

RESEARCH

Open Access



# Modelling virtual radio resource management in full heterogeneous networks

Sina Khatibi<sup>1,2\*</sup>  and Luis M. Correia<sup>1</sup>

## Abstract

Virtual radio access networks (RANs) is the candidate solution for 5G access networks, the concept of virtualised radio resources completing the virtual RAN paradigm. This paper proposes a new analytical model for the management of virtual radio resources in full heterogeneous networks. The estimation of network capacity and data rate allocation are the model's two main components. Based on the probability distribution of the signal-to-interference-plus-noise-ratio observed at the user terminal, the model leads to the probability distribution for the total network data rate. It considers different approaches for the estimation of the total network data rate, based on different channel qualities, i.e., optimistic, realistic and pessimistic. The second component uses the outcome of the first one in order to maximise the weighted data rate subject to the total network capacity, the SLAs (service level agreements) of Virtual Network Operators (VNOs), and fairness. The weights for services in the objective function of the resource allocation component enable the model to have prioritisation among services. The performance of the proposed model is evaluated in a practical heterogeneous access network. Results show an increase of 2.5 times in network capacity by implementing an access point at the centre of each cell of a cellular network. It is shown that the cellular network capacity itself can vary from 0.9 Gbps in the pessimistic approach up to 5.5 Gbps in the optimistic one. Finally, the isolation of service classes and VNOs by means of virtualisation of radio resources is clearly demonstrated.

**Keywords:** Virtualisation of radio resources, Virtual radio resource management, Radio access networks, Network function virtualisation, 5G

## 1 Introduction

The monthly global data traffic is going to surpass 10 EB in 2017, as the result of the proliferation of smart devices and of traffic-hungry applications [1]. In order to address this issue, operators have to find a practical, flexible and cost-efficient solution for their networks expansion and operation, using the scarce available radio resources. The increment of cellular networks' capacity by deploying dense base stations (BSs) is the ground-work in any candidate solution.

In addition, traffic offloading, e.g., to Wi-Fi access points (APs), has proven to be a valuable complementary approach. According to [2, 3], an acceptable portion of traffic can be offloaded to APs, just by deferring delay-

tolerant services for a pre-specified maximum interval until reaching an AP. Offloading approaches are generally based on using other connectivity capabilities of mobile terminals, whenever it is possible, instead of using further expensive cellular bands. The authors in [4] discuss the economics of traffic offloading, and in [5] address an energy-saving analysis.

Nevertheless, drastic temporal and geographical variations of traffic, in addition to the shortage of network capacity, make the situation for operators even worse [6]. The usual provisioning of radio access networks (RANs) for busy hours leads to an inefficient resource usage with relatively high CAPEX and OPEX costs, which is not acceptable anymore. Instead, operators are in favour of flexible and elastic solutions, where they can also share their infrastructure.

\* Correspondence: khatibi@nomor.de

<sup>1</sup>Nomor Research GmbH, Munich, Germany

<sup>2</sup>University of Lisbon, Lisbon, Portugal

Lately, the sharing of network infrastructure using network function virtualisation (NFV) has become an active research topic to transform the way operators architect their networks [7]. In the same research path, the concept of virtualisation of radio resources for cellular networks is proposed in [8–10]. The key idea is to aggregate and manage all the available physical radio resources in a set of infrastructures, offering pay-as-you-go connectivity-as-a-service (CaaS) to virtual network operators (VNOs). Virtual radio resource management (VRRM) is a non-trivial task, since it has to serve multiple VNOs with different requirements and service level agreements (SLAs) over the same infrastructure. Furthermore, wireless links are always subject to fading and interference, hence, their performance is variable [11]. The proposed model for virtual radio resource management in [8–10] has two key parts: (i) estimation of available radio resources and (ii) allocation of the available resources estimated in the first step to the services of VNOs. In this paper an analytical description of the model is provided followed by the evaluation for a practical scenario. The novelty of this paper can be summarised as follows:

- I. This paper extends the analytical model for the management of virtual radio resources, considering full heterogeneous access networks, and including both non-cellular (e.g., Wi-Fi) and cellular (e.g., GSM, UMTS, LTE and whatever comes next in 5G–5th generation) networks. The key point in extending the model to non-cellular networks is the consideration of the effect of collision on the total network throughput. Consequently, the model has to consider the number of connected terminals to the Wi-Fi network, while optimising the other objectives.
- II. The techniques for estimating network capacity are improved with three extra approaches, i.e., optimistic, realistic and pessimistic ones, which approximate the model to a real network operation.
- III. A comprehensive study of the proposed model for VRRM in full heterogeneous networks under different channel quality conditions and different traffic load scenarios is given.

This paper is organised as follows: Section II addresses the background and related works, and Section III describes the proposed model for VRRM. The scenario for model evaluation is stated in Section IV. In Section V, numeric results are presented and discussed. The paper is concluded in Section VI.

## 2 Background and related works

Based on [12, 13], infrastructure sharing solutions can be categorised into three main types: geographical,

passive and active sharing. In geographical sharing or national roaming, a federation of operators can achieve full coverage in a short time, by dividing the service area into several regions, over which each of the operators provides coverage [14]. Passive sharing refers to the sharing agreement of fundamental infrastructures, such as tower masts, equipment houses and power supply, in order to reduce operational costs. Active sharing, however, is the sharing of transport infrastructures, radio spectrum and baseband processing resources. In [15], two types of sharing are introduced: multi-operator RAN and multi-core network. In the former, operators maintain a maximum level of independent control over their traffic quality and capacity, by splitting BSs and their controller nodes into logically independent units over a single physical infrastructure. In the latter, however, operators give up their independent control, by sharing the aforementioned entities in conjunction with the pooling of radio resources. Although the cost items in multi-core network are identical to multi-operator RAN, radio resources pooling leads to further savings in extremely low-traffic areas over equipment-related costs. Moreover, a network-wide radio resource management framework is proposed [16], in order to achieve isolation in addition to the optimal distribution of resources across the network.

Despite of RAN sharing benefits, surprisingly few sharing agreements have been made, especially in mature markets. The reasons offered by operators for not engaging into sharing deals are often the up-front transformation costs, the potential loss of control over their networks and the challenge of operational complexity [17, 18]. Sharing deals may be too expensive, and the initial cost of a network-sharing deal can be daunting; hence, operators without a comfortable margin of funds to make the necessary investment are likely to assume that they simply cannot afford to participate in such operation. 3rd Generation Partnership Project (3GPP) standards also limit the shared RAN to serve only four operators [19]. Moreover, many operators, particularly incumbent ones whose early entrance into markets has given them the best coverage and network qualities, assume that sharing their network with rivals would dilute their competitive advantage. Some of them may feel that they would not be able to control the direction for the development of their network in future rollout strategies and choices about hardware and vendors. Last, but not the least, having a shared RAN running properly is an elaborate and a complex task. Some operators believe that having a shared network operation puts many technological and operational challenges, which may lead to little financial benefits and great potential of chaos [19]. However, some studies, e.g. [20], show that the pros of sharing are larger than the cons, and that

this approach can really be seen as a very promising solution for the future, namely, in a broader perspective, i.e., looking at RAN virtualisation instead of RAN sharing as the candidate solution.

NFV has captured the attention of many researchers, e.g. [7], and some studies have also considered RAN as well. By introducing an entity called “hypervisor” on the top of physical resources, the authors in [21] addressed the concept of a virtualised eNodeB. The hypervisor allocates the physical resources among various virtual instances and coordinates multiple virtual eNodeBs over the same physical one; the LTE spectrum is shared among them using the concept of RAN sharing, i.e., each virtual eNodeB receives a portion of the available frequency bands. The virtualisation of BSs in LTE is also addressed in [22], by considering the resource allocation to be static or dynamic spectrum sharing among virtual operators. The authors in [23] look into the advantages of a virtualised LTE system, via an analytical model for FTP (File Transfer Protocol) transmissions; the evaluation is done with considerations on realistic situations to present multiplexing gain in addition to the analytical analysis.

As the next step in RAN virtualisation, the concept of radio resource virtualisation and a management model is proposed in [8], and the extension of the model to

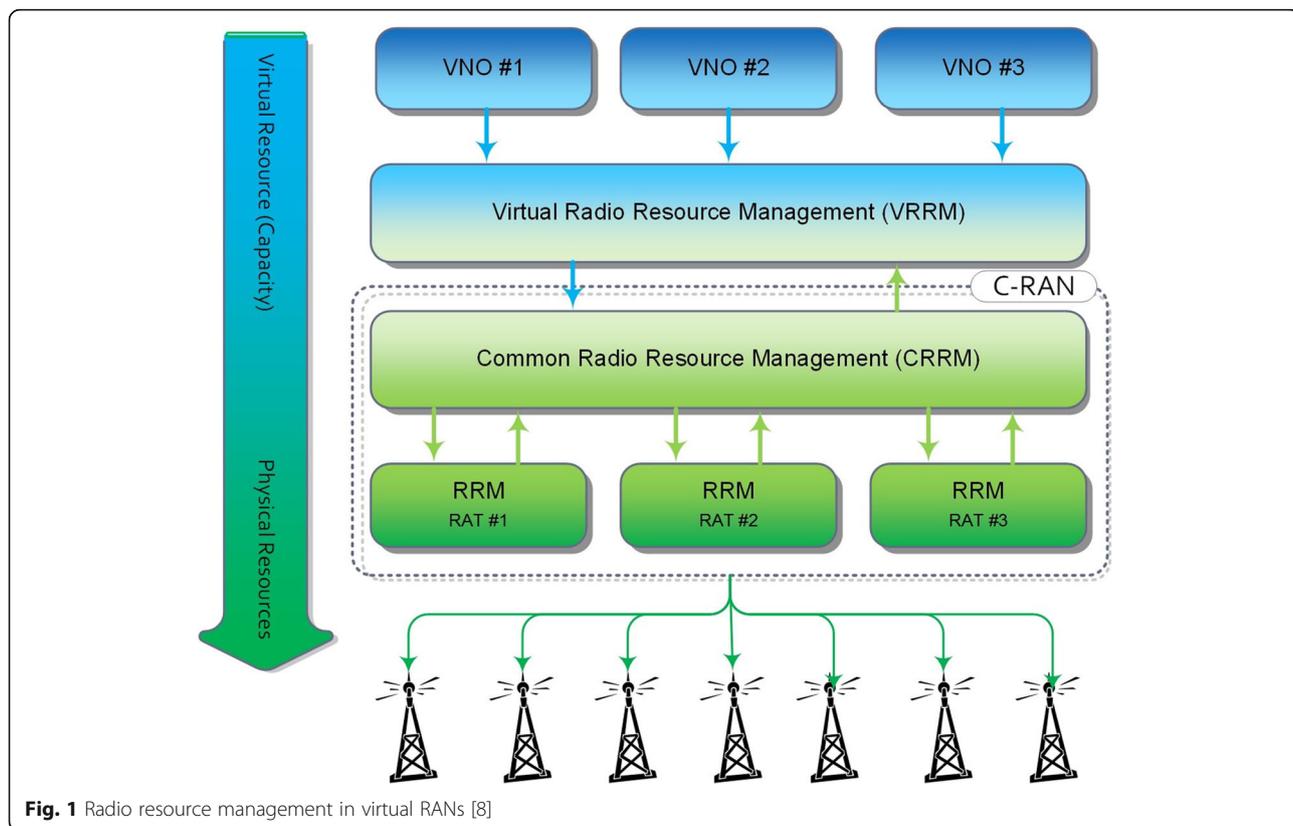
support the shortage of radio resources is presented in [9]. In the same research path, the current paper, as well as [10], considers the virtualisation of radio resources over a full heterogeneous access network (i.e., a combination of cellular networks and WLANs (wireless local area networks)), over which pay-as-you-go CaaS is offered to VNOs.

