Q model.

As a general rule, the selection of $n_k^* >> E(n_k)$ for all service-classes decreases the effect of the PrTH policy on arriving calls and therefore leads to TC (or CC) probabilities that are close to those obtained by applying the CS policy.

If $N_k \to \infty$ for $k = 1, \ldots, 2K$ and the total offered traffic remains constant, then a Poisson process arises and we have the PrTH-R model whose PFS is the following [18]

$$P(\boldsymbol{n}) = G^{-1} \left(\prod_{k=1}^{2K} \prod_{x=n_k^*}^{n_k - 1} p_k(x) \frac{\alpha_k^{n_k}}{n_k!} \right), \quad (16)$$

where $\alpha_k = \lambda_k \mu_k^{-1}$ is the offered traffic-load (in erl) of service-class k calls and $G \equiv G(\boldsymbol{\Omega}) = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}} \left(\prod_{k=1}^{2K} \prod_{x=n_k^*}^{n_k - 1} p_k(x) \frac{\alpha_k^{n_k}}{n_k!} \right)$ is the normalization constant.

For the CBP calculation, we exploit the PFS of the PrTH-R model, and use the abovementioned 3-step convolution algorithm, whereby the only change is in (8) which becomes [18]

$$q_k(j)=\begin{cases} q_k(0)\frac{\alpha_k^i}{i!}, \quad \text{for } 1\le i\le n_k^* \text{ and } j=ib_k \\ q_k(0)\dfrac{\prod_{x=n_k^*}^{i-1} p_k(x)\alpha_k^i}{i!}, \text{ for } n_k^*<i\le \lfloor C/b_k\rfloor \\ \qquad\qquad\qquad\qquad \text{and } j=ib_k \end{cases}, \quad (17)$$

CBP and system's utilization can be determined by (12) and (13), respectively, while $E(n_k)$ is determined by (14) where $y_k(j)$ can be calculated by

$$y_k(j)=\alpha_k\frac{\sum_{x=0}^{j-b_k} p_k(x)q_k(x)Q_{(-k)}(j-b_k-x)}{q(j)}. \quad (18)$$

IV. The Proposed PrTH-RQ Model

A special case of the PrTH-Q and PrTH-R models is the PrTH-RQ model whereby new calls follow a Poisson process and handover calls a quasi-random process.

The GB equation for state $\boldsymbol{n}$ in the PrTH-RQ model is

$$\begin{aligned}&\sum_{k=1}^{K}\Big[\lambda_k\delta_k^-(\boldsymbol{n})p_k(n_k-1)P_{inf,fin}(\boldsymbol{n}_k^-)+(n_k+1)\mu_k\delta_k^+(\boldsymbol{n})\\&P_{inf,fin}(\boldsymbol{n}_k^+)\Big]+\sum_{k=K+1}^{2K}\Big[(N_k-n_k+1)v_k\delta_k^-(\boldsymbol{n})p_k(n_k-1)\\&P_{inf,fin}(\boldsymbol{n}_k^-)+(n_k+1)\mu_k\delta_k^+(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}_k^+)\Big]\\&=\sum_{k=1}^{K}\Big[\lambda_k\delta_k^+(\boldsymbol{n})p_k(n_k)P_{inf,fin}(\boldsymbol{n})+n_k\mu_k\delta_k^-(\boldsymbol{n})\\&P_{inf,fin}(\boldsymbol{n})\Big]+\sum_{k=K+1}^{2K}\Big[(N_k-n_k)v_k\delta_k^+(\boldsymbol{n})p_k(n_k)\\&P_{inf,fin}(\boldsymbol{n})+n_k\mu_k\delta_k^-(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n})\Big],\end{aligned} \quad (19)$$

where $P_{inf,fin}(\boldsymbol{n}), P_{inf,fin}(\boldsymbol{n}_k^-), P_{inf,fin}(\boldsymbol{n}_k^+)$ are the probability distributions of states $\boldsymbol{n}, \boldsymbol{n}_k^-, \boldsymbol{n}_k^+$, respectively.

Besides, the following LB equations are extracted

$$\begin{aligned}&\lambda_k\delta_k^-(\boldsymbol{n})p_k(n_k-1)P_{inf,fin}(\boldsymbol{n}_k^-)=\\&n_k\mu_k\delta_k^-(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}), 1\le k\le K,\end{aligned} \quad (20)$$

$$\begin{aligned}&(N_k-n_k+1)v_k\delta_k^-(\boldsymbol{n})p_k(n_k-1)P_{inf,fin}(\boldsymbol{n}_k^-)=\\&n_k\mu_k\delta_k^-(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}), K+1\le k\le 2K,\end{aligned} \quad (21)$$

$$\begin{aligned}&\lambda_k\delta_k^+(\boldsymbol{n})p_k(n_k)P_{inf,fin}(\boldsymbol{n})=\\&(n_k+1)\mu_k\delta_k^+(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}_k^+), 1\le k\le K,\end{aligned} \quad (22)$$

$$\begin{aligned}&(N_k-n_k)v_k\delta_k^+(\boldsymbol{n})p_k(n_k)P_{inf,fin}(\boldsymbol{n})=\\&(n_k+1)\mu_k\delta_k^+(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}_k^+), K+1\le k\le 2K.\end{aligned} \quad (23)$$

The system of LB equations (20)-(23) is satisfied by the following PFS

$$\begin{aligned}P_{inf,fin}(\boldsymbol{n}) =&G^{-1}\Bigg[\prod_{k=1}^{K}\left(\frac{\alpha_k^{n_k}}{n_k!}\prod_{x=n_k^*}^{n_k-1}p_k(x)\right)\\&\prod_{k=K+1}^{2K}\left(\binom{N_k}{n_k}\alpha_{k,fin}^{n_k}\prod_{y=n_k^*}^{n_k-1}p_k(y)\right)\Bigg],\end{aligned} \quad (24)$$

where G is the normalization constant given by

$$\begin{aligned}G\equiv G(\boldsymbol{\Omega})=\sum_{\boldsymbol{n}\in\boldsymbol{\Omega}}&\Bigg[\prod_{k=1}^{K}\left(\frac{\alpha_k^{n_k}}{n_k!}\prod_{x=n_k^*}^{n_k-1}p_k(x)\right)\\&\prod_{k=K+1}^{2K}\left(\binom{N_k}{n_k}\alpha_{k,fin}^{n_k}\prod_{y=n_k^*}^{n_k-1}p_k(y)\right)\Bigg].\end{aligned} \quad (25)$$

For an efficient calculation of the various performance measures, we can exploit the PFS of the PrTH-RQ model, and use the 3-step convolution algorithm of Section III, whereby, $q_k(j)$'s for new and handover service-class k calls can be determined by (17) and (8), respectively. CBP and system's utilization can be determined by (12) and (13), respectively, while $E(n_k)$ can be determined by (14) where $y_k(j)$ is calculated by (18) for new calls and (15) for handover calls.

V. The Proposed MFCR-Q Model

Consider a cell of capacity C channels that accommodates calls of $2K$ service-classes under the MFCR policy, while the input traffic is quasi-random (MFCR-Q model). A service-class k call is new if $1\le k\le K$ and handover if $K+1\le k\le 2K$. Calls of service class k ($k=1,\ldots,2K$) come from a finite source population N_k. The mean arrival rate of service-class k idle sources is $\lambda_{k,fin}=(N_k-n_k)v_k$ where n_k is the number of in-service calls and v_k is the arrival rate per idle source. The offered traffic-load per idle source of service-class k is $\alpha_{k,fin}=v_k/\mu_k$ (in erl). If $N_k\to\infty$ for $k=1,\ldots,2K$, and the total offered traffic-load remains constant, then we have a Poisson process and the MFCR-R model [4].

The MFCR policy is described as follows: A call of service class k ($k=1,\ldots,2K$) requests b_k channels and has an MFCR parameter $t_{r,k}$ that expresses the real number of channels reserved to benefit calls of all other service-classes except from k. The reservation of $t_{r,k}$ channels is achieved because $\lfloor t_{r,k}\rfloor+1$ channels are reserved with probability $t_{r,k}-\lfloor t_{r,k}\rfloor$ while $\lfloor t_{r,k}\rfloor$ channels are reserved with probability $1-(t_{r,k}-\lfloor t_{r,k}\rfloor)$. As an example, calls of service-class k may have an MFCR parameter of $t_{r,k}=2.4$ channels. The reservation of 2.4 channels is achieved by assuming that $\lfloor 2.4\rfloor+1=3$ channels are reserved with probability 0.4 while $\lfloor 2.4\rfloor=2$ channels are reserved with probability 0.6. Let j be the occupied system's bandwidth ($j=0,1,\ldots,C$) when a service-class k call arrives in the cell. Then: a) if $C-j-\lfloor t_{r,k}\rfloor>b_k$, the call is accepted in the cell, b) if $C-j-\lfloor t_{r,k}\rfloor=b_k$, the call is accepted in the cell with probability $1-(t_{r,k}-\lfloor t_{r,k}\rfloor)$ and c) if $C-j-\lfloor t_{r,k}\rfloor<b_k$, there is no available bandwidth and the call is blocked and lost. An accepted call remains in the cell for a generally distributed service time with mean μ_k^{-1}.

The GB equation for state $\boldsymbol{n} = (n_1, \ldots, n_k, \ldots, n_{2K})$ in the MFCR-Q model is given by

$$\sum_{k=1}^{2K}\Big[(N_k - n_k + 1)v_k(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-)p_k(n_k+1)\mu_k P_{fin}(\boldsymbol{n}_k^+)\Big] = \sum_{k=1}^{2K}\Big[(N_k - n_k)v_k(\boldsymbol{n})P_{fin}(\boldsymbol{n}) + n_k\mu_k P_{fin}(\boldsymbol{n})\Big], \quad (26)$$

where

$$v_k(\boldsymbol{n}) = \begin{cases} v_k, & \text{for } C - \boldsymbol{nb} > b_k + \lfloor t_{r,k} \rfloor \\ \Big(1 - (t_{r,k} - \lfloor t_{r,k} \rfloor)\Big)v_k, & \text{for } C - \boldsymbol{nb} = b_k + \lfloor t_{r,k} \rfloor \\ 0, & \text{otherwise} \end{cases}, \quad (27)$$

$\boldsymbol{n}_k^- = (n_1, \ldots, n_k - 1, ..., n_{2K})$, $\boldsymbol{n}_k^+ = (n_1, \ldots, n_k + 1, \ldots, n_{2K})$ and $P_{fin}(\boldsymbol{n}), P_{fin}(\boldsymbol{n}_k^-), P_{fin}(\boldsymbol{n}_k^+)$ are the probability distributions of states $\boldsymbol{n}, \boldsymbol{n}_k^-, \boldsymbol{n}_k^+$, respectively.

The proposed model does not have a PFS for the determination of the steady state probabilities $P_{fin}(\boldsymbol{n})$ since LB can be destroyed between states $\boldsymbol{n}_k^-$, $\boldsymbol{n}$ or $\boldsymbol{n}$, $\boldsymbol{n}_k^+$. This means that $P_{fin}(\boldsymbol{n})$'s (and consequently all performance measures) can be determined by solving the GB equations, a realistic task only for cells of very small capacity and two or three service-classes.

To circumvent this problem, we prove an approximate but recursive formula for the calculation of the occupancy distribution, $q_{fin}(j)$, of the proposed MFCR-Q model. By definition

$$q_{fin}(j) = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} P_{fin}(\boldsymbol{n}), \quad (28)$$

where $\boldsymbol{\Omega}_j$ is the set of states whereby exactly j channels are occupied by all in-service calls, i.e. $\boldsymbol{\Omega}_j = \{\boldsymbol{n} \in \boldsymbol{\Omega} : \boldsymbol{nb} = j\}$. Since $j = \boldsymbol{nb} = \sum_{k=1}^{2K} n_k b_k$ we write (28) as follows

$$j q_{fin}(j) = \sum_{k=1}^{2K} b_k \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} n_k P_{fin}(\boldsymbol{n}). \quad (29)$$

To determine the $\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} n_k P_{fin}(\boldsymbol{n})$ in (28), we assume (this is an approximation) that LB exists between the adjacent states $\boldsymbol{n}_k^-$, $\boldsymbol{n}$ and has the form

$$(N_k - n_k + 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = n_k P_{fin}(\boldsymbol{n}), \quad (30)$$

where $\alpha_{k,fin}(\boldsymbol{n}_k^-) = v_k(\boldsymbol{n}_k^-)/\mu_k$.

Summing both sides of (30) over $\boldsymbol{\Omega}_j$ we have

$$\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j}(N_k - n_k + 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} n_k P_{fin}(\boldsymbol{n}). \quad (31)$$

The left hand side of (31) can be written as

$$\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j}(N_k - n_k + 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = N_k \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} \alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) - \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j}(n_k - 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-). \quad (32)$$

Since $\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} \alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = \alpha_{k,fin}(j - b_k)q_{fin}(j - b_k)$ the first term of the right hand side of (32) becomes

$$N_k \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} \alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = N_k \alpha_{k,fin}(j - b_k)q_{fin}(j - b_k), \quad (33)$$

where

$$\alpha_{k,fin}(j - b_k) = \begin{cases} \alpha_{k,fin}, & \text{for } j < C - \lfloor t_{r,k} \rfloor \\ \Big(1 - (t_{r,k} - \lfloor t_{r,k} \rfloor)\Big)\alpha_{k,fin}, & \text{for } j = C - \lfloor t_{r,k} \rfloor \\ 0, & \text{for } j > C - \lfloor t_{r,k} \rfloor. \end{cases} \quad (34)$$

Note: Equation (34) incorporates the MFCR policy in the calculation of $q_{fin}(j)$'s. E.g., assume that the occupied system's bandwidth is $j - b_k$ at the time of arrival of a service-class k call which requires b_k channels and has an MFCR parameter $t_{r,k}$. Then: a) the call will surely be accepted in the system and the new state will be j, if $j < C - \lfloor t_{r,k} \rfloor$, b) the call will be accepted in the system with probability $\big(1 - (t_{r,k} - \lfloor t_{r,k} \rfloor)\big)$ and the new state will be j, if $j = C - \lfloor t_{r,k} \rfloor$ and c) the call will be blocked and lost if $j > C - \lfloor t_{r,k} \rfloor$. **End of note**.

The second term of the right hand side of (32) is written as

$$\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j}(n_k - 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = \alpha_{k,fin}(j - b_k)y_{k,fin}(j - b_k)q_{fin}(j - b_k), \quad (35)$$

where $y_{k,fin}(j - b_k)$ is the average number of service-class k calls in state $j - b_k$.

Based on (33)-(35), (32) takes the form

$$\sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j}(N_k - n_k + 1)\alpha_{k,fin}(\boldsymbol{n}_k^-)P_{fin}(\boldsymbol{n}_k^-) = \alpha_{k,fin}(j - b_k)(N_k - y_{k,fin}(j - b_k))q_{fin}(j - b_k). \quad (36)$$

Equation (31) due to (36) is written as

$$(N_k - y_{k,fin}(j - b_k))\alpha_{k,fin}(j - b_k)q_{fin}(j - b_k) = \sum_{\boldsymbol{n} \in \boldsymbol{\Omega}_j} n_k P_{fin}(\boldsymbol{n}). \quad (37)$$

Equation (29) due to (37) is written as

$$j q_{fin}(j) = \sum_{k=1}^{2K}(N_k - y_{k,fin}(j - b_k))\alpha_{k,fin}(j - b_k)b_k q_{fin}(j - b_k). \quad (38)$$

In (38), the values of $y_{k,fin}(j - b_k)$ are not known. To determine them, we use a lemma of [48]: Two stochastic systems are equivalent and result in the same CBP, if they have the same: (a) traffic parameters $(2K, N_k, \alpha_{k,fin})$ where $k = 1, \ldots, 2K$ and (b) set of states.

Our purpose is therefore to find a new stochastic system, whereby we can calculate $y_{k,fin}(j - b_k)$. The channel requirements of calls and the capacity in the new system are chosen according to two criteria: 1) conditions (a) and (b) are valid and 2) each state has a unique occupancy j.

Now, state j is reached only via the previous state $j - b_k$. Thus, $y_{k,fin}(j-b_k) = n_k - 1$. Based on the above, (38) is given by

$$q_{fin}(j) = \begin{cases} 1, \text{ for } j = 0 \\ \frac{1}{j}\sum_{k=1}^{2K}(N_k - n_k + 1)\alpha_{k,fin}(j-b_k) \\ b_k q_{fin}(j-b_k), \text{ for } j = 1, \ldots, C, \end{cases} \tag{39}$$

where $\alpha_{k,fin}(j-b_k)$ is given by (34).

The calculation of $q_{fin}(j)$'s via (39) requires the unknown value of n_k. The determination of n_k's requires the state space determination of the equivalent system, a complex procedure especially for large systems. Thus, we approximate n_k in state j, $n_k(j)$, as $y_k(j)$, when Poisson arrivals are considered, i.e., $n_k(j) \approx y_k(j)$, and determine $q_{fin}(j)$'s via the formula

$$q_{fin}(j) = \begin{cases} 1, \text{ for } j = 0 \\ \frac{1}{j}\sum_{k=1}^{2K}(N_k - y_k(j-b_k))\alpha_{k,fin}(j-b_k) \\ \quad b_k q_{fin}(j-b_k), \text{ for } j = 1, \ldots, C, \end{cases} \tag{40}$$

where the values of $y_k(j)$'s are given by the MFCR-R model [4]

$$y_k(j) = \begin{cases} \frac{\alpha_k q(j-b_k)}{q(j)}, \text{ for } j < C - \lfloor t_{r,k} \rfloor \\ \frac{(1-(t_{r,k}-\lfloor t_{r,k}\rfloor))\alpha_k q(j-b_k)}{q(j)}, \text{ for } j = C - \lfloor t_{r,k} \rfloor \\ 0, \text{ for } j > C - \lfloor t_{r,k} \rfloor. \end{cases} \tag{41}$$

The values of $q(j)$'s in (41) are determined by [4]

$$q(j) = \begin{cases} 1, \text{ for } j = 0 \\ \frac{1}{j}\sum_{k=1}^{2K}\alpha_k(j-b_k)b_k q(j-b_k), \text{for } j = 1, \ldots, C, \end{cases} \tag{42}$$

where

$$\alpha_k(j-b_k) = \begin{cases} \alpha_k, \text{ for } j < C - \lfloor t_{r,k} \rfloor \\ (1-(t_{r,k}-\lfloor t_{r,k}\rfloor))\alpha_k, \text{ for } j = C - \lfloor t_{r,k} \rfloor \\ 0, \text{ for } j > C - \lfloor t_{r,k} \rfloor \end{cases} \tag{43}$$

and $\alpha_k = \lambda_k \mu_k^{-1}$ is the total offered traffic-load of service-class k calls (in erl).

Having determined $q(j)$'s we calculate the TC probabilities of service-class k calls, B_k, as

$$B_k = \sum_{j=C-b_k-\lfloor t_{r,k}\rfloor+1}^{C} G^{-1}q_{fin}(j) + (t_{r,k} - \lfloor t_{r,k}\rfloor)G^{-1} q_{fin}(C - b_k - \lfloor t_{r,k}\rfloor), \tag{44}$$

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

CC probabilities of service-class k can be determined via (44) where $q_{fin}(j)$'s are calculated (via (40)) for a system with $N_k - 1$ traffic sources. The system's utilization can be determined by (13) while $E(n_k)$ can be determined by (14).

VI. The Proposed MFCR-RQ Model

In the proposed MFCR-RQ model, we assume that new calls of service-class k follow a Poisson process while handover calls of the same service-class k follow a quasi-random process.

The GB equation for state $\boldsymbol{n} = (n_1, \ldots, n_k, \ldots, n_{2K})$ in the MFCR-RQ model is expressed by

$$\begin{aligned} &\sum_{k=1}^{K}\Big[\lambda_k(\boldsymbol{n}_k^-)P_{inf,fin}(\boldsymbol{n}_k^-) + (n_k+1)\mu_k P_{inf,fin}(\boldsymbol{n}_k^+)\Big] \\ &+ \sum_{k=K+1}^{2K}\Big[(N_k - n_k + 1)v_k(\boldsymbol{n}_k^-)P_{inf,fin}(\boldsymbol{n}_k^-) + \\ &(n_k+1)\mu_k P_{inf,fin}(\boldsymbol{n}_k^+)\Big] \\ &= \sum_{k=1}^{K}\Big[\lambda_k(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}) + n_k\mu_k P_{inf,fin}(\boldsymbol{n})\Big] + \\ &\sum_{k=K+1}^{2K}\Big[(N_k - n_k)v_k(\boldsymbol{n})P_{inf,fin}(\boldsymbol{n}) + n_k\mu_k P_{inf,fin}(\boldsymbol{n})\Big], \end{aligned} \tag{45}$$

where $v_k(\boldsymbol{n})$ can be determined by (27).

Based on Section V, we propose the following formula for the calculation of $q_{inf,fin}(j)$'s

$$q_{inf,fin}(j) = \begin{cases} 1, \text{ for } j = 0 \\ \frac{1}{j}\sum_{k=1}^{K}\alpha_k(j-b_k)b_k q_{inf,fin}(j-b_k) + \\ \frac{1}{j}\sum_{k=K+1}^{2K}(N_k - y_k(j-b_k))\alpha_{k,fin}(j-b_k)b_k \\ \quad q_{inf,fin}(j-b_k), \text{ for } j = 1, \ldots, C, \end{cases} \tag{46}$$

where $\alpha_k(j)$'s and $y_k(j)$'s are given by (43) and (41), respectively.