### 3 Radio resource management in virtual RANs

#### 3.1 Model

Figure 1 presents the hierarchy for the management of virtual radio resources, consisting of a VRRM entity on the top of the usual radio resource management (RRM) entities of heterogeneous access networks [24], i.e., common RRM (CRRM) and local RRMs (for each of the physical networks), the latter managing different RATs (Radio Access Technologies)

VNOs, placed at the top of the hierarchy, require wireless connectivity to be offered to their subscribers, not owning any radio access infrastructure [8, 9]. VNOs ask for RAN-as-a-service (RANaaS) from the RAN provider with physical infrastructure [25]. VNOs do not have to deal with the management of virtual RANs; they just define requirements, such as contracted capacity, in their SLAs with RAN providers. The role of VRRM is to translate VNOs’ requirements and SLAs into a set of



policies for the lower levels [8]. These policies contain data rates for different services, in addition to their priorities. Although VRRM optimises the usage of virtual radio resources, it does not deal with physical ones. However, VRRM has to consider practical issues, as the effect of the collision rate in WLANs on network data rate, in order to have an effective management of virtual radio resources. Reports and monitoring information provided by CRRM enables VRRM to improve policies. Load balancing among RANs, controlling the offloading procedure, is the duty of CRRM also known as Join RRM. Finally, the local RRRMs are in charge of managing the physical resources based on the policies of VRRM and CRRM.

The VNOs' SLAs can generally be categorised into three main groups:

- Guaranteed bitrate (GB), in which the RAN provider guarantees the VNO a minimum and a maximum level of data rates, regardless of the network status. Allocating the maximum guaranteed data rate to the VNO leads to its full satisfaction. The upper boundary in this type of SLA enables VNOs to have full control on their networks. For instance, a VNO offering VoIP (voice over IP) to its subscribers may foresee to offer this service to only 30 up 50% of its subscribers simultaneously, hence, the VNO can put this policy into practice by choosing a guaranteed SLA for its VoIP service. It is expected that subscribers always experience a good quality of service (QoS) in return of relatively more expensive services.
- Best effort with minimum guaranteed (BG), where the VNO is guaranteed with a minimum level of service. The request for data rates higher than the guaranteed level is served in the best effort manner, hence, the minimum guaranteed data rate is the one received during busy hours. In this case, although VNOs do not invest as much as former ones, they can still guarantee the minimum QoS to their subscribers. From the subscribers' viewpoint, the acceptable service (not as good as the previous ones) is offered with a relatively lower cost.
- Best effort (BE), in which the VNO is served in the pure best effort approach. In this case, operators and their subscribers may suffer from low QoS and resource starvation during busy hours, but the associated cost will be lower as well.

VRRM can be considered as a decision-making problem, under uncertainty for a dynamic environment; the decision is on the allocation of resources to the different services of VNOs, by considering the set of available

radio resources. In what follows, the VRRM problem is discussed in detail.

The first step is forming the set of radio resources containing the number of available radio resource units (RRUs) per RAT (e.g., time-slots in GSM, codes in UMTS, resource-blocks in LTE and channels in Wi-Fi), as follows:

$$s_t^{RRU} = \{N_{SRRU}^{RAT_1}, N_{SRRU}^{RAT_2}, \dots, N_{SRRU}^{RAT_{N_{RAT}}}\} \quad (1)$$

where:

- $s_t^{RRU}$ : the set of radio resources at  $t$ ,
- $N_{SRRU}^{RAT_i}$ : number of spare RRUs in the  $i$ th RAT,
- $N_{RAT}$ : number of RATs.

Mapping the set of radio resources onto the total network data rate is the next step. Since radio resources' performance is not deterministic, it is not possible to have a precise prediction of capacity, requiring the estimation of the total network data rate via a probability distribution, as a function of the available RRUs for further decisions,

$$s_t^{RRU} \xrightarrow{\text{mapping}} R_b^{CRRM} [Mbps] (t) \quad (2)$$

where:

- $R_b^{CRRM}$ : total network data rate.

The third step is the allocation of the available resources to the services of VNOs. Sets of policies are designed in this step, in order to assign a portion of the total network capacity to each service of each VNO. Meeting the guaranteed service levels and increasing the resources' usage efficiency are the primary objectives, but other goals, such as fairness, may be considered. The resource allocation mapping of the total network capacity onto different services' data rates is given by:

$$R_b^{CRRM} [Mbps] (t) \xrightarrow{\text{map}} \{R_{b_{ji}}^{Srv} [Mbps] (t) | j = 1, \dots, N_{VNO}, i = 1, \dots, N_{srv}\} \quad (3)$$

where:

- $R_{b_{ji}}^{Srv}$ : serving (allocated) data rate for service  $j$  of VNO  $i$ ;
- $N_{VNO}$ : number of VNOs;
- $N_{srv}$ : number of services.

Finally, monitoring the used resources and updating their status is the observation part of this decision-making

problem. It enables the manager to evaluate the accuracy of its decisions and to modify former ones. Updating changes in the set of radio resources helps the manager to cope with dynamic changes in the environment and in VNOs requirements. In summary, it can be claimed that resource management solutions generally have two main components; estimation of available resources and optimisation of their allocation, which are addressed next.

### 3.2 Estimation of available resources

The estimation of available resources, and their allocation to the different services, are the two key steps in VRRM procedures. Obtaining a probabilistic relationship in the form of a probability density function (PDF), between the set of available RRUs and network capacity is the goal in a first step, then, by having an estimation of network capacity, VRRM allocates a portion of this capacity to each service of each VNO.

Depending on various parameters, such as RAT, modulation and coding, the allocation of an RRU can provide different data rates to mobile terminals. However, the data rate of an RRU is generally a function of the signal-to-interference-plus-noise-ratio (SINR), [8, 9]. Since SINR is a random variable, given the channel characteristics from path loss, fading and mobility, among other things, one needs to express the data rate also as a random variable:

$$R_{bRAT_i[\text{Mbps}]}(\rho_{in}) \in \left[ 0, R_{bRAT_i[\text{Mbps}]}^{\max} \right] \quad (4)$$

where:

- $R_{bRAT_i}$ : data rate of an RRU from the  $i$ th RAT,
- $\rho_{in}$ : SINR,
- $R_{bRAT_i}^{\max}$ : maximum data rate of an RRU from the  $i$ th RAT.

Based on [8], the PDF of  $R_{bRAT_i}$  can be given as:

$$p_{R_b}(R_{bRAT_i[\text{Mbps}]}) = \frac{\frac{0.46}{a_p} \left( \sum_{k=1}^5 k a_k (R_{bRAT_i})^{k-1} \right) \exp\left(-\frac{0.46}{a_p} \sum_{k=0}^5 a_k (R_{bRAT_i})^k\right)}{\exp\left(-\frac{0.46}{a_p} a_0\right) - \exp\left(-\frac{0.46}{a_p} \sum_{k=0}^5 a_k (R_{bRAT_i}^{\max})^k\right)} \quad (5)$$

where:

- $a_p \geq 2$ : path loss exponent,
- $a_k$ : coefficients in a polynomial approximation of SINR, as a function of data rate in each RAT (presented in [8] for cellular networks, and in [10] for Wi-Fi).

The total data rate for a single RAT pool is

$$R_{b_{tot}}^{RAT_i} = \sum_{n=1}^{N_{RRU}^{RAT_i}} R_{b_n}^{RAT_i} \quad (6)$$

where:

- $N_{RRU}^{RAT_i}$ : number of RRUs of the  $i$ th RAT,
- $R_{b_{tot}}^{RAT_i}$ : data rate from the  $i$ th RAT pool,
- $R_{b_n}^{RAT_i}$ : data rate from the  $n$ th RRU of the  $i$ th RAT.

The PDF of a RAT's data rate is equal to the convolution of all RRUs' PDFs when the channels, and consequently, the data rates' random variables ( $R_{bi}$ ) are independent [26]. In the deployment of heterogeneous access networks, the resource pools of the different RATs can be aggregated under the supervision of CRRM. The total data rate aggregated from all RATs is then the summation of the total data rate from each individual:

$$R_{b[\text{Mbps}]}^{CRRM} = \sum_{i=1}^{N_{RAT}} R_{b_{tot}[\text{Mbps}]}^{RAT_i} \quad (7)$$

By having the number of available resources mapped onto probability functions, the VRRM has an estimation of the total network capacity. In the estimation procedure, the total network data rate is highly dependent on the channel quality at the mobile terminal, to which the radio resources are allocated. A higher network data rate can be achieved when the RRUs are allocated to mobile terminals with a high SINR. Thus, the allocation of the radio resources to mobile terminals with a low SINR leads to a lower network data rate. In a very low network capacity, VRRM may not be able to meet the minimum guaranteed requirements. The aforementioned estimation approach does not consider any assumption on the channel quality of the mobile terminals, this approach being referred to as the general (G) one. By adding assumptions about the mobile terminals' channel quality, three additional approaches for the estimation of network capacity are considered:

- Optimistic approach (OP): all RRUs are assigned to users with very good channel quality (i.e., high SINR), therefore, it is assumed that the data rate of each RRU satisfies:

$$0.5 R_{b[\text{Mbps}]}^{\max} \leq R_{b[\text{Mbps}]} \leq R_{b[\text{Mbps}]}^{\max} \quad (8)$$