TC probabilities of service-class k can be determined via (41), the system's utilization via (13) and the mean number of service-class k calls in the system via (14).

VII. Applicability of the Proposed Models in 4G OFDM Wireless Networks

We consider the downlink of an OFDM-based cell that has C_s subcarriers and let R, P_{cell} and B be the average data rate per subcarrier, the available power in the cell and the system's bandwidth, respectively. Following [49], let the entire range of channel gains or signal to noise ratios (SNRs) per unit power be partitioned into M consecutive (but non-overlapping) intervals and denote as g_m, $m = 1, \ldots, M$ the average channel gain of the mth interval.

To calculate the power $p_{R,m}$ required to achieve the data rate R of the subcarrier assigned to a call whose average channel gain is g_m we can use the Shannon-Hartley theorem

$$R = \frac{B}{C_s}\log_2(1 + g_m p_{R,m}). \tag{47}$$

A newly arriving service-class (m, k) call ($m = 1, \ldots, M$ and $k = 1, \ldots, K$) requires b_k subcarriers in order to be accepted in the cell (i.e., it has a data rate requirement $b_k R$) and has an average channel gain g_m. If these subcarriers are not available then the call is blocked and lost. Otherwise, the call remains in the cell for a generally distributed service time with mean μ_k^{-1}. Assuming that calls follow a Poisson process with rate λ_{mk} and that n_{mk} is the number of in-service calls of service-class (m, k) then the model has a PFS for the steady-state probabilities $P(\boldsymbol{n})$ [49]

$$P(\boldsymbol{n}) = G^{-1}\left(\prod_{m=1}^{M}\prod_{k=1}^{K}\frac{\alpha_{mk}^{n_{mk}}}{n_{mk}!}\right), \tag{48}$$

where $\boldsymbol{n} = (n_{11}, \ldots, n_{m1}, \ldots, n_{M1}, \ldots, n_{1K}, \ldots, n_{mK}, \ldots, n_{MK})$, $G = \sum_{\boldsymbol{n}\in\boldsymbol{\Omega}}\prod_{m=1}^{M}\prod_{k=1}^{K}\frac{\alpha_{mk}^{n_{mk}}}{n_{mk}!}$, $\boldsymbol{\Omega}$ is the state space of the system, $\boldsymbol{\Omega} = \{\boldsymbol{n} : 0 \le \sum_{m=1}^{M}\sum_{k=1}^{K} n_{mk}b_k \le C_s, 0 \le \sum_{m=1}^{M}\sum_{k=1}^{K} p_{R,m}n_{mk}b_k \le P_{cell}\}$ and $\alpha_{mk} = \lambda_{mk}/\mu_k$ is the offered traffic-load (in erl) of service-class (m, k) calls.

According to [49], all performance measures (e.g., CBP) are based on the determination of $P(\boldsymbol{n})$'s via (48). However, since the cardinality of $\boldsymbol{\Omega}$ grows as $(C_sP_{cell})^{MK}$, the applicability of (48) is limited to problems of moderate size. To circumvent this problem, we denote by $j_1 = \sum_{m=1}^{M}\sum_{k=1}^{K} n_{mk}b_k$ the occupied subcarriers, i.e., $j_1 = 0, \ldots, C_s$, and by $j_2 = \sum_{m=1}^{M}\sum_{k=1}^{K} p_{R,m}n_{mk}b_k$ the occupied power in the cell, $j_2 = 0, \ldots, P_{cell}$. Then, we propose an accurate and recursive formula for the determination of the occupancy distribution $q(j_1, j_2)$

$$q(j_1, j_2) = \begin{cases} 1, \text{ for } j_1 = j_2 = 0 \\ \dfrac{1}{j_1}\displaystyle\sum_{m=1}^{M}\sum_{k=1}^{K}\alpha_{mk}b_k q(j_1-b_k, j_2-p_{R,m}b_k), \\ \text{for } j_1 = 1, \ldots, C_s \text{ and } j_2 = 1, \ldots, P_{cell}. \end{cases} \tag{49}$$

The proof of (49) is similar to the K-R formula [33] and thus is omitted. Note that if $P_{cell} \to \infty$ then (49) coincides with the K-R formula and can be used for the CBP determination in the multirate loss model for OFDM wireless networks of [50].

Having obtained $q(j_1, j_2)$ we may calculate the CBP of service-class (m, k) via the formula

$$B_{m,k} = \sum_{\{(j_1+b_k>C_s)\cup(j_2+p_{R,m}b_k>P_{cell})\}} q(j_1, j_2). \tag{50}$$

Modifications of (49) give us the possibility to consider the case of quasi-random traffic (or a mixture of random and quasi-random traffic) and apply in the model of [49] the BR, the MFCR or the TH policy. As far as the PrTH policy is concerned, it can also be applied in the model of [49], by exploiting the PFS of (48) and the 3-step convolution algorithm of Section III.

VIII. Applicability of the Proposed Models in 5G Networks

While the development and standardization of 5G networks are still at their early stage, it is widely acknowledged that 5G systems will extensively rely on SDN and NFV [51]. Using these technologies, the network intelligence can be pushed towards network edges, e.g., by embracing the concept of mobile edge computing (MEC) [52], or kept highly centralized, e.g., in the form of a C-RAN [53]. Furthermore, the incorporation of self-organising network (SON) features enables more autonomous and automated cellular network planning, deployment, and optimization [54].

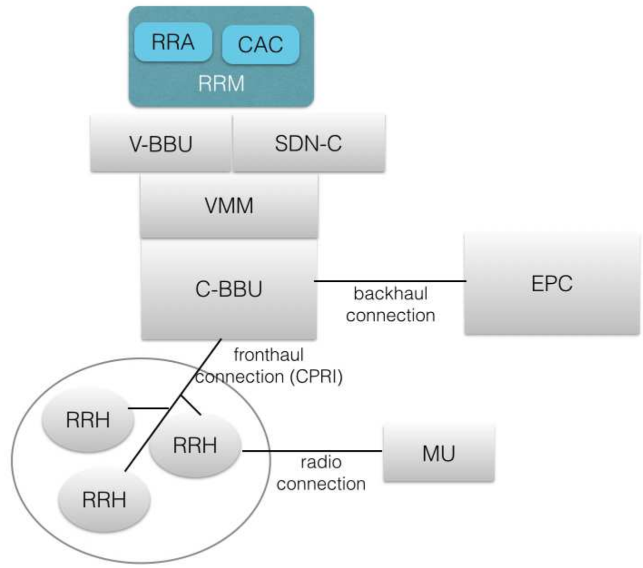

Fig. 2: The reference architecture.

In this section, we discuss the applicability of our proposed models in the context of new architectural and functional enhancements of next-generation cellular networks.

A. Reference Architecture

The considered reference architecture which is appropriate for the application of our models is presented in Fig. 2. This is in line with the C-RAN architecture, although it can also support a more distributed, MEC-like functionality, by incorporating, e.g., the SON features. At the RAN level, the architecture includes an SDN controller (SDN-C) and a virtual machine monitor (VMM) to enable NFV. Three main parts are distinguished: a pool of remote radio heads (RRHs), a pool of baseband units (BBUs), and the evolved packet core (EPC). The RRHs are connected to the BBUs via the common public radio interface (CPRI) with a high-capacity fronthaul. The BBUs form a centralized pool of data center resources (denoted as C-BBU). The C-BBU is connected to the EPC via the backhaul connection. To benefit from NFV, we consider virtualized BBU resources (V-BBU) [53] where the BBU functionality and services have been abstracted from the underlying infrastructure and virtualized in the form of virtual network functions (VNFs) [55]. To realize the virtualization, the VMM manages the execution of BBUs. Finally, the SDN-C is responsible for routing decisions and configuring the packet forwarding elements [56]. Among the BBU functions that could be virtualized in the form of a VNF, we focus on the RRM, which is responsible for call admission control (CAC) and radio resource allocation (RRA). The PrTH and MFCR policies could be implemented at the RRM level and enable sharing of V-BBU resources among the RRHs. An analytical framework is described in the next subsection.

B. Analytical Framework for Single-Cluster C-RAN

In this subsection, we adopt [57] and present the analysis for the case where all RRHs in the C-RAN form a single cluster. The analysis for the multi-cluster case is similar and is proposed in [58]. In both [57], [58], the C-RAN accommodates Poisson arriving calls of a single service-class under the CS policy. We propose a convolution algorithm for the CBP calculation of [57], thus facilitating the applicability of more complicated policies in the models of [57], [58] such as the MFCR and the PrTH policies and the extension of [57], [58] to include multiple service-classes.

Consider the C-RAN model of Fig. 3 where the RRHs are separated from the V-BBU, which performs the centralized baseband processing (of accepted calls). The total number of RRHs is L and each RRH has C subcarriers, which essentially represent units of the radio resource and can be allocated to the accepted calls. The V-BBU consists of T units (servers) of the computational resource, which are consumed for baseband processing.

Arriving calls follow a Poisson process with rate λ. An arriving call requires a subcarrier from the serving RRH and a unit of the computational resource. If these are available, then the call is accepted and remains in the system for a generally distributed service time with mean μ^{-1}. Otherwise, the call is blocked and lost. Based on this CAC mechanism which is actually the CS policy, the set of all possible states is given by: $\mathbf{\Omega} = \{\boldsymbol{n}\colon 0 \le n_1, \ldots, n_L \le C, 0 \le \sum_{l=1}^{L} n_l \le T\}$, where n_l is the number of in-service calls in the l-th RRH.

The number of in-service calls in all RRHs can be described by the steady-state vector $\boldsymbol{n} = (n_1, \ldots, n_l, \ldots, n_L)$. We further denote the vectors $\boldsymbol{n}_l^- = (n_1, \ldots, n_l - 1, \ldots, n_L)$, $\boldsymbol{n}_l^+ = (n_1, \ldots, n_l+1, \ldots, n_L)$. Then the probability distributions of $\boldsymbol{n}$, $\boldsymbol{n}_l^-$, and $\boldsymbol{n}_l^+$ are $P(\boldsymbol{n})$, $P(\boldsymbol{n}_l^-)$, and $P(\boldsymbol{n}_l^+)$, respectively. According to [57], the Markov chain of this model is reversible and therefore LB exists between the adjacent states $\boldsymbol{n}_l^-$ and $\boldsymbol{n}$: $\lambda P(\boldsymbol{n}_l) = n_l \mu P(\boldsymbol{n})$.

The system of LB equations is satisfied by the following PFS

$$P(\boldsymbol{n}) = G^{-1}\left(\prod_{l=1}^{L} \frac{\alpha^{n_l}}{n_l!}\right), \tag{51}$$

where $\alpha = \lambda/\mu$ is the offered traffic-load and $G \equiv G(\mathbf{\Omega}) = \sum_{\boldsymbol{n}\in\mathbf{\Omega}} \prod_{l=1}^{L} \frac{\alpha^{n_l}}{n_l!}$.