- Realistic approach (RL): it is assumed that the RRUs of each RAT are divided into two equal groups, and

that the data rate of the RRU from each group is as follows:

$$0 \leq R_{b[\text{Mbps}]} < 0.5 R_{b[\text{Mbps}]}^{\max} \quad \text{Low SINR Group} \quad (9)$$

$$0.5 R_{b[\text{Mbps}]}^{\max} \leq R_{b[\text{Mbps}]} < R_{b[\text{Mbps}]}^{\max} \quad \text{High SINR Group} \quad (10)$$

- Pessimistic approach (PE): it is assumed that all the RRUs in the system are assigned to users with low SINR so that the boundaries are

$$0 \leq R_{b[\text{Mbps}]} \leq 0.5 R_{b[\text{Mbps}]}^{\max} \quad (11)$$

Equation (5) can be further developed for these special case studies, where data rate is bounded between high and low values, the conditional PDF of a single RRU in this case being calculated as follows [26]:

$$p_{R_b} \left( R_{b\text{RAT}_i[\text{Mbps}]} \mid R_{b\text{Low}[\text{Mbps}]} \leq R_{b\text{RAT}_i[\text{Mbps}]} \leq R_{b\text{High}[\text{Mbps}]}^{\max} \right) = \frac{\frac{0.46}{a_p} \left( \sum_{k=1}^5 k a_k (R_{b\text{RAT}_i})^{k-1} \right) \exp \left( -\frac{0.46}{a_p} \sum_{k=0}^5 a_k (R_{b\text{RAT}_i})^k \right)}{\exp \left( -\frac{0.46}{a} \sum_{k=0}^5 a_k (R_{b\text{Low}})^k \right) - \exp \left( -\frac{0.46}{a} \sum_{k=0}^5 a_k (R_{b\text{High}})^k \right)} \quad (12)$$

where:

- $R_{b\text{Low}}$ : low boundary for the RRU data rate;
- $R_{b\text{High}}$ : high boundary for the RRU data rate.

One should note that there is a relationship in between the various parameters related to the data rate:

$$0 \leq R_{b\text{Low}[\text{Mbps}]} \leq R_{b\text{High}[\text{Mbps}]} \leq R_{b[\text{Mbps}]}^{\max} \quad (13)$$

Furthermore, one should note that this approach is not scenario-dependent, i.e., it can be applied to any network/system fitting into the general network architecture previously presented, which encompasses all current cellular networks and WLANs, as well as to other approaches for the estimation of network capacity (just by considering the corresponding conditions).

### 3.3 Allocation of resources in cellular networks

In the next step, the services of the VNOs have to be granted with a portion of the network capacity. The allocation of resources has to be based on services' priority and SLAs, e.g., the services from conversation (e.g., VoIP) and streaming (e.g., video) classes are delay-sensitive, but they have almost constant data rates, thus, the allocation of data rates higher than the ones

contracted for these services does not increase the QoS, in contrast to interactive (e.g., FTP) and background (e.g., email) classes; as a consequence, operators offering the former set of services are not interested in allocating higher data rates.

The primary goal in the allocation procedure is to increase the total network data rate, while considering the priority of different services, subject to the constraints. On the ground of this fact, the objective function for VRRM is the total weighted network data rate, being expressed for cellular RATs as:

$$f_{R_b}^{\text{cell}}(\mathbf{R}_b^{\text{cell}}) = \sum_{i=1}^{N_{\text{VNO}}} \sum_{j=1}^{N_{\text{srv}}} W_{ji}^{\text{Srv}} R_{b_{ji}[\text{Mbps}]}^{\text{cell}} \quad (14)$$

where:

- $\mathbf{R}_b^{\text{cell}}$ : vector of serving data rates from cellular networks,
- $N_{\text{VNO}}$ : number of served VNOs by this VRRM,
- $N_{\text{srv}}$ : number of services for each VNO,
- $W_{ji}^{\text{Srv}}$ : weight of serving unit of data rate for service  $j$  of VNO  $i$  by VRRM, where  $W_{ji}^{\text{Srv}} \in [0, 1]$ .

The weights in (14) are used to prioritise the allocation of data rates to services, being a common practice to have the summation of all them equal to unit. The choice of these weights is based on the SLAs between VNOs and VRRM, and they can be modified depending on the agreed KPIs (key performance indicators) during runtime.

### 3.4 Allocation of resources in WLAN

It is desirable that the services with the higher serving weights receive data rates higher than the ones with the lower serving weights. The equivalent function for WLANs is

$$f_{R_b}^{\text{WLAN}}(\mathbf{R}_b^{\text{WLAN}}) = \sum_{i=1}^{N_{\text{VNO}}} \sum_{j=1}^{N_{\text{srv}}} \left( W_{ji}^{\text{Srv}} R_{b_{ji}[\text{Mbps}]}^{\text{WLAN}} + W^{\text{SRb}} \frac{\overline{R}_{b_j}}{R_b^{\max}} R_{b_{ji}[\text{Mbps}]}^{\text{WLAN}} \right) \quad (15)$$

where:

- $\mathbf{R}_b^{\text{WLAN}}$ : vector of serving data rates from APs,
- $W^{\text{SRb}}$ : weight for session average data rate, where  $W^{\text{SRb}} \in [0, 1]$ ,
- $\overline{R}_b^{\max}$ : maximum average data rate among all services,
- $\overline{R}_{b_j}$ : average data rate for service  $j$ .

In (15),  $W^{\text{SRb}}$  is introduced to give priority to services with a higher data rate per session. Assigning these

services to a Wi-Fi network reduces collision rates, leading to a higher network data rate. Obviously, assigning zero to this weight completely eliminates the average data rate effect (i.e., the effect of collision on network data rate) and converts the objective function of WLANs in (15) into the cellular one in (14). The aforementioned weight is chosen by the VRRM approach, based on the SLAs and the Wi-Fi and LTE coverage maps, in addition to applied network planning and load-balancing policies. It can also be subject to modifications during runtime, based on measurements and reports.

### 3.5 Fairness

In addition to increasing network data rate, a fair resource allocation is another objective in VRRM. On the one hand, the model is expected to allocate more resources to services with a higher serving weight, while on the other hand, services with a lower weight not being served at all or being served in very poor conditions are not acceptable. A fair allocation of resources is achieved when the deviation from the weighted average for all services is minimised:

$$\min_{R_{b_{ji}}^{Srv}} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \left| \frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} \right| \right\} \quad (16)$$

This concept is addressed as a fairness function, being written as:

$$f_{R_b}^{fr}(\mathbf{R}_b^f) = \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \left( R_{b_{ji}}^f [\text{Mbps}] \right) \quad (17)$$

where:

- $R_{b_{ji}}^f$ : the boundary for deviation data rate from the normalised average for service  $j$  of VNO  $i$ , defined as:

$$\begin{cases} \frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} - \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{N_{VNO} N_{srv} W_{ji}^{Srv}} \leq R_{b_{ji}}^f [\text{Mbps}] \\ -\frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{W_{ji}^{Srv}} + \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \frac{R_{b_{ji}}^{Srv} [\text{Mbps}]}{N_{VNO} N_{srv} W_{ji}^{Srv}} \leq R_{b_{ji}}^f [\text{Mbps}] \end{cases} \quad (18)$$

In order to better discuss the balance between these two objectives, i.e., the fairness and the total weighted network data rate, the boundaries of these two objectives have to be compared. The highest resource efficiency (i.e., the highest weighted data rate) is obtained when all resources are allocated to the service(s) with the highest

serving weight, hence, the maximum of the first objective can be written as:

$$\max_{R_{b_{ji}}^{Srv}} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^{Srv} R_{b_{ji}}^{Srv} [\text{Mbps}] \right\} = \max \left\{ W_{ji}^{Srv} \right\} R_b^{CRRM} [\text{Mbps}] \quad (19)$$

This means that, as the network capacity increases, the summation of the weighted data rate in (14) increases as well. In the same situation, the fairness objective function also reaches its maximum:

$$\max_{R_{b_{ji}}^{Srv}} \left\{ \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} R_{b_{ji}}^f [\text{Mbps}] \right\} = \frac{R_b^{CRRM} [\text{Mbps}]}{\max \left\{ W_{ji}^{Srv} \right\}} \quad (20)$$

Based on (19) and (20), the complete objective function for the management of virtual radio resources is defined as:

$$f_{R_b}^v(\mathbf{R}_b^{Srv}) = f_{R_b}^{\text{cell}}(\mathbf{R}_b^{\text{cell}}) + f_{R_b}^{\text{WLAN}}(\mathbf{R}_b^{\text{WLAN}}) - \alpha_f(W_f) f_{R_b}^f(\mathbf{R}_b^f) \quad (21)$$

where:

- $\mathbf{R}_b^f$ : vector of intermediate fairness variables,
- $\mathbf{R}_b^{Srv}$ : vector of serving data rates,
- $\alpha_f$ : fairness coefficient as a function of fairness weight:

$$\alpha_f(w_f) = \frac{W_f R_b^{CRRM} [\text{Mbps}]}{(1-w_f) R_b^{CRRM} [\text{Mbps}] N^{SmaxRb} + W_f R_b^{\text{max}} [\text{Mbps}]} \quad (22)$$

where:

- $N^{SmaxRb}$ : number of subscribers using the service with maximum data rate,

$$N^{SmaxRb} = \frac{R_b^{CRRM} [\text{Mbps}]}{R_b^{\text{max}} [\text{Mbps}]} \quad (23)$$

In (18) and (21), the allocated data rate for a specific service is defined as:

$$R_{b_{ji}}^{Srv} [\text{Mbps}] = R_{b_{ji}}^{\text{cell}} [\text{Mbps}] + R_{b_{ji}}^{\text{WLAN}} [\text{Mbps}] \quad (24)$$

### 3.6 Constraints

In addition, there are more constraints for VRRM to allocate data rates to various services, which should not be violated. The very fundamental constraint is the total network capacity estimated in the last section. The

summation of all assigned data rates to all services cannot be higher than the total estimated capacity of the network:

$$\sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} R_{b_{ji}}^{Srv} \leq R_b^{CRRM} \quad (25)$$

The data rate offered to GB and BG services imposes the next constraints. The data rate allocated to these services has to be higher than a minimum guaranteed level (for both GB and BG) and lower than the maximum guaranteed one (for GB only):

$$R_{b_{ji}}^{Min} \leq R_{b_{ji}}^{Srv} \leq R_{b_{ji}}^{Max} \quad (26)$$

where:

$R_{b_{ji}}^{Min}$ : minimum data rate for service  $j$  of VNO  $i$

$R_{b_{ji}}^{Max}$ : maximum data rate for service  $j$  of VNO  $i$ .

Based on this model, the objective function presented in (21) has to be optimised subject to constraints addressed in (18), (24), (25) and (26).