Having determined $P(\boldsymbol{n})$'s, the total CBP, B_{tot}, can be calculated by distinguishing two types of blocking events: 1) those that are caused due to insufficient subcarriers and are represented by the probability, B_{sub}, and 2) those that are caused due to insufficient units of the computational resource and are represented by the probability, B_{res}. Based on the above, we have [57]

$$B_{tot} = B_{sub} + B_{res}. \tag{52}$$

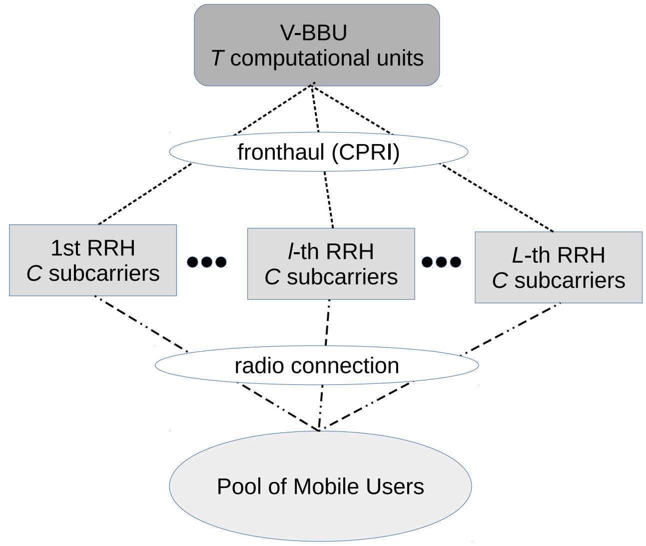

Fig. 3: C-RAN model (single cluster).

The values of B_{sub} can be determined via (51) by the following formula

$$B_{sub} = G\frac{\alpha^C}{C!} \sum_{\boldsymbol{n}\in\mathbf{\Omega}^{1,C}_{<T}} \prod_{l=2}^{L} \frac{\alpha_{n_l}}{n_l!}, \tag{53}$$

where $G = \left(\sum_{\boldsymbol{n}\in\mathbf{\Omega}} \prod_{l=1}^{L} \frac{\alpha_{n_l}}{n_l!}\right)^{-1}$, $\mathbf{\Omega}^{1,C}_{<T} = \{\mathbf{\Omega}^{1,C} \cap \mathbf{\Omega}_{<T}\}$, $\mathbf{\Omega}^{1,C} = \{\boldsymbol{n} : n_1 = C\}$, $\mathbf{\Omega}_{<T} = \{\boldsymbol{n} : n_1 + \cdots + n_L < T\}$. Note that (53) gives the values of B_{sub} for the 1st RRH. However, since the RRHs are identical and have the same capacities of C subcarriers, (53) refers to the B_{sub} of any RRH.

Similarly, by denoting $\mathbf{\Omega}_{=T} = \{\boldsymbol{n} : n_1 + \cdots + n_L = T\}$, the values of B_{res} are given by

$$B_{res} = \sum_{\boldsymbol{n}\in\mathbf{\Omega}_{=T}} P(\boldsymbol{n}). \tag{54}$$

For an efficient calculation of B_{tot}, we exploit (51) and propose a 3-step convolution algorithm:

Step 1) Determine the occupancy distribution of each of the L RRHs, $q_l(j)$, where $j = 1, \ldots, C$ and $l = 1, \ldots, L$

$$q_l(j) = q_l(0)\alpha^j/j!. \tag{55}$$

Step 2) Determine the aggregated occupancy distribution $Q_{(-l)}$ based on the successive convolution of all RRHs apart from the l-th RRH: $Q_{(-l)} = q_1 * q_2 * \cdots * q_{l-1} * q_{l+1} * \cdots * q_L$.

The convolution operation between two occupancy distributions q_k and q_r is given by

$$q_k * q_r = \left\{q_k(0)q_r(0), \sum_{x=0}^{1} q_k(x)q_r(1-x), \ldots, \sum_{x=0}^{T} q_k(x)q_r(T-x)\right\}. \tag{56}$$

Step 3) Calculate the total CBP, B_{tot}, based on the normalized values of the convolution operation of step 2, as follows

$$B_{tot} = B_{sub} + B_{res} = q_1(C)Q_{(-1)}(0) + q(T), \tag{57}$$

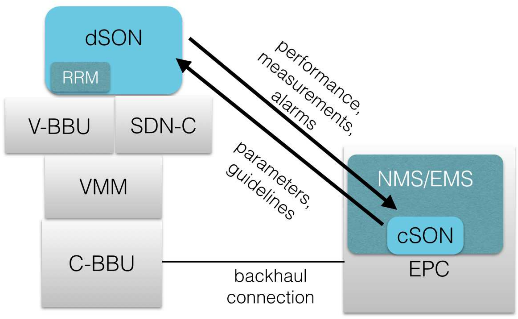

Fig. 4: Enabling a hybrid Self-Organising Network.

where $q(T) = G^{-1} \sum_{x=0}^{T} Q_{(-1)}(x) q_1(T-x)$ and G the normalization constant.

Based on the PFS of (51) and the convolution algorithm of (55)-(57) the model of [57] (and consequently of [58]) can be extended to include: a) multiple service-classes where calls may have different subcarrier and computational resource requirements per service-class, b) different call arrival processes per RRH or group of RRHs, thus allowing for a mixture of arrival processes (e.g., random and quasi-random traffic) and c) different sharing policies (e.g. the PrTH, the BR or the MFCR policy) for the allocation of resources in the RRHs or the V-BBU.

C. On the Implementation of Bandwidth Sharing Policies with SON in the C-RAN

Having formally described an analytical framework for the C-RAN, we now discuss the use of the SON technology for the implementation of the proposed policies. As an implementation example consider a virtualized RRM function in the C-RAN of Fig. 2.

Traditionally, SON functions refer to: self-planning, self-optimization, and self-healing. Implementing the RRM function as a SON function would mainly target the self-optimization objective, although this could also greatly facilitate the self-planning objective. In particular, the goal of self-optimization is as follows: During the cellular network operation, the self-optimization intends to improve the network performance or keep it at an acceptable level. The optimization could be performed in terms of QoS, coverage, and/or capacity improvements and is achieved by intelligently tuning various network settings of the base station (BS) as well as of the RRM function (e.g., CAC thresholds and RRA parameters).

In the literature, the following architectural approaches for implementing the SON functions have been proposed [54]:

- centralized SON (cSON): the SON functions are executed at the network management system (NMS) level or at the element management system (EMS) level. This will be particularly suitable for highly-centralized C-RAN solutions.
- distributed SON (dSON): the SON functions are executed at the BS level. They can be implemented either within a BS or in a distributed manner among cooperating BSs. This is beneficial for scenarios that require pushing the network intelligence closer to MUs.
- hybrid SON (hSON): a combination of cSON and dSON concepts (as shown in Fig. 4). According to this approach, some SON functionality is distributed and executed at the BS level, whereas other SON functionality is centralized at the NMS or EMS level.

The PrTH and MFCR policies, although they can be used under all three approaches, may bring most of the benefits under the hSON. Focusing on SON's self-optimization function, we consider the optimization of CAC. The CAC function (part of the dSON) admits or rejects a call, while the cSON function performs the selection of the optimization parameters for the CAC algorithm. These parameters are the CAC thresholds of the PrTH policy and the MFCR parameters. In fact, the thresholds of the PrTH policy and the MFCR parameters can be used not only for CAC, but also for determining the allocated bandwidth to both new and handover calls. When the cSON selects the CAC optimisation parameters, it considers the overall resource utilization in the cell, the QoS requirements of already accepted calls and the requirements of the new call. According to the PrTH policy, the parameters that are sent from the cSON to the dSON/RRM are the current values of the probability vector of (1). The cSON sends an updated probability vector if any changes in the performance guarantees are required (e.g., different acceptable levels of CBP). In the simplified case (e.g., when operating under tight resource or energy constraints) only the current values of the thresholds are sent to the dSON/RRM. In particular, the cSON determines the configuration parameters (e.g., CAC thresholds) based on a number of objectives (e.g., acceptable handover failure probabilities). This set of CAC thresholds can be easily determined, e.g., via (1), in the case of the PrTH-Q model. The dSON at the RAN receives the configuration parameters (e.g., CAC thresholds) and acts accordingly (e.g., rejects connection requests that do not conform to the specified parameters). If the measurements reported from the dSON to the cSON violate the objectives (e.g., the handover failure probability for a particular service is too high), the cSON will re-calculate and send updated configuration parameters to the dSON. This approach can result in a more autonomous and automated cellular network functionality and enables a simpler and faster decision-making and operation.

IX. Evaluation

The proposed models are evaluated through a comparison against other models and through simulation. To this end, we present three application examples. Simulation results are mean values of 7 runs. Each run is based on the generation of two million calls. To account for a warm-up period, the blocking events of the first 5% of these generated calls are not considered in the results. Due to the fact that the confidence intervals are very small, they are not presented in the figures that follow. The simulation tool used is Simscript III [59].

In the first example, we consider a cell of capacity $C = 150$ channels that accommodates two service-classes, with the traffic characteristics as shown in Table I. We provide analytical

TABLE I: Traffic characteristics (1st example)

Service-class	Traffic-load (erl)	Bandwidth (channels)	Threshold	Sources	Traffic-load per idle source (erl)
1st (new)	α_1 = 20.0	b_1 = 2	n_1^* = 35	100	$\alpha_{1,fin}$ = 0.20
2nd (new)	α_2 = 5.0	b_2 = 7	n_2^* = 10	100	$\alpha_{2,fin}$ = 0.05
1st (handover)	α_3 = 6.0	b_3 = 2	n_3^* = 70	100	$\alpha_{3,fin}$ = 0.06
2nd (handover)	α_4 = 1.0	b_4 = 7	n_4^* = 20	100	$\alpha_{4,fin}$ = 0.01

and simulation TC probabilities results for the proposed PrTH-RQ model considering two scenarios: (1) New calls of the 1st service-class behave as in the ordinary TH policy, i.e., $p_1(35) = p_1(36) = \ldots = p_1(75) = 0$, while new calls of the 2nd service-class are accepted in the system with probability $p_2(10) = p_2(11) = \ldots = p_2(20) = 0.5$, and $p_2(21) = 0$, (2) New calls of the 1st service-class are accepted in the system with probability $p_1(35) = p_1(36) = \ldots = p_1(74) = 0.7$, and $p_1(75) = 0$ while new calls of the 2nd service-class are accepted as in scenario 1. For both scenarios, we assume that $p_3(\cdot) = p_4(\cdot) = 0.95$, for all possible states equal or above the corresponding thresholds. These TC probabilities results are compared with the TC probabilities: a) for random new and handover traffic and the CS policy [33], [34] as well as the BR policy [60] and the PrTH policy (PrTH-R model) [18] b) for random new and quasi-random handover traffic and the CS or the BR policy [8], c) for random new and handover traffic and the MFCR-R model [4], d) for quasi-random new and handover traffic and the proposed MFCR-Q model and e) for random new and quasi-random handover traffic and the proposed MFCR-RQ model. In the MFCR policy, the MFCR parameters are $t_{r,1} = t_{r,3} = 4.7$ channels and $t_{r,2} = t_{r,4} = 0$. In the BR policy, the BR parameters are $t_1 = t_3 = 5$ channels and $t_2 = t_4 = 0$ so as to achieve TC probabilities equalization among calls of both service-classes. The BR parameters of a service-class k denote the number of channels reserved to benefit calls of all service-classes, apart from k. In the x-axis of Figs. 5-8 the offered traffic load of new and handover calls of both service-classes increases in steps of 1.0, 0.2, 0.5 and 0.1 erl, respectively. So, point 1 refers to: $(\alpha_1,\alpha_2,\alpha_3,\alpha_4)$ = (20.0, 5.0, 6.0, 1.0) while point 11 to: $(\alpha_1,\alpha_2,\alpha_3,\alpha_4) = (30.0, 7.0, 11.0, 2.0)$.