### 3.7 Resource allocation with violation

In the allocation process, there are situations where resources are not enough to meet all guaranteed capacity, and the allocation optimisation is no longer feasible. A simple approach in these cases is introduced in [9], which is to relax the constraints by the introduction of violation (also known as slack) variables. In case of VRRM, the relaxed constraint is given by:

$$R_{b_{ji}}^{Min} \leq R_{b_{ji}}^{Srv} + \Delta R_{b_{ji}}^v \quad (27)$$

where:

- $\Delta R_{b_{ji}}^v$ : non-negative violation variable for the minimum guaranteed data rate of service  $j$  of VNO  $i$ .

By introducing the violation parameter, the former infeasible optimisation problem turns into a feasible one. The optimal solution maximises the objective function and minimises the weighted average constraints violations. The weighted average constraints violation is defined as follows:

$$\Delta \overline{R}_b^v = \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^v \Delta R_{b_{ji}}^v \quad (28)$$

where:

- $\Delta \overline{R}_b^v$ : average constraint violation.
- $W_{ji}^v$ : weight of violating minimum guaranteed data rate of service  $j$  of VNO  $i$ , where  $W_{ji}^v \in [0, 1]$ .

The objective function presented in (21) has also to be changed as follows:

$$f_{R_b}^v(R_b) = f_{R_b}^{cell}(R_b^{cell}) + f_{R_b}^{WLAN}(R_b^{WLAN}) - \alpha_f(W_f) f_{R_b}^f(R_b^f) - f_{R_b}^{vi}(\Delta \overline{R}_b^v) \quad (29)$$

where  $f_{R_b}^{vi}$  is the constraint violation function, given by:

$$f_{R_b}^{vi}(\Delta \overline{R}_b^v) = \frac{R_b^{CRRM}}{R_b^{min}} \Delta \overline{R}_b^v \quad (30)$$

However, the definition of fairness in a congestion situation is not the same, i.e., in this case, fairness is to make sure that the weighted violation of all services is the same. Fairness constraints are changed as follows:

$$\begin{cases} W_{ji}^v \Delta R_{b_{ji}}^v - \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} W_{ji}^v \Delta R_{b_{ji}}^v \leq R_{b_{ji}}^f \\ -W_{ji}^v \Delta R_{b_{ji}}^v + \frac{1}{N_{VNO} N_{srv}} \sum_{i=1}^{N_{VNO}} \sum_{j=1}^{N_{srv}} \Delta R_{b_{ji}}^v \leq R_{b_{ji}}^f \end{cases} \quad (31)$$

The management of virtual radio resources is a complex optimisation problem, since network status and constraints vary in time. In [8, 9], it is proposed to divide the time axis into decision windows and to maximise the objective function in each of these intervals, independently. However, the output of partial VRRM may only be a local optimum and not the global one, since the effect of each decision on the network state and other dependencies are neglected. Nevertheless, partial VRRM is a simple solution, which can be used as the starting step and reference point.

The optimisation problem is solved by using MATLAB Linear Programming solver (i.e., the linprog function) [27], using the interior-point approach [28], which is a variant of Mehrotra's predictor-corrector algorithm [29], a primal-dual interior-point method. The termination tolerance on the function is chosen to be  $10^{-8}$ .

## 4 Scenario

The realistic scenario for evaluating the proposed model offers cellular networks coverage by means of a set of radio remote heads (RRHs) [14], supporting OFDMA (LTE-Advanced), CDMA (UMTS/HSPA+) and FDMA/TDMA (GSM/EDGE). VRRM is in charge of a service area, as described in Table 1:

- The OFDMA cells, with a 400 m radius, are the smallest ones; based on the 100 MHz LTE-Advanced feature, each cell has 500 RRUs to be assigned to traffic bearers.

**Table 1** Different RAT cell radius

| RAT   | Number of cells | Cell radius [km] | System | $N_{RRU}^{RAT}$ |
|-------|-----------------|------------------|--------|-----------------|
| OFDM  | 16.0            | 0.08             | Wi-Fi  | 40              |
| OFDMA | 16.0            | 0.4              | LTE    | 8 000           |
| CDMA  | 1.7             | 1.2              | UMTS   | 80              |
| TDMA  | 1.0             | 1.6              | GSM    | 75              |

- The configurations of CDMA cells are chosen according to UMTS/HSPA+, at 2.1 GHz, each cell with a 1.2 km radius and 3 carriers (each carrier has 16 codes); only 45 codes, out of all 48 in each cell, are assigned to users' traffic.
- The FDMA/TDMA cells are the biggest ones, with a 1.6 km radius, based on GSM900, each cell having 10 radio channels (each one has 8 timeslots), being assumed that 75 timeslots out of the total 80 available ones in each cell are used for users' traffic.

In addition to cellular networks, in this scenario Wi-Fi (OFDM) coverage is provided by means of IEEE802.11ac APs, configured to work with an 80 MHz channel bandwidth. It is assumed that each AP covers a cell with an 80 m radius, being facilitated with beamforming and MU-MIMO to support up to 8 spatial streams. Five radio channels are taken for the 80 MHz APs, following European Union regulations, [30]. In contrast to former RATs, APs use the same set of links for up- and down-link streams. In order to achieve coherency among various RATs, the total data rate of APs is equally divided between down- and uplinks, therefore, in Table 1, the number of RRUs in each Wi-Fi cell is indicated as half of the total number of available channels. This table also presents the maximum data rate for each RAT in down-link. It is also assumed that the APs are only deployed on the OFDMA BS, hence, not providing full coverage; however, the entire Wi-Fi capacity can be used for traffic offloading.

The path-loss exponent,  $\alpha_p$ , corresponding to signal propagation in the various environments, is considered to be 3.8 for regular urban environments, a value between 2 for free space and 5 for high attenuation dense urban ones [31].

Three VNOs are assumed to operate in this area, each one with a different SLA, i.e., GB, BG and BE. All of them offer the same set of services, as in Table 2. The serving weights in (14) and (15) are based on the general service classes: 0.4 for conversational (Con), 0.3 for streaming (Str), 0.2 for interactive (Int) and 0.1 for background (Bkg). Besides the usual human interaction services, one is also considering several machine-to-machine (M2M) applications, as this is one of the areas foreseen for large development of VNOs. In order not to

**Table 2** Network traffic mixture

| Service                              | Volume [%] | $W_{ji}^{Str}$ | $W_{ji}^{Int}$ | $\bar{R}_{b[jkps]}$ |        |
|--------------------------------------|------------|----------------|----------------|---------------------|--------|
| Video calling (ViC)                  | 59.7       | 95.4           | 0.3            | 0.27                | 5120.0 |
| Video streaming (ViS)                |            | 4.6            | 0.4            | 0.36                | 384.0  |
| File sharing (FTP)                   | 3.5        |                | 0.2            | 0.18                | 1024.0 |
| Web browsing (WWW)                   | 11.9       |                | 0.2            | 0.18                | 500.0  |
| Social networking (SoN)              | 14.4       |                | 0.2            | 0.18                | 384.0  |
| M2M Smart metres (MMM)               | 5.5        | 25             | 0.01           | 0.09                | 200.0  |
| e-Health (MME)                       |            | 25             | 0.2            | 0.18                | 200.0  |
| Intelligent transport services (MMI) |            | 25             | 0.4            | 0.36                | 200.0  |
| Surveillance (MMS)                   |            | 25             | 0.3            | 0.27                | 200.0  |
| Email (Ema)                          | 1          |                | 0.01           | 0.09                | 100.0  |
| Music streaming (MuS)                | 3          |                | 0.03           | 0.27                | 64.0   |
| VoIP (Vol)                           | 1          |                | 0.04           | 0.36                | 12.2   |

compromise the objective function for achieving fairness, the fairness weight,  $W^f$ , in (22) is considered to be unit, leading to a maximum fairness, while  $W^{SRb}$  in (15) is heuristically chosen to be 0.02.

Each VNO is assumed to have 500 subscribers, where each one requires the average data rate of 6.375 Mbps [32]. Hence, the contracted data rate,  $R_b^{Con}$ , for all operators is 3.11 Gbps, and each service receives a portion based on a volume percentage, in Table 2. In the second step, the number of subscribers for each VNO is swept from 300 (low load) up to 1400 (high load), in order to observe how VNOs capacity and their services are affected by this increase of load.

On the ground of each service data rate, the SLAs of these VNOs are defined as follows:

- VNO GB: the data rates allocated to services are guaranteed to be in the range of 50 to 100% of the corresponding service data rate.
- VNO BG: it has the best effort, with a minimum 25% of the service data rate guaranteed by the SLA.
- VNO BE: it has all services served in the best effort approach, without any guarantee.

## 5 Results

### 5.1 Network capacity

For the network capacity estimation, in addition to the general approach, the other three, i.e., pessimistic, realistic and optimistic, are also considered. The minimum and maximum data rates from each RAT, considering different approaches, are presented in Table 3. Equation (7) is used to obtain the PDF for the general approach and (12) for the other ones.

Using (5) and (12), the PDF and the cumulative distribution functions (CDFs) of the considered network

**Table 3** Minimum and maximum data rate of each RAT in different approaches

| RAT |  | OFDMA | CDMA | TDMA  | OFDM  |
|-----|--|-------|------|-------|-------|
| PE  | Min. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 0     | 0    | 0     | 0     |
|     | Max. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 2.73  | 1.68 | 0.002 | 25.35 |
| RL  | Min. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 1.37  | 0.84 | 0.001 | 12.67 |
|     | Max. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 4.10  | 2.52 | 0.003 | 38.08 |
| OP  | Min. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 2.73  | 1.68 | 0.002 | 25.35 |
|     | Max. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 5.47  | 3.36 | 0.004 | 50.78 |
| G   | Min. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 0     | 0    | 0     | 0     |
|     | Max. $R_{b, \text{tot}}^{\text{RAT}}$ [Gbps] | 5.47  | 3.36 | 0.004 | 50.78 |