Figures 5-8 show that: (a) The PrTH policy clearly affects the TC probabilities of both service-classes; thus, it allows for a fine congestion control aiming at guaranteeing certain QoS to each service-class (particularly for handover calls). (b) The TC probabilities obtained for random handover traffic are higher compared to the corresponding results obtained for quasi-random handover traffic. This is anticipated due to the finite number of traffic sources in the case of quasi-random handover traffic. (c) Simulation results are almost identical to the corresponding analytical results of the PrTH loss models. This was anticipated since the PrTH loss models have a PFS. (d) The existing CS and BR policies fail to approximate the results obtained from the PrTH or the MFCR policy; this fact reveals the necessity of both policies.

In the second example, we consider a cell of $C = 200$ channels that accommodates random new and handover calls of $K = 2$ service-classes with the following traffic characteristics: α_1=10 erl, α_3=1 erl, $b_1 = b_3$=1 channel, and α_2=4 erl, α_4= 1 erl, $b_2 = b_4$=17 channels. The TH (or the PrTH) policy is applied only to new calls of the 2nd service-class (PrTH-R model). We consider the values $n_2^* = 8, 6$ calls and let $p_2(\cdot) = 0.7$ and 0.4 for all possible states equal or above the corresponding thresholds. We do not consider the case of quasi-random handover traffic since the conclusions are similar. The motivation behind this example is the following: the high difference between the channel requirements of the two service-classes ($b_1 = 1$ and b_2=17 channels) results in TC probabilities oscillations of the 1st service-class calls (new or handover), when α_1 increases. In the x-axis of Figs. 9-11 the offered traffic-load of α_1 increases in steps of 1 erl up to 40 erl while the offered traffic-loads of α_2, α_3 and α_4 remain constant. In Figs. 9-11, we present the TC probabilities of the 1st service-class calls (new or handover), 2nd service-class new calls and 2nd service-class handover calls, respectively. As a comparison we present the corresponding TC probabilities for the CS, the BR and the MFCR policies. In the MFCR policy, the MFCR parameters are $t_{r,1} = t_{r,3} = 6.4$ channels and $t_{r,2} = t_{r,4} = 0$. Figure 9 shows that TC probabilities oscillations appear not only in the CS policy but also in the PrTH and the MFCR policies. To intuitively explain oscillations, consider an instant where a call of the 1st service-class arrives in the cell and finds 17 available channels. In that case, the call is accepted and the cell has 16 available channels. If now a call of the 2nd service-class arrives in the cell it will be blocked, leaving the 16 channels for calls of the 1st service-class. In such a case, an increase in α_1 will not lead to a TC probabilities increase. As α_1 continues to increase, TC probabilities of the 1st service-class calls will increase until another block of 17 channels becomes available to 1st service-class calls. Such oscillations show that attention is needed when dimensioning a system, especially when calls of a service-class require much more bandwidth than others. Figures 10, 11 show that decreasing n_2^*, increases the TC probabilities of the new 2nd service-class calls and decreases the corresponding TC probabilities of handover calls. In addition, changing the values of $p_2(\cdot)$ affects TC probabilities, assuming a constant n_2^*.

The motivation of the third example is to show that the MFCR-Q model: i) provides quite accurate TC probability results compared to simulation and ii) is consistent; the TC probabilities of the MFCR-Q model approach those of the MFCR-R model as sources increase. Since these conclusions are not affected by the existence of handover traffic, we only consider new traffic. Consider a cell of capacity $C = 60$ channels, that accommodates only new calls of three service-classes, with the traffic characteristics of Table II. In the case of quasi-random traffic, we consider the sets 1) $N_1 = N_2 = N_3 = 10$ and 2) $N_1 = N_2 = N_3 = 30$ sources. In both sets, $\alpha_{k,fin} = \alpha_k / N_k$.

In the x-axis of Figs. 12-14 the offered traffic load of the 1st, 2nd and 3rd service-class increases in steps of 0.5, 0.2 and

TABLE II: Traffic characteristics (3rd example)

Service-class	Traffic-load for Poisson traffic (erl)	Bandwidth (channels)	MFCR parameter (channels)
1st	$\alpha_1 = 1.0$	$b_1 = 1$	$t_{r,1} = 9.4$
2nd	$\alpha_2 = 1.0$	$b_2 = 5$	$t_{r,2} = 5.3$
3rd	$\alpha_3 = 1.0$	$b_3 = 10$	$t_{r,3} = 0.0$

0.1 erl, respectively. So, point 1 is: $(\alpha_1, \alpha_2, \alpha_3) = (1.0, 1.0, 1.0)$ while point 8 is: $(\alpha_1, \alpha_2, \alpha_3) = (4.5, 2.4, 1.7)$.

In Figs. 12-14, we present analytical TC probabilities results of the MFCR-Q, the MFCR-R and the models of [33], [34], [60] together with the MFCR-Q simulation results, for each service-class, respectively. All figures show that the analytical results of the MFCR-Q model: a) are close to the corresponding simulation results, a fact that validates the proposed formulas, b) are lower than those of the MFCR-R model, especially for $N_1 = N_2 = N_3 = 10$ sources, due to the finite number of traffic sources. In addition, Figs. 12-14, show that TC probabilities of the 3rd service-class are reduced due to the MFCR policy at the cost of substantially increasing the TC probabilities of the other two service-classes.

X. Conclusion and Future Work

We propose four teletraffic multirate loss models for a single cell that accommodates random or quasi-random traffic under the PrTH and the MFCR policies. The system is analysed as a multirate loss system. The calculation of performance measures in the proposed models is done via accurate convolution algorithms (PrTH models) or via recursive formulas (MFCR models). The high accuracy of the proposed models is verified through simulation. Comparison against other models (under the CS or the BR policy) reveals the necessity of the new models. We also discuss the applicability of our models in OFDM and 5G cellular networks.

As a future work we intend to extend the proposed models to include the case of multiple cells. A possible framework of how this can be achieved is described in various papers (see e.g., [61]–[64]) whose models are related either to the Erlang-B formula [61] or the K-R formula [62], [63] or convolution algorithms [64]. In [61], an approximate loss model is proposed for the CBP determination in a system of two access links that cooperate with each other by interchanging capacities using an offloading scheme. This system of links accommodates only one service-class while each link adopts a variant of the CS policy. The load balancing mechanism of [61] can easily be incorporated in our proposed models. In [62], a multirate loss model that includes a load balancing mechanism is proposed for the CBP calculation in a group of LTE cells that supports non-elastic traffic. In [63], the work of [62] has been extended to include the case of elastic and adaptive traffic. The analysis of [62], [63] can be extended to include our proposed models. Finally, in [64], a multirate loss model is proposed for the CBP calculation in hierarchical loss networks. The main application example of this model is a two-tier hierarchical cellular network, where a macro-cell, overlays a set of microcells and the capacity of the macro-cells is shared among the overlaid micro-cells. In all cells, the

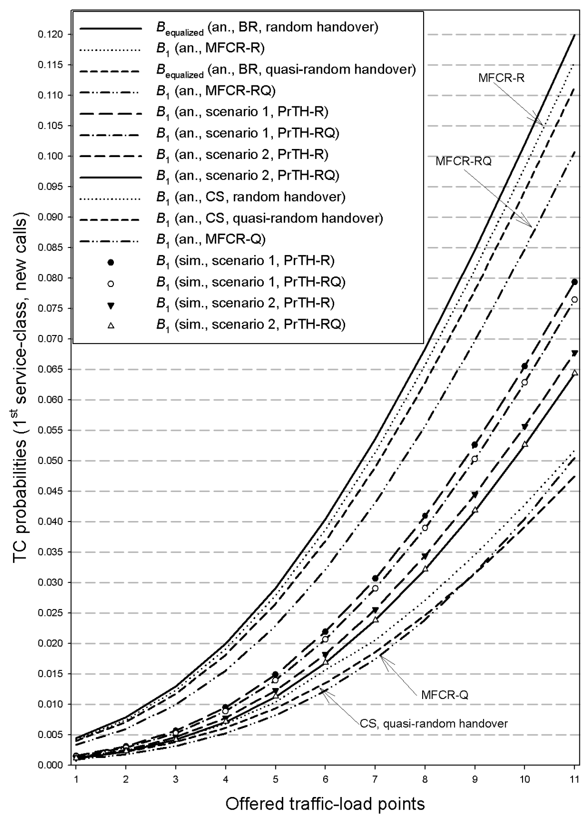

Fig. 5: TC probabilities - 1st service-class (new calls).

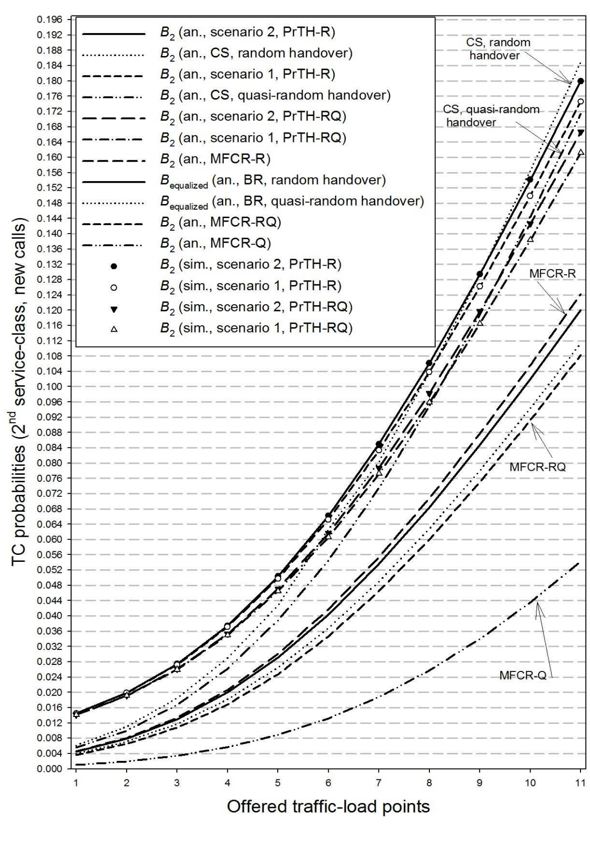

Fig. 6: TC probabilities - 2nd service-class (new calls).