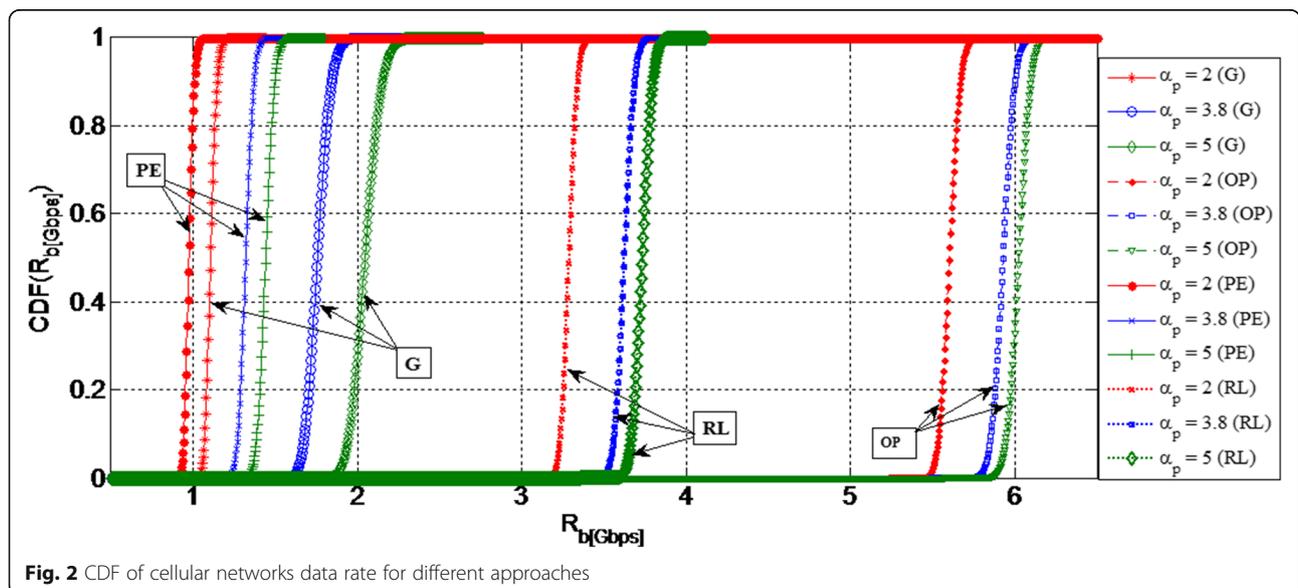
capacity are obtained. Figure 2 compares the differences in between the three approaches in conjunction with the general one. As expected, the lowest network capacity estimation is achieved by applying the pessimistic (PE) approach, since the assumption is the allocation of RRUs to mobile terminals with the lowest SINR. In this PE approach, the median capacity of the network for regular urban environments is 1.32 Gbps, while the general approach (G) leads to 1.76 Gbps, i.e., the former is 75.0% of this one; however, the realistic (RL) and the optimistic (OP) approaches provide a quite higher median network capacity of 3.60 and 5.93 Gbps, respectively, i.e., 2.0 and 3.4 times higher. Moreover, Fig. 2 also shows that a higher path-loss exponent yields a higher network capacity: the higher the path loss, the higher the attenuation, implying that interference is more attenuated than the signal, hence, increasing the carrier-interference-ratio, ultimately, yielding a higher capacity.

When adding the capacity offered by traffic offloading to Wi-Fi APs in regular urban environments (i.e., with a path-loss exponent of 3.8 for all systems, which is an approximation for a real scenario), one gets the CDF of the network capacity as plotted in Fig. 3. The comparison of Figs. 2 and 3 shows that the median values increase to 3.7 Gbps (1.8 times) in PE, 7.2 Gbps (3.0 times) in G, 19.5 Gbps (4.4 times) in RL and 35.3 Gbps (2.3 times) in OP, which is quite an increase in capacity. The total network capacity, according to Fig. 3, is 9.5 times higher in the OP approach than in the PE one, which without any doubt affects the allocation of resources to different services of the VNOs. The interdecile intervals range in between 2.14 Gbps in PE and 4.3 Gbps in G, with 2.93 Gbps and 3.52 Gbps for RL and OP, respectively, showing that the type of approach does not have a monotonic impact on the probability of obtaining a given capacity.

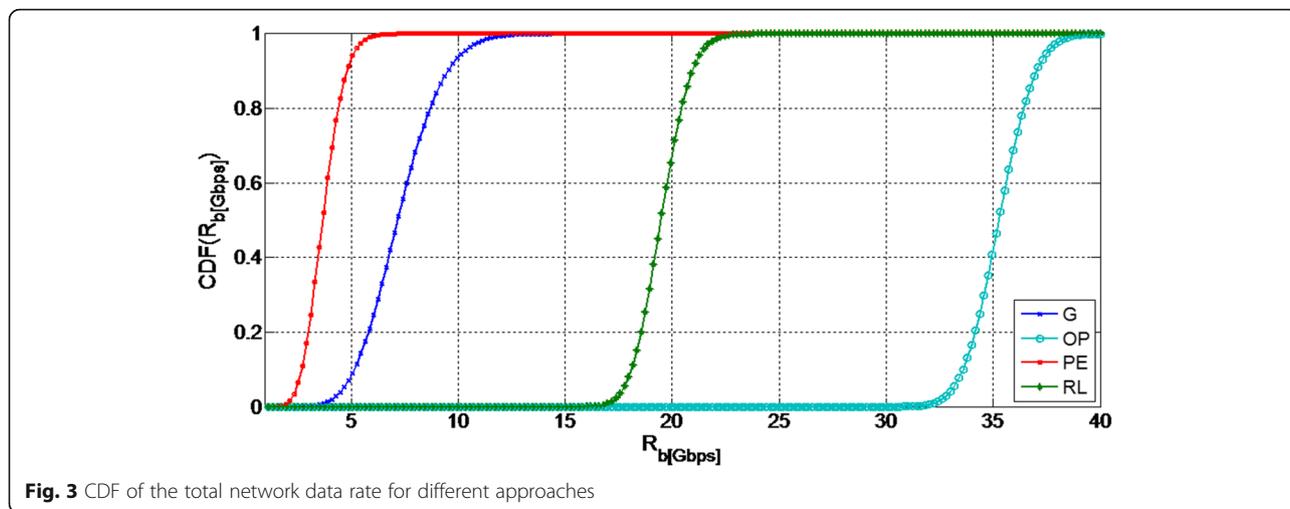
**5.2 Resource allocation with traffic offloading**

Figure 4 presents the data rates allocated to each of the services from cellular networks and WLANs in the G case, when there are 500 subscribers per VNO. As expected, conversational services (VoIP, Video Call and M2M/MMI), which are the ones with the highest serving weights, receive the highest data rates, streaming (music, M2M/MMS and video streaming) being placed second; the services of the background class (email and M2M/MMM) are the ones that are allocated the smallest portion of the available capacity.

The effect of the weight for session average data rate in WLAN,  $W^{\text{SRb}}$ , can be observed in the data rates balance between WLANs and cellular networks, e.g., the data rate allocated to video streaming (ViS) from



**Fig. 2** CDF of cellular networks data rate for different approaches



**Fig. 3** CDF of the total network data rate for different approaches

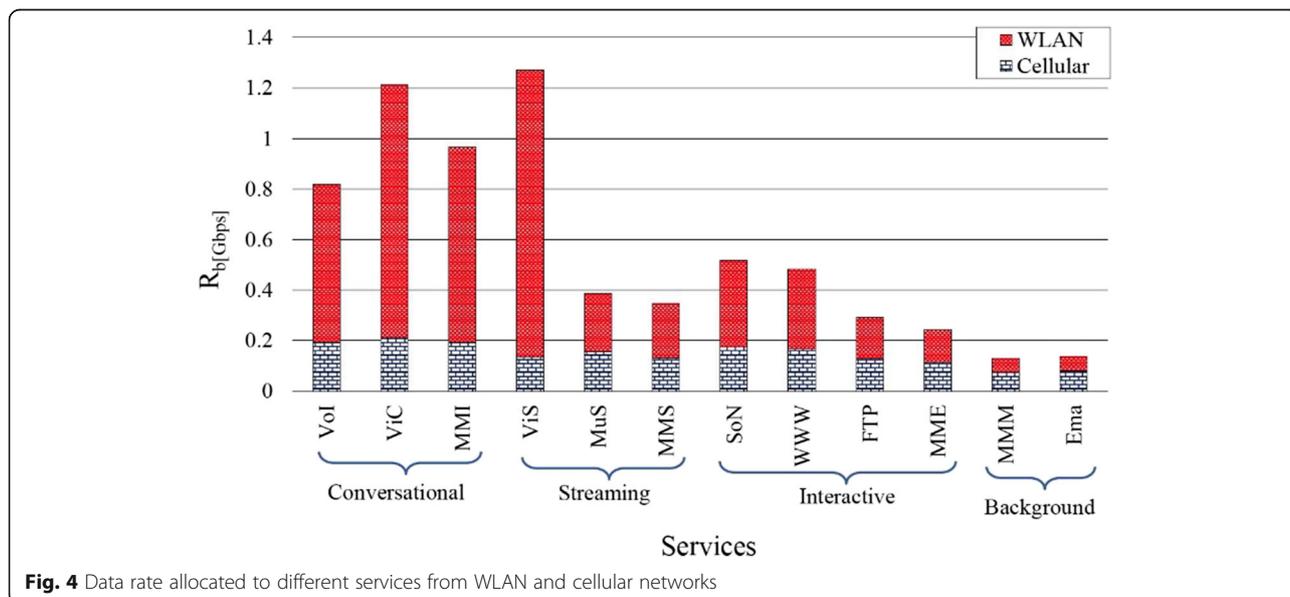
WLANs is 6.5 times higher than the one from cellular networks, while, in contrast, email, a service with a low average data rate, is allocated a higher data rate in cellular networks than in WLANs. However, VoIP is not following the same rule, since it has a relatively high serving weight, which overcomes the effect of the average data rate in (15), hence, being allocated a comparatively high data rate from both type of networks; the same phenomenon can be observed among M2M services, i.e., the ones with high serving weights, e.g., M2M/MMI, received relatively high capacity from WLANs comparing to the other ones.