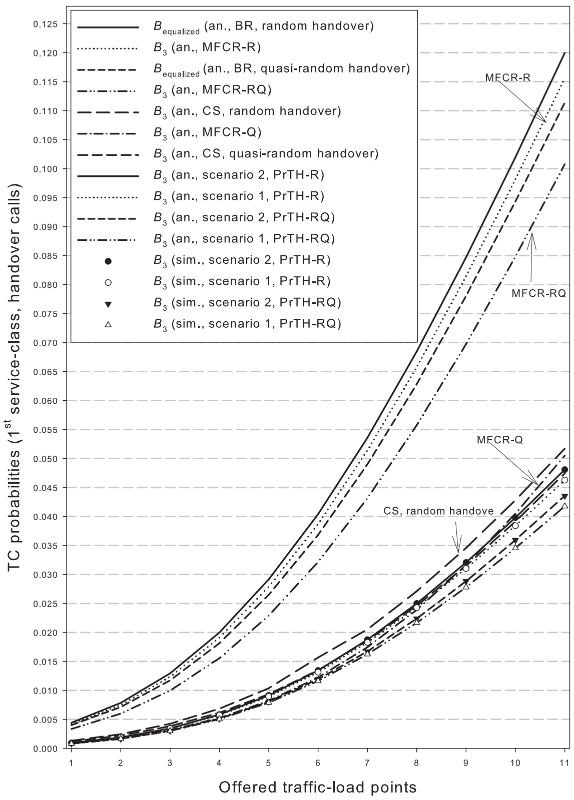

Fig. 7: TC probabilities - 1st service-class (handover calls).

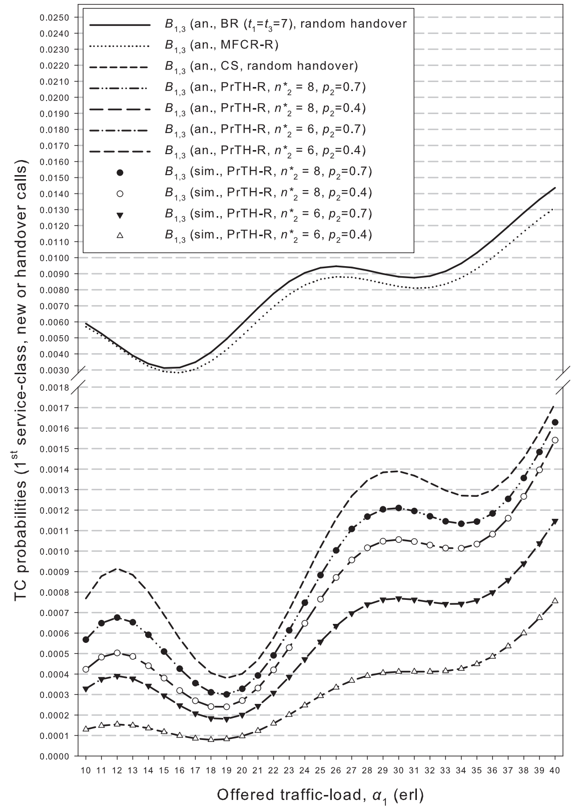

Fig. 9: TC probabilities - 1st service-class (new or handover calls, 2nd example).

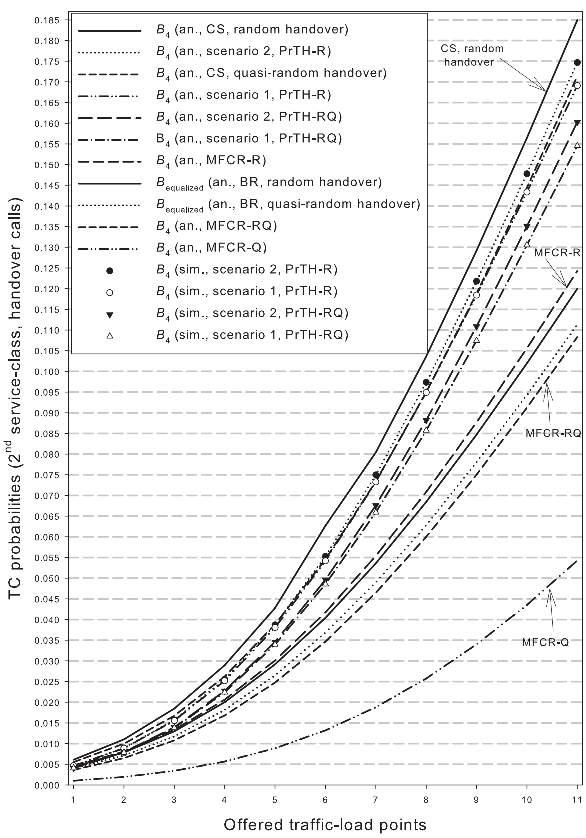

Fig. 8: TC probabilities - 2nd service-class (handover calls).

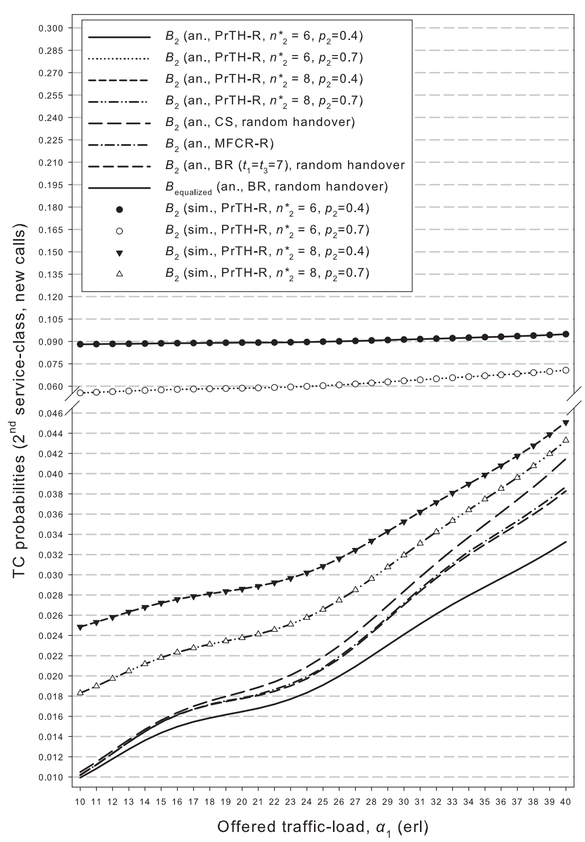

Fig. 10: TC probabilities - 2nd service-class (new calls, 2nd example).

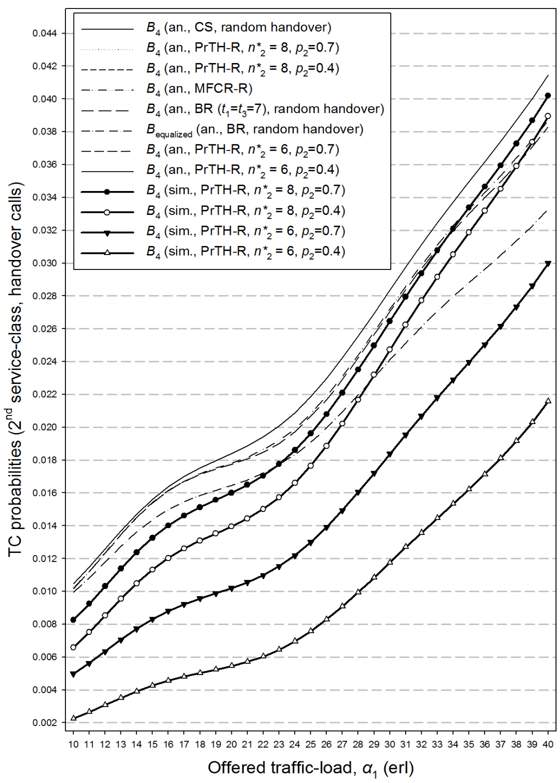

Fig. 11: TC probabilities - 2nd service-class (handover calls, 2nd example).

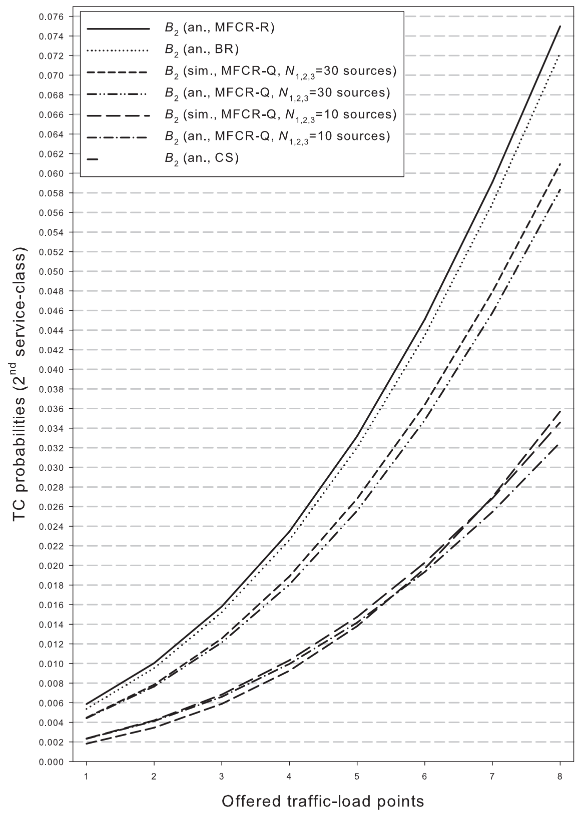

Fig. 13: TC probabilities - 2nd service-class (new calls, 3rd example).