Figure 5 demonstrates the allocation of virtual radio resources to the services of each VNO. Maximum guaranteed data rates are provided to almost all the services of

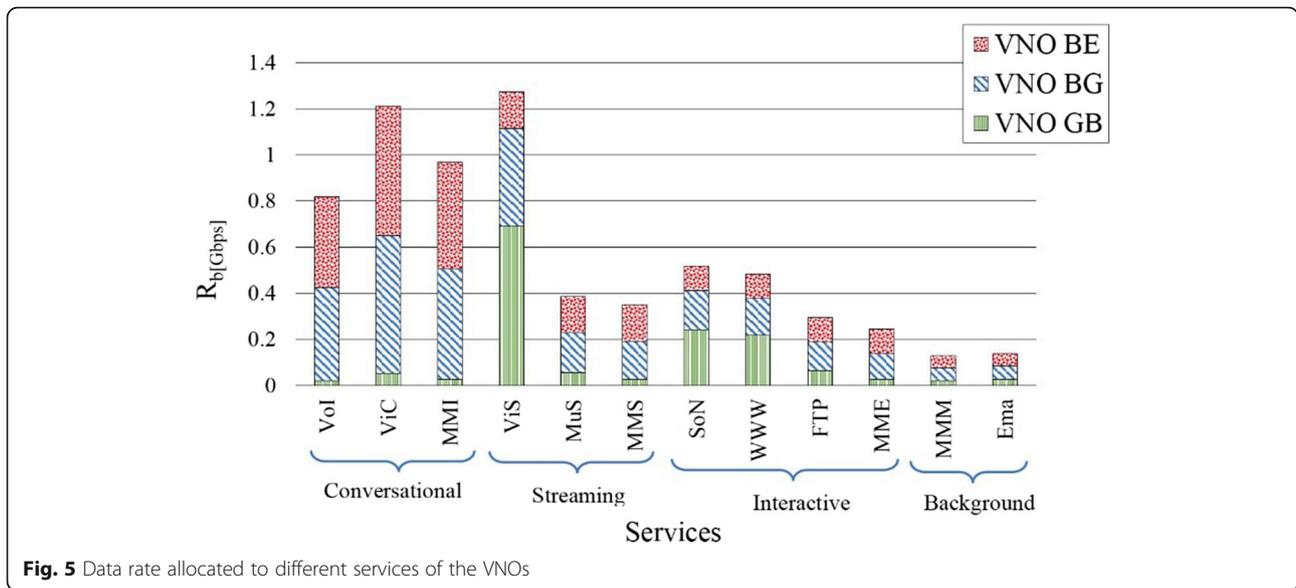
VNO GB, e.g., VoIP and music with the relative assigned data rate of 31.87 and 95.62 Mbps (which are the maximum requested). The upper boundary in the allocation of virtual resources to the services is the primary difference between the services of VNO GB and BG, in other words, while the data rates allocated to services of the guaranteed VNO are bounded by maximum guaranteed values, the services of VNO BG have no limitation. In contrast, the capacity offered to VNO BG in a resource shortage situation can be smaller than the VNO GB one.

**5.3 Resource allocation for different number of subscribers**

In this section, one analyses the performance of the proposed model under different network traffic loads. The number of subscribers is swept between 300 and 1 400



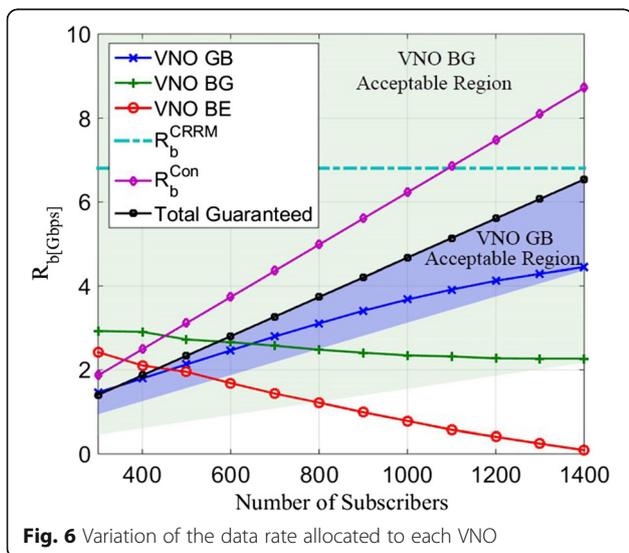
**Fig. 4** Data rate allocated to different services from WLAN and cellular networks



**Fig. 5** Data rate allocated to different services of the VNOs

per VNO (i.e., low and high loads). Figure 6 illustrates the distribution of the available virtual resources among VNOs, in addition to the total network capacity ( $R_b^{CRRM}$ ), the total minimum guaranteed and the contracted data rates ( $R_b^{Con}$ ). The contracted data rate for each VNO increases from 1.91 Gbps (low load) to 8.92 Gbps (high load).

The acceptable regions for VNOs GB and BG in the plot are shown by solid blue and light green colours. The total minimum guaranteed data rate, i.e., the summation of minimum guaranteed data rates of VNOs GB and BG, in the low load is 20.6% of the network capacity (i.e., 1.4 Gbps).

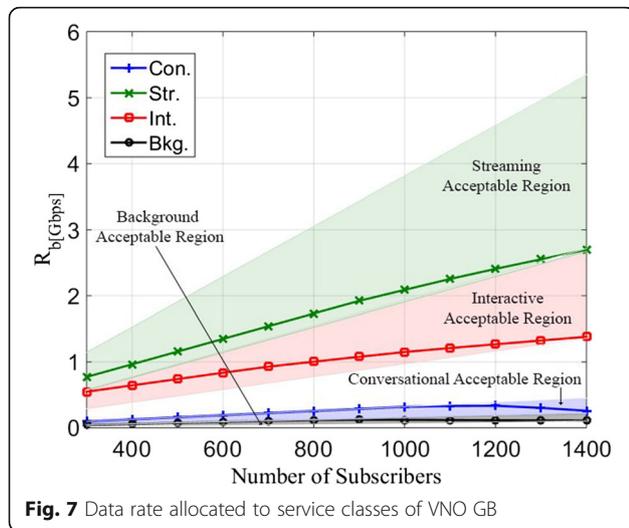


**Fig. 6** Variation of the data rate allocated to each VNO

Since network capacity is considerably higher than the minimum guaranteed data rates, best effort services are also served well; the allocation of 2.39 Gbps to VNO BE is the evidence to this claim. It is worth noting that the share of VNO BE is 35.1% of the whole network capacity, which is 1.6 times higher than VNO GB. The reason behind this observation is the maximum guaranteed data rate of the guaranteed services. Although the assigned portion of available resources to VNO GB is not as big as the other two VNOs, it is served up to its maximum satisfaction.

In contrast, VNO BG has a minimum guaranteed data rate, but the maximum received is 43.3% of the network capacity (2.94 Gbps). The guarantee data rates grow up to 6.53 Gbps (i.e., 96.1% of the whole available capacity) as the load increases. Obviously, the share of the best effort services in this situation considerably decreases. The allocated capacity to VNO BE reduces to only 65.6 Mbps, which is 0.9% of the total available capacity and 97.3% of its initial value. As shown in Fig. 6, the total network capacity is only enough for serving the contracted data rate of only one of the VNOs. In addition, the increase of the subscribers to 1400 makes the total minimum guaranteed data rate of the three VNOs equal to the total network capacity, which means that the data rates allocated to the services of VNO BE reach zero.

Furthermore, Fig. 7 illustrates the effect of demand variation on the allocation of data rates to the service classes of VNO GB: this VNO is a guaranteed one, therefore, each service class has a minimum and a maximum guaranteed data rate, presented in the figure with the solid colour. By increasing the number of subscribers, demand increases 4.7 times.

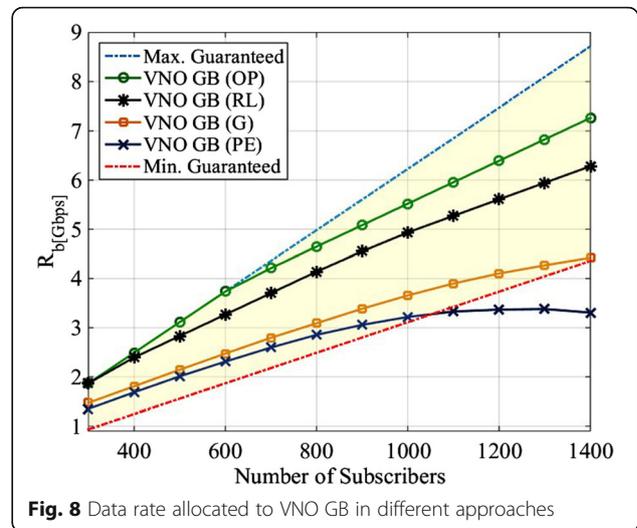


**Fig. 7** Data rate allocated to service classes of VNO GB

It can be seen that the streaming services are the ones with the highest volume, having the highest data rate; the minimum guaranteed data rate varies between 0.58 and 2.71 Gbps. The data rate allocated to this class in low load (when there are only 300 subscribers) is 67% of the maximum guaranteed one, but it reduces to the minimum one (i.e., 50% of the contracted data rate) for the maximum load case. The other service classes (i.e., interactive, conversational and background) are served according to capacity needs. It can be seen that, in the low load situation, the maximum guaranteed data rates are assigned, but as demand increases, data rates move towards the lower boundary. The interactive service class is a very good example for this behaviour: while it receives the maximum guaranteed data rate of 0.54 Gbps in low load, the allocated capacity for the high one is reduced from the maximum to 1.45 Gbps, the minimum guaranteed data rate. Considering the slope of allocated data rates in various services, the effect of serving weights and the service volume can be seen. Since the interactive class has a lower serving weight compared to the conversational one, it receives almost the minimum acceptable data rate with 1100 subscribers; in the same situation, conversational services are still provided by the highest acceptable data rate.

**5.4 The effect of channel quality effect on VRRM**

The effect of channel quality on the management of virtual radio resources by considering the three approaches (i.e., OP, RL and PE) is studied as well. Figure 8 presents the data rates allocated to VNO GB in conjunction with minimum and maximum guaranteed ones. As long as the data rates are in the acceptable region (shown by the solid colour), there is no violation of the SLAs and guaranteed data rates.

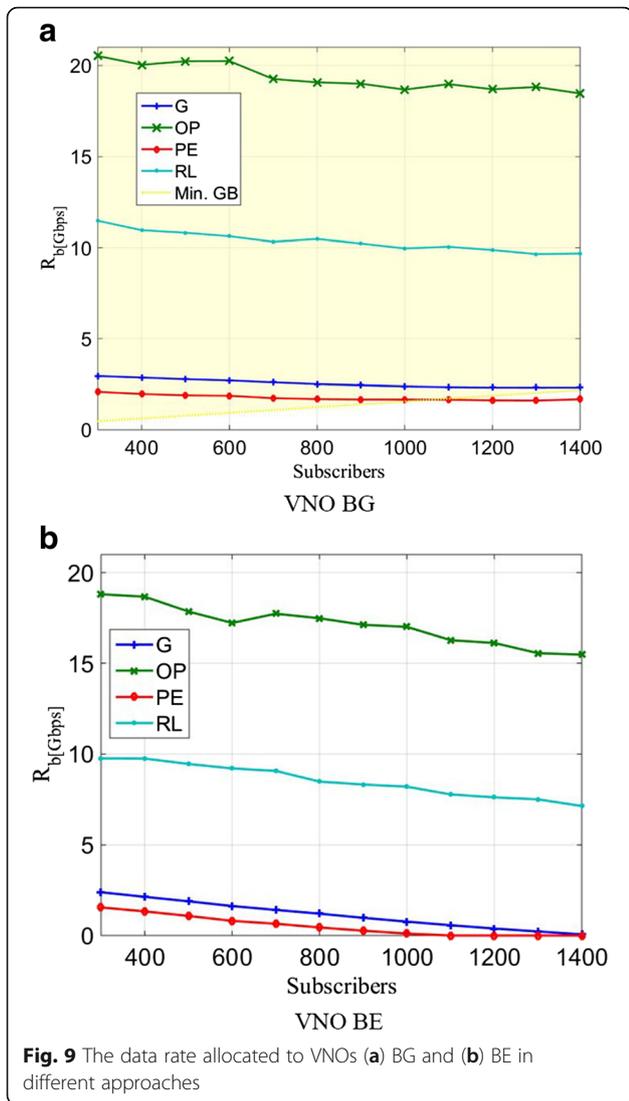


**Fig. 8** Data rate allocated to VNO GB in different approaches

Figure 8 shows that the maximum guaranteed data rate reaches 3.74 Gbps when there are 600 subscribers, being possible to allocate all of it in the OP approach, and that for 1400 subscribers it reaches 8.7 Gbps, but with only 83.3% of it being allocated to the OP approach. However, VNO GB faces the violation on the minimum guaranteed data rate in the PE approach, as the number of subscribers passes 1100: while at least 3.42 Gbps is required, only 97.3% of it is allocated to the VNO; this means that the network capacity in this approach is lower than the total minimum guaranteed data rates, and the resource management entity has to violate some of the minimum guaranteed data rates. The VNO requires at least 4.36 Gbps for the heavy load, the allocated data rate being 6.28 Gbps in RL and 4.42 Gbps in G, which are still enough to fulfil the SLA, but it goes down to 3.3 Gbps, i.e., 76.8% of the minimum required data rate, in PE, in clear violation of the SLA. This clearly shows the effect of SINR on resources usage efficiency and QoS offered to VNOS.

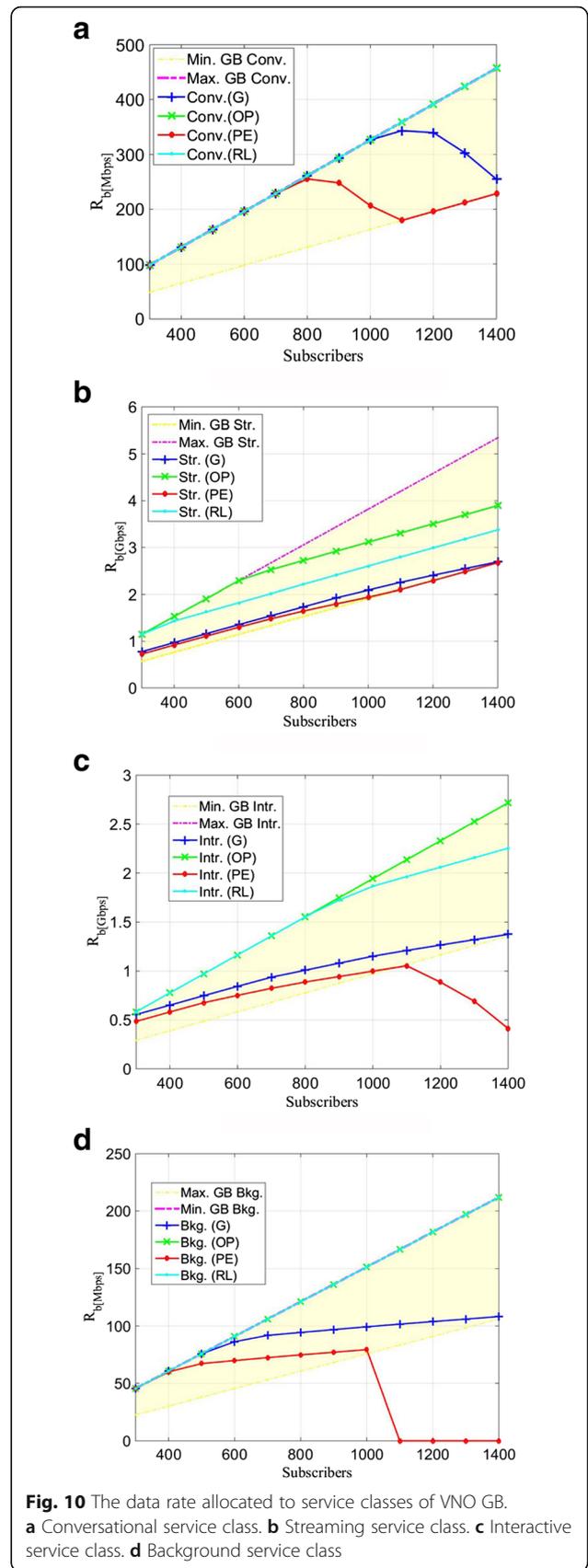
The data rates allocate to VNOS BG and BE are plotted in Fig. 9. Just as VNO GB, it can be seen that a high data rate is allocated to these VNOS in the OP and RL approaches (i.e., 18.46 and 9.68 Gbps). In these cases, the high SINR leads to the high network capacity, and the model is not only able to serve the minimum guaranteed data rates, but it can also serve acceptable data rates to the BE and BG VNOS. Consequently, VNOS BG and BE suffer more from resources shortage in the high load situations. The allocation of resources to VNO BE even stops for more than 1100 subscribers when the PE approach is considered.

Regarding the distribution of data rates allocated to the service classes of VNO GB, Fig. 9 illustrates the variation of the capacity assigned to each one in different



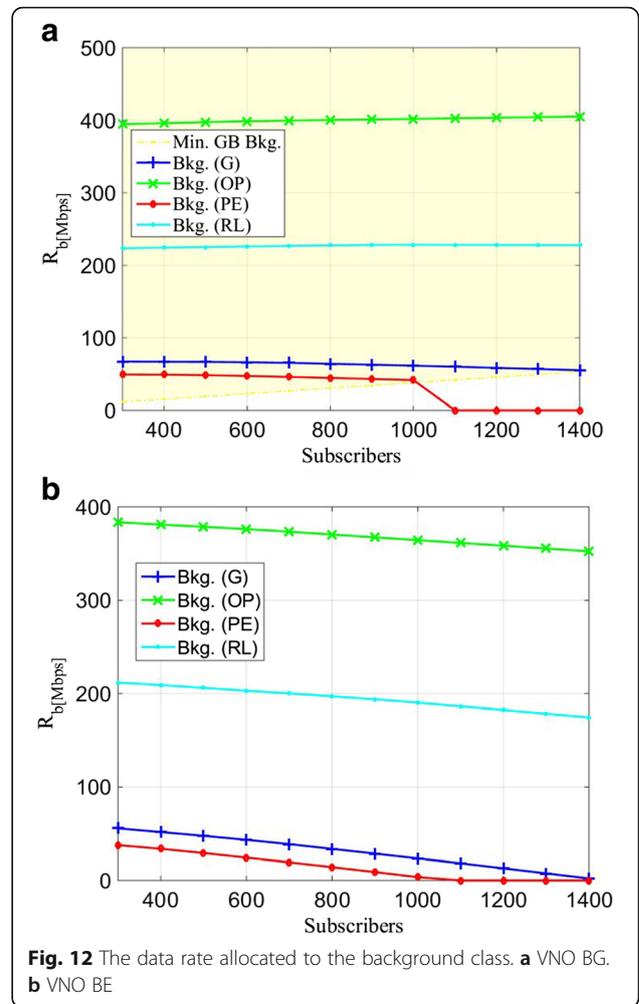
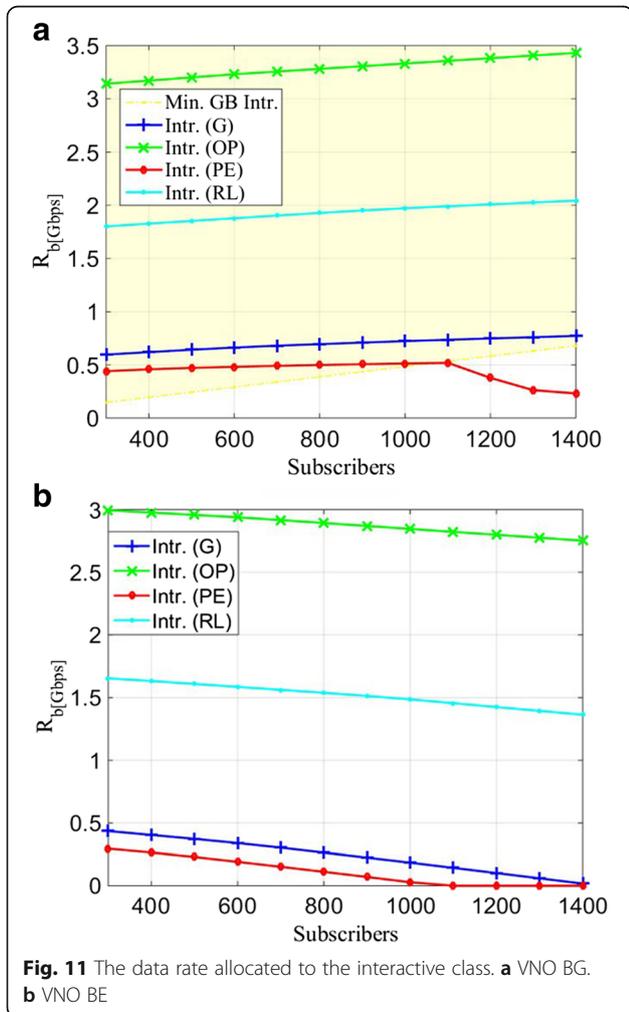
approaches. It can be seen that the conversational class (i.e., the class with the highest service weight) receives the maximum guaranteed data rate for the OP and RL approaches. The data rate allocated for the G and PE approaches is more than 50% of the contracted data rate. In the PE case, although for high-density situations the data rate decreases to a minimum guaranteed data rate, the services of this class never experience violation of the guaranteed data rate. Likewise, the streaming class is always served with a data rate higher than the minimum guaranteed. The maximum guaranteed data rate in heavy load reaches 5.34 Gbps and 72.9% in OP, 63.2% in RL, 50.4% in G and 50.0% in PE is allocated (Fig. 10).

For interactive and background classes, it is shown that they face violation of minimum guaranteed data rate in the PE approach. The violation situation in the background class is, to such an extent, that no capacity



is allocated to its services when there are more than 1100 subscribers per VNO. The data rate allocated to the interactive class reaches 15.1% of the contracted data rate in heavy load, while the minimum guaranteed is 50%.