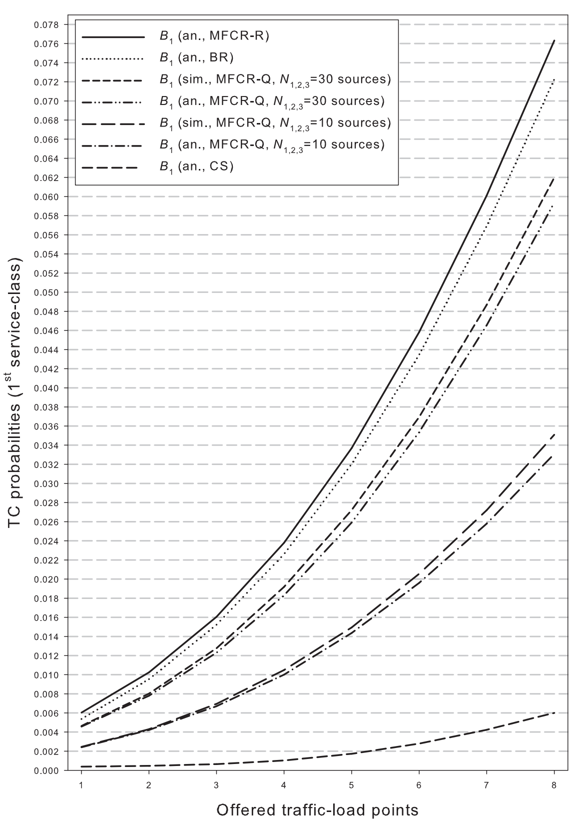

Fig. 12: TC probabilities - 1st service-class (new calls, 3rd example).

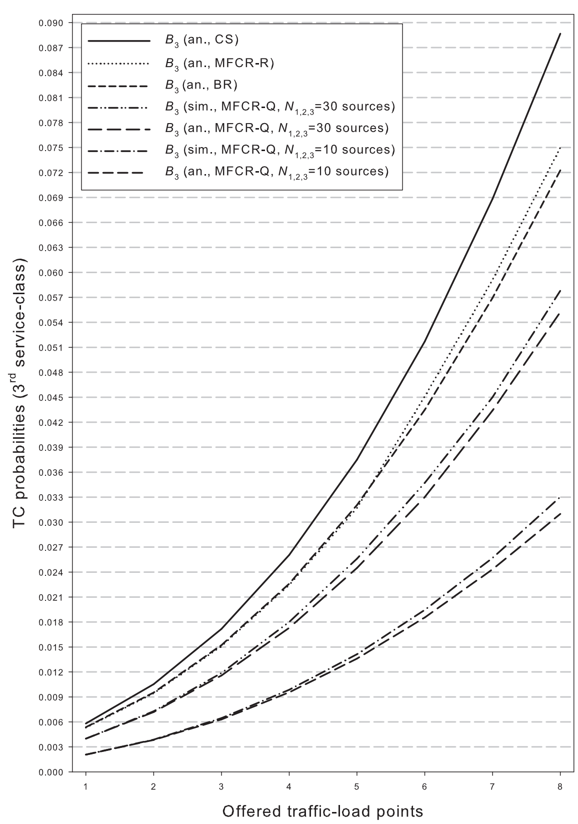

Fig. 14: TC probabilities - 3rd service-class (new calls, 3rd example).

CS policy is used. Clearly, the proposed policies (MFCR and PrTH) could be applied in such loss networks following the analysis of [64]. Another future direction can be the analysis of optimal resource allocation in 4G/5G networks under the proposed policies. Such an analysis could be based on the recent works of [65], [66].

REFERENCES

[1] L. Pierucci, "The quality of experience perspective toward 5G technology," IEEE Wireless Commun., vol. 22, no. 4, pp. 10-16, Aug. 2015.

[2] M. Yang, Y. Li, D. Jin, L. Zeng, X. Wu, and A. V. Vasilakos, "Software-defined and virtualized future mobile and wireless networks: A survey," Mobile Networks and Applications, vol. 20, no. 1, pp. 4-18, Feb. 2015.

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

[4] F. Cruz-Pérez, J. Vázquez-Ávila, and L. Ortigoza-Guerrero, "Recurrent formulas for the multiple fractional channel reservation strategy in multi-service mobile cellular networks," IEEE Commun. Letters, vol. 8, no. 10, pp. 629-631, Oct. 2004.

[5] J. Vázquez-Ávila, F. Cruz-Pérez, and L. Ortigoza-Guerrero, "Performance analysis of fractional guard channel policies in mobile cellular networks," IEEE Trans. Wireless Commun., vol. 5, no. 2, pp. 301-305, Feb. 2006.

[6] E. Bernal-Mor, D. Garcia-Roger, J. Martinez-Bauset, and V. Pla, "Optimal design of multiple fractional guard channel policy in multiservice cellular networks", Proc. 2nd UBICOMM, Valencia, Spain, Oct. 2008.

[7] M. Cello, G. Gnecco, M. Marchese and M. Sanguineti, "Optimality conditions for coordinate-convex policies in CAC with nonlinear feasibility boundaries," IEEE/ACM Trans. Networking, vol. 21, no. 5, pp. 1363-1377, Oct. 2013.

[8] M. Glabowski, A. Kaliszan, and M. Stasiak, "Asymmetric convolution algorithm for blocking probability calculation in full-availability group with bandwidth reservation," IET Circuits, Devices & Systems, vol. 2, issue 1, pp. 87-94, Feb. 2008.

[9] I. Moscholios and M. Logothetis, "The Erlang multirate loss model with batched Poisson arrival processes under the bandwidth reservation policy," Computer Commun., vol. 33, supplement 1, pp. S167-S179, Nov. 2010.

[10] A. Al Daoud, M. Alanyali, and D. Starobinski, "Reservation policies for revenue optimization from secondary spectrum access in cellular networks," Computer Networks, vol. 55, issue 1, pp. 214-224, Jan. 2011.

[11] M. Glabowski, "Modelling of state-dependent multirate systems carrying BPP traffic," Annals Telecommun., vol. 63, issue 7, pp. 393-407, Aug. 2008.

[12] S. Castellanos-López, F. Cruz-Pérez, M. Rivero-Ángeles, and G. Hernández-Valdez, "Performance analysis of coordinated cognitive radio networks under fixed-rate traffic with hard delay constraints," Journal of Communications and Networks, vol. 16, no. 2, pp. 130-139, April 2014.

[13] Y. Shen, T. Feng, W. Tang and Y. Lan, "Maximum achievable arrival rate of secondary users under GoS constraints in cognitive radio networks," Proc. IWCMC, Nicosia, Cyprus, Aug. 2014.

[14] S. Castellanos-López, F. Cruz-Pérez, M. Rivero-Ángeles, and G. Hernández-Valdez, "Call admission control strategy for cognitive radio networks with VoIP-traffic," Proc. Veh. Tech. Conf. (VTC Spring), Glasgow, U.K., May 2015.

[15] X. Lu, B. Yin, H. Zhang, "A reinforcement-learning approach for admission control in distributed network service systems," Journal of Combinatorial Optimization, vol. 31, issue 3, pp. 1241-1268, April 2016.

[16] I. Moscholios, M. Logothetis, J. Vardakas, and A. Boucouvalas, "Performance metrics of a multirate resource sharing teletraffic model with finite sources under both the threshold and bandwidth reservation policies," IET Networks, vol. 4, issue 3, pp. 195-208, May 2015.

[17] I. Moscholios, M. Logothetis, J. Vardakas and A. Boucouvalas, "Congestion probabilities of elastic and adaptive calls in Erlang-Engset multirate loss models under the threshold and bandwidth reservation policies," Computer Networks, vol. 92, part 1, pp. 1-23, Dec. 2015.

[18] I. Moscholios, V. Vassilakis, M. Logothetis, and A. Boucouvalas, "A probabilistic threshold-based bandwidth sharing policy for wireless multirate loss networks", IEEE Wireless Commun. Letters, vol. 5, issue 3, pp. 304-307, June 2016.

[19] A. Al Daoud, M. Alanyali, and D. Starobinski, "Pricing strategies for spectrum lease in secondary markets," IEEE/ACM Trans. Netw., vol. 18, no. 2, pp. 462-475, April 2010.

[20] X. Y. Yu and H. B. Zhu, "An efficient method for loss performance modeling of hierarchical heterogeneous wireless networks," Int. Journal of Commun. Systems, vol. 27, issue 6, pp. 956-968, June 2014.

[21] Z. Wang, P. Mathiopoulos, R. Schober, "Performance analysis and improvement methods for channel resource management strategies of LEO-MSS with multiparty traffic," IEEE Trans. Vehic. Tech., vol. 57, no. 6, pp. 3832-3842, Nov. 2008.

[22] Z. Wang, P. Mathiopoulos, and R. Schober, "Channel partitioning policies for multi-class traffic in LEO-MSS," IEEE Trans. Aerosp. Electron. Syst., vol. 45, no. 4, pp. 1320-1334, Oct. 2009.

[23] G. Boggia, P. Camarda and N. Di Fonzo, "Teletraffic analysis of hierarchical cellular communication networks," IEEE Trans. Vehic. Tech., vol. 52, issue 4, pp. 931-946, July 2003.

[24] M. Glabowski, M. Stasiak, A. Wisniewski and P. Zwierzykowski, "Blocking probability calculation for cellular systems with WCDMA radio interface servicing PCT1 and PCT2 multirate traffic," IEICE Trans. Commun., vol. E92-B, no.4, pp. 1156-1165, April 2009.

[25] V. Vassilakis, I. Moscholios and M. Logothetis, "Uplink blocking probabilities in priority-based cellular CDMA networks with finite source population," IEICE Trans. Commun., vol. E99-B, no. 6, pp. 1302-1309, June 2016.

[26] J. Juarez, R. Paulo, and F. Velez, "Modelling and simulation of B3G multi-service traffic in the presence of mobility," Wireless Networks, vol. 16, issue 3, pp. 659-669, April 2010.

[27] V. Goswami and P. Swain, "Analysis of finite population limited fractional guard channel call admission scheme in cellular networks," Procedia Engineering, vol. 30, pp. 759-766, 2012.

[28] C. Kim, V. Klimenok, and A. Dudin, "Analysis and optimization of guard channel policy in cellular networks with account of retrials," Computers & Operations Research, vol. 43, pp. 181-190, March 2014.

[29] D. Staehle and A. Mäder, "An analytic approximation of the uplink capacity in a UMTS network with heterogeneous traffic," Proc. 18th ITC, Berlin, pp. 81-91, Sept. 2003.

[30] G. Fodor and M.Telek, "Bounding the blocking probabilities in multirate CDMA networks supporting elastic services," IEEE/ACM Trans. Netw., vol. 15, issue 4, pp. 944-956, Aug. 2007.

[31] V. Vassilakis, G. Kallos, I. Moscholios, and M. Logothetis, "The wireless Engset multi-rate loss model for the call-level analysis of W-CDMA networks," Proc. 18th IEEE PIMRC, Athens, Greece, Sept. 2007.

[32] V. Vassilakis, G. Kallos, I. Moscholios, and M. Logothetis, "Call-level analysis of W-CDMA networks supporting elastic services of finite population," Proc. IEEE ICC, Beijing, China, May 2008.

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

[34] J. Roberts, "A service system with heterogeneous user requirements," Performance of Data Communications Systems and their Applications, North Holland, Amsterdam, pp.423-431, 1981.

[35] V. Vassilakis, G. Kallos, I. Moscholios, and M. Logothetis, "On the handoff-call blocking probability calculation in W-CDMA cellular networks," Proc. 4th AICT 2008, Athens, Greece, June 2008.

[36] M. Stasiak, P. Zwierzykowski, and D. Parniewicz, "Modelling of the WCDMA interface in the UMTS network with soft handoff mechanism," Proc. IEEE Globecom, Honolulu, USA, Dec. 2009.

[37] P. Garcia, A. Gonzalez, A. Alonso, B. Martinez, J. Perez, and A. Esguevillas, "Automatic UMTS system resource dimensioning based on service traffic analysis," EURASIP Journal on Wireless Commun. and Netw., pp. 1-10, Dec. 2012.

[38] S. Hanczewski, M. Stasiak, and P. Zwierzykowski, "A new model of the soft handover mechanism in the UMTS network," Proc. 9th IEEE/IET CSNDSP, Manchester, U.K., July 2014.

[39] M. Glabowski, S. Hanczewski, and M. Stasiak, "Modelling of cellular networks with traffic overflow," Mathematical Problems in Engineering, vol. 2015, article ID 286490, 15 pages, 2015, doi: 10.1155/2015/286490.

[40] S. Yashkov and A. Yashkova, "Processor sharing: A survey of the mathematical theory," Automation and Remote Control, vol. 68, issue 9, pp. 1662-1731, Sept. 2007.

[41] L. Lei, C. Lin, J. Cai, and X. Shen, "Flow-level performance of opportunistic OFDM-TDMA and OFDMA networks," IEEE Trans. Wireless Commun., vol. 7, no. 12, pp. 5461-5472, Dec. 2008.

[42] I. Moscholios, J. Vardakas, M. Logothetis, and A. Boucouvalas, "Congestion probabilities in a batched Poisson multirate loss model supporting elastic and adaptive traffic," Annals Telecommun., vol. 68, issue 5, pp. 327-344, June 2013.

[43] L. Yun, L. Man, C. Bin, W. Yong, and L. Wenjing, "Dynamic optimization of handover parameters adjustment for conflict avoidance in long term evolution," China Commun., vol. 10, no. 1, pp. 56-71, Jan. 2013.

[44] Y. Li, S. Xia, X. Xiong, S. Liu, X. Liu, and M. Daneshmand, "Throughput-delay trade-off scheduling in multi-channel downlink wireless networks," Wireless Pers. Commun., vol. 92, no. 2, pp. 681-694, Jan. 2017.

[45] B. Cao, F. He, Y. Li, C. Wang, and W. Lang, “Software defined virtual wireless network: Framework and challenges,” IEEE Network, vol. 29, no. 4, pp. 6-12, July-August 2015.
[46] M. Gerasimenko, D. Moltchanov, R. Florea, S. Andreev, Y. Koucheryavy, N. Himayat, S. P. Yeh, and S. Talwar, “Cooperative radio resource management in heterogeneous cloud radio access networks,” IEEE Access, vol. 3, pp. 397-406, April 2015.
[47] V. Iversen, Teletraffic Engineering and Network Planning, Technical University of Denmark, May 2011.
[48] G. Stamatelos and J. Hayes, “Admission control techniques with application to broadband networks,” Comput. Commun., 17 (9), pp. 663-673, Sept. 1994.
[49] C. Paik and Y. Suh, “Generalized queueing model for call blocking probability and resource utilization in OFDM wireless networks,” IEEE Commun. Letters, vol. 15, no. 7, pp. 767-769, July 2011.
[50] V. Pla, J. Martinez-Bauset, and V. Casares-Giner, “Comments on call blocking probability and bandwidth utilization of OFDM subcarrier allocation in next-generation wireless networks,” IEEE Commun. Lett., vol. 12, no. 5, pp. 349, May 2008.
[51] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” IEEE Commun. Magazine, vol. 52, no. 11, pp. 65-75, Nov. 2014.
[52] ETSI Technical White Paper, “Mobile Edge Computing,” Sept. 2014.
[53] A. Checko, H. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. Berger, and L. Dittmann, “Cloud RAN for mobile networks - a technology overview,” IEEE Commun. Surveys & Tutorials, vol. 17, no. 1, pp. 405-26, March 2015.
[54] 3GPP 32.500 v12.1.0, “Self-Organizing Networks (SON); Concepts and requirements (Release 12),” Sept. 2014.
[55] ETSI GS NFV-MAN 001 (V1.1.1) “Network Function Virtualisation (NFV); Management and Orchestration,” Dec. 2014.
[56] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, “Software defined mobile networks: Concept, survey, and research directions,” IEEE Commun. Mag., vol. 53, no. 11, pp. 126-133, Nov. 2015.
[57] J. Liu, S. Zhou, J. Gong, Z. Niu and S. Xu, “On the statistical multiplexing gain of virtual base station pools,” Proc. IEEE Globecom, Austin, TX, Dec. 2014.
[58] J. Liu, S. Zhou, J. Gong, Z. Niu and S. Xu, “Statistical multiplexing gain analysis of heterogeneous virtual base station pools in cloud radio access networks,” IEEE Trans. Wireless Commun., vol. 15, no. 8, pp. 5681-5694, Aug. 2016.
[59] Simscript III, http://www.simscript.com (last accessed: June 2017)
[60] J. Roberts, “Teletraffic models for the Telecom 1 Integrated Services Network,” Proc. 10th ITC, paper 1.1-2, Montreal 1983.
[61] V. Burger, M. Seufert, T. Hossfeld, and P. Tran-Gia, “Performance evaluation of backhaul bandwidth aggregation using a partial sharing scheme,” Physical Communication, vol. 19, pp. 135-144, June 2016.
[62] M. Glabowski, S. Hanczewski, and M. Stasiak, “Modelling load balancing mechanisms in self-optimizing 4G mobile networks,” Proc. 21st APCC, Kyoto, Japan, 2015.
[63] M. Glabowski, S. Hanczewski, and M. Stasiak, “Modelling load balancing mechanisms in self-optimizing 4G mobile networks with elastic and adaptive traffic,” IEICE Trans. Commun., vol. E99-B, no. 8, pp. 1718-1726, Aug. 2016.
[64] Q. Huang, K. Ko, and V. Iversen, “Approximation of loss calculation for hierarchical networks with multiservice overflows,” IEEE Trans. Commun., vol. 56, no. 3, pp. 466-473, March 2008.
[65] L. Tan, Z. Zhu, F. Ge, and N. Xiong, “Utility maximization resource allocation in wireless networks: Methods and algorithms,” IEEE Trans. Systems, Man and Cybernetics: Systems, vol. 45, no. 7, pp. 1018-1034, July 2015.
[66] L. Tan and Y. Zhang, “Optimal resource Allocation with principle of equality and diminishing marginal utility in wireless networks,” Wireless Personal Commun., vol. 84, no. 1, pp. 671-693, Sept. 2015.

Ioannis D. Moscholios was born in Athens, Greece, in 1976. He received the Dipl.-Eng. degree in Electrical & Computer Engineering (ECE) 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, ECE Department, University of Patras. From 2009 to 2013 he was a Lecturer in the Dept. of Telecommunications Science and Technology, University of Peloponnese, Tripolis, Greece. From 2013, he is an Assistant Professor in the Dept. of Informatics & Telecommunications, University of Peloponnese, Tripolis, Greece. His research interests include teletraffic engineering, simulation and performance analysis of communication networks. He has published over 130 papers in international journals/conferences. He has served as a Guest Editor in IET Communications, IET Networks, Mobile Information Systems and as an Associate Editor in IEICE Transactions on Communications. He is an IARIA Fellow and a member of the Technical Chamber of Greece (TEE).

Vassilios G. Vassilakis is a lecturer in the Department of Computer Science at the University of York, UK. He received his Ph.D. degree in Electrical & Computer Engineering from the University of Patras, Greece, in 2011. From 2011 to 2013, he was with the Network Convergence Laboratory, University of Essex, where he conducted research on information-centric networking and network security, and contributed to the EC FP7 PURSUIT project. In 2013, he joined the Institute for Communication Systems, University of Surrey, and conducted research on 5G wireless networks. After that, he was with the Computer Laboratory, University of Cambridge, where he conducted research on future Internet technologies and contributed to the EC H2020 RIFE project. In 2015, he joined the Secure and Dependable Software Systems group at the University of Brighton, where he conducted research on 5G network security and contributed to the EC H2020 SESAME project. His main research interests are in the areas of 5G wireless and mobile networks, future Internet technologies, network security, software-defined networks, and IoT. He has published over 80 journal/conference papers. He has served as a Guest Editor in IEICE Transactions on Communications, Elsevier Optical Switching & Networking, and IET Networks, and for the TPC of IEEE ICC and IEEE GLOBECOM.

Michael D. Logothetis 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 and 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 ECE Department of the University of Patras, and in 2009 elected (Full) 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 200 conference/journal papers and has over 770 third-part citations. He has published a teletraffic book in Greek. He has organised (General Chair) the IEEE/IET CSNDSP 2006, and IEICE ICTF 2016. He has served as TPC member of several international conferences while he has chaired many technical sessions. He has become a Guest Editor in five journals: a) Mediterranean Journal of Electronics and Communications, b) Mediterranean Journal of Computers and Networks, c) IET Circuits, Devices & Systems, d) IET Networks, and e) Ubiquitous Computing and Communication Journal. He is a member of the IARIA (Fellow), IEEE (Senior), IEICE, FITCE and the Technical Chamber of Greece (TEE).

Anthony C. Boucouvalas received the B.Sc. degree in electrical and electronic engineering from Newcastle University in 1978, and the M.Sc. and D.I.C. degrees in communications engineering and the Ph.D. degree in fiber optics from Imperial College, in 1979 and 1982, respectively. He was with GEC Hirst Research Centre, and became Group Leader and Divisional Chief Scientist, where he was involved in fiber optic components, measurements, and sensors. In 1987, he was with Hewlett Packard Laboratories as Project Manager, where he was involved in optical communication systems, optical networks, and instrumentation, until 1994, when he joined Bournemouth University. In 1996, he was a Professor of Multimedia Communications and in 1999 the Director of the Microelectronics and Multimedia Research Centre with Bournemouth University. In 2007, he was with the Department of Telecommunication Science, University of Peloponnese, where he served for six years as Head of Department. He is currently a Professor of Communications Networks and Applications with the University of Peloponnese. He was involved with research in fiber optic communications, wireless communications, and multimedia, and has an accumulated 35 years experience in well-known academic and industrial research centers. He has published over 310 papers. His current research interests include optical wireless communications, fibre optic communications, inverse fibre optic problems, network protocols, human-computer interfaces, and Internet applications. He is a Fellow of IEEE, IET and the Royal Society for the encouragement of Arts, Manufacturers and Commerce.