For the sake of comparison, the data rate allocated to the interactive and background classes of VNOs BG and BE is shown respectively in Figs. 11 and 12. It can be seen that for VNO BG the situation is very much similar to VNO GB, the main difference being the high boundary of allocated data rate. VNO BG does not have a maximum guaranteed data rate or high boundary for allocation of data rates. Consequently, the services of this VNO are served by comparatively higher data rates comparing to VNO GB when a high network capacity is available, e.g., in the OP situation. As an example, consider the conversational class on both VNOs: for the OP approach with 400 subscribers, VNO GB is granted with 0.1 Gbps while VNO BG receives 1.3 Gbps; on the other hand, in the case of resource shortage, VNO BG receives



**Fig. 12** The data rate allocated to the background class. **a** VNO BG. **b** VNO BE

data rates lower than VNO GB, e.g., the share of interactive class of VNO BG when there are 1200 subscribers in the PE approach is only 0.387 Gbps while VNO GB is allocated 0.908 Gbps.

In conclusion, the effect of channel quality on the total available resource, and consequently on the performance of VRRM, is studied in this section. Through numeric results, one shows that the proposed model for managing virtual radio resources can serve different service classes of VNOs with different requirements, while offering an acceptable level of isolation. As evidence to this claim, one can consider services of the conversational class, i.e., services with the highest priority and serving weight, which are always allocated with a satisfactory amount of resources as the demands and the network capacity changes. Likewise, the minimum guaranteed data rates are offered to the relevant VNOs. Moreover, the prioritising of service classes offered by VRRM enables to serve the more important services, even when there are not enough resources.

## 6 Conclusions

A model for the management of virtual radio resources in a full heterogeneous network (i.e., a network with both cellular and WLANs) is proposed, which has two key components: the estimation of available resources and the allocation of resources. In the first step, the model maps the number of the available RRUs from different RATs onto the total network capacity by obtaining a probabilistic relationship. The model is able to consider multiple channel quality assumptions for the terminals through different estimation approaches. The allocation of resources to maximise the weighted data rate of the network based on the estimated network capacity is the next step. The serving weights in the objective function make possible to prioritise services. The resource allocation in a shortage of resources (i.e., when there are not enough resources to meet the minimum guaranteed data rates) tries to minimise the violation of the guaranteed data rates. In addition, the model also considers fairness among services.

Moreover, the performance of the proposed model is evaluated for a set of practical scenarios, and numeric results are achieved. It is shown that by adding an AP to each OFDMA cell, the network capacity increases up to 2.8 times. As a result of traffic offloading, the VRRM model is able to properly serve not only guaranteed services, but best effort ones are allocated with a relatively high data rate. Services with a higher serving weight, such as services of the conversational class, are provided with a higher data rate.

Furthermore, the results show that the network capacity increases from 0.9 Gbps in the pessimistic approach to 5.5 Gbps in the optimistic one. The effect of these capacity changes on the allocated data rates for different VNOs, and their service classes are presented through a series of plots. It is shown that when there is enough capacity, not only the guaranteed VNO is satisfied, but also best effort VNOs are well served. However, as the network capacity decreases due the channel quality (G and PE approaches), best effort VNOs are affected more than the guaranteed one. The same situation is shown at the service class level too. Conversational and streaming classes are the ones with the highest serving priority (i.e., serving weights), being generally allocated with data rates higher than the other two classes. When there is a shortage of resources, i.e., in the G and PE approaches, violations start relatively by background and interactive classes.

The model performance under different loads is also evaluated. The results confirm that the model is able to realise an acceptable level of isolation. When there is a shortage of radio resources, the model for resource management starts violating the guaranteed levels of services with lower serving weights, which are the background

and interactive services. As evidence to this claim, the VNO GB, and particularly its conversational class, can be considered where the requested service quality regardless of network situation is offered. In addition, the flexibility of the applied model makes the equilibrium among different services or different VNOs possible.

In conclusion, it is shown that our model achieves the desired goals: (a) on-demand wireless capacity offering by serving three VNOs with different SLAs; (b) isolation by considering changes in the networks into almost one tenth of its original value (from OP to PE), in addition to changing the demand from 300 UE per VNO to 1 400; (c) element abstraction and multi-RAT support by providing wireless connectivity using both cellular networks and WLANs, while the VNOs do not have to deal with the details. In the future, the aforementioned concept of virtual radio resources and the proposed model will be implemented in realistic test-beds. In addition, the modelling of service demands and user behaviour in full heterogeneous networks will be taken in the next extension of the model.

### Acknowledgements

The research leading to these results was partially funded by the European Union's Seventh Framework Programme Mobile Cloud Networking project (FP7-ICT-318109).

### Authors' contributions

The authors have contributed jointly in all the parts for preparing this manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 September 2016 Accepted: 15 March 2017

Published online: 19 April 2017

### References

1. Cisco Systems, Global mobile data traffic forecast update, 2012–2017, in *from Visual Network Index (VNI) White Paper* (Cisco Systems, CA, 2013)
2. L Kyunghan, L Joohyun, Y Yung, R Injong, C Song, Mobile data offloading: how much can Wi-Fi deliver? *IEEE/ACM Trans Networking* **21**, 536–550 (2013)
3. A Balasubramanian, R Mahajan, A Venkataramani, Augmenting mobile 3G using Wi-Fi, in *8th international conference on Mobile, Systems, Applications, and Services* (ACM, San Francisco, 2010), pp. 209–222
4. L Joohyun, Y Yung, C Song, J Youngmi, Economics of Wi-Fi offloading: trading delay for cellular capacity. *IEEE Trans Wirel Commun* **13**, 1540–1554 (2014)
5. A Kliks, N Dimitriou, A Zalonis, O Holland, Wi-Fi traffic offloading for energy saving, in *20th International Conference on Telecommunications (IEEE, Casablanca, 2013)*, pp. 1–5
6. H Guan, T Kolding, P Merz, Discovery of cloud-RAN, in *NSN Cloud-RAN Workshop* (Nokia Siemens Networks, Beijing, 2010)
7. M Chiosi, D Clarke, P Willis, A Reid, J Feger, M Bugenhagen, W Khan, M Fargano, C Cui, H Deng, J Benitez, U Michel, H Damker, K Ogaki, T Matsuzaki, *Network function virtualisation: an introduction, benefits, enabler, challenges, and call for action* (European Telecommunications Standards Institute, Darmstadt, 2012)
8. S Khatibi, LM Correia, Modelling of virtual radio resource management for cellular heterogeneous access networks, in *IEEE 25th Annual International*

*Symposium on Personal, Indoor, and Mobile Radio Communications* (Washington, 2014)

9. S Khatibi, LM Correia, A model for virtual radio resource management in virtual RANs. *EURASIP J Wirel Commun Netw* **68**, 2015 (2015)
10. S Khatibi, LM Correia, Modelling virtual radio resource management with traffic offloading support, in *IEEE 24th European Conference on Networks and Communications* (IEEE, Paris, 2015)
11. S Khatibi, LM Correia, The effect of channel quality on virtual radio resource management, in *IEEE 82nd Vehicular Technology Conference* (IEEE, Boston, 2015)
12. JA Village, KP Worrall, DI Crawford, 3G shared infrastructure, in *3rd International Conference on 3G Mobile Communication Technologies* (IEEE, London, 2002), pp. 10–16
13. X Costa-Perez, J Swetina, G Tao, R Mahindra, S Rangarajan, Radio access network virtualization for future mobile carrier networks. *IEEE Commun Mag* **51**, 27–35 (2013)
14. L Samson, *LTE network sharing: some operational and management aspects. ITU cross regional seminar for CIS, ASP and EUR regions on broadband access (fixed, wireless including mobile)*, Chisinau, Moldova, 2011
15. T Frisanco, P Tafertshofer, P Lurin, R Ang, *Infrastructure sharing and shared operations for mobile network operators: from a deployment and operations view* (IEEE International Conference on Communications, Beijing, 2008), pp. 2193–2200
16. R Mahindra, MA Khojastepour, Z Honghai, S Rangarajan, Radio access network sharing in cellular networks, in *21st IEEE International Conference on Network Protocols* (IEEE, Göttingen, 2013), pp. 1–10
17. R Friedrich, S Pattheeuws, D Trimmel, H Geerdes, *Sharing mobile networks—why the pros outweigh the cons* (Booz & Company Inc., New York, 2012)
18. A Khan, W Kellerer, K Kozu, M Yabusaki, Network sharing in the next mobile network: TCO reduction, management flexibility, and operational independence. *IEEE Commun Mag* **49**, 134–142 (2011)
19. 3GPP, *Universal Mobile Telecommunications System (UMTS); LTE; Network sharing; architecture and functional description* (European Telecommunications Standards Institute (ETSI), France, 2013)
20. YK Song, H Zo, AP Ciganek, Multi-criteria evaluation of mobile network sharing policies in Korea. *ETRI J* **36**, 572–580 (2014)
21. Y Zaki, Z Liang, C Goerg, A Timm-Giel, LTE wireless virtualization and spectrum management, in *3rd Joint IFIP Wireless and Mobile Networking Conference* (IEEE, Budapest, 2010), pp. 1–6
22. Y Zaki, L Zhao, C Goerg, A Timm-Giel, LTE mobile network virtualization. *Mob Netw Appl* **16**, 424–432 (2011)
23. Z Liang, L Ming, Y Zaki, A Timm-Giel, C Gorg, LTE virtualization: from theoretical gain to practical solution, in *23rd International Teletraffic Congress* (IEEE, San Francisco, 2011), pp. 71–78
24. J Pérez-Romero, X Gelabert, O Sallent, Radio resource management for heterogeneous wireless access, in *Heterogeneous wireless access networks*, ed. by E Hossain (Springer US, New York, 2009), pp. 1–33
25. J Carapinha, C Parada (eds.), *Reference scenarios and technical system requirements definition* (Mobile Cloud Networking Project, 2013)
26. A Papoulis, SU Pillai, *Probability, random variables, and stochastic processes* (McGraw-Hill, NY, 2002)
27. *MMATLAB and Statistics Toolbox*, The MathWorks, Inc., Natick, Massachusetts, United States, 2015
28. Y Zhang, Solving large-scale linear programs by interior-point methods under the MATLAB environment, in *Technical Report TR96-01* (Department of Mathematics and Statistics, University of Maryland, Baltimore, 1995)
29. S Mehrotra, On the implementation of a primal-dual interior point method. *Soc Ind Appl Math J Optim* **2**, 575–601 (1992)
30. O Bejarano, EW Knightly, P Minyoung, IEEE 802.11ac: from channelization to multi-user MIMO. *IEEE Commun Mag* **51**, 84–90 (2013)
31. E Damosso, LM Correia (eds.), *COST 231 Final report—digital mobile radio: evolution towards future generation systems* (COST Secretariat, European Commission, Brussels, 1999)
32. Cisco Systems, *The Zettabyte Era - Trends and Analysis*, in *from Visual Network Index (VNI) White Paper* (Cisco Systems, CA, 2013)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)