A Capacity Broker Architecture and Framework for Multi-tenant support in LTE-A Networks

Georgia Tseliou*, Konstantinos Samdanis†, Ferran Adelantado*, Xavier Costa-Perez† and Christos Verikoukis‡
* Open University of Catalonia (UOC), Barcelona, Spain
† NEC Europe Ltd., Heidelberg, Germany
‡ Telecommunications Technological Centre of Catalonia (CTTC), Castelldefels, Spain
{gtseliou, ferranadelantado}@uoc.edu, {Konstantinos.Samdanis, Xavier.Costa}@neclab.eu, cveri@cttc.es

***Abstract*—This paper introduces the concept of capacity broker in the 3GPP network management architecture, reusing the existing interfaces with a minimum set of enhancements. This element allocates resources towards Mobile Virtual Network Operators (MVNOs) upon request in flexible, short-term time scale. In addition, we introduce a framework for slicing the capacity, considering two types of traffic: guaranteed where resources are locked for explicit use of a particular MVNO and Best-Effort (BE) where resources are pooled and shared by all participants. To accomplish this we take a two-step approach: (i) analyze short-term forecasting algorithms that can facilitate the capacity broker with accurate information regarding non-uniformities of the expected traffic and (ii) slice the available resources into these two types of traffic classes based on the result of the forecasting, considering the Confidence Degree (CD) of the prediction.**

I. Introduction

Mobile communications are entering a new era with the popularity of smart-phone devices, which gave rise to a plethora of new services with ever-increasing resource demands. This encouraged the early adoption of LTE, but lately operators' revenues cannot keep pace, considering the cost needed to operate and upgrade their network infrastructure. Network sharing can recover up to 20% of operational costs for typical European operators and significantly reduce capital expenditures in developing countries e.g., up to 70% in India [1]. These economic considerations drive the adoption of network sharing alongside network operational observations, e.g., 50% of the sites carry traffic that yields less than 10% of the revenue [2]. These under-utilized resources can be shared, providing another revenue means for Mobile Network Operators (MNOs).

The first generation of network sharing concentrated on passive and active solutions based on long-term contractual agreements. This paper takes a step further considering on-demand network sharing, where resources are acquired in the scale of minutes, while allocations are configured via signaling. To assure synchronization in resource sharing for such short-time scales, while satisfying Service Level Agreements (SLAs), this paper introduces a centralized resource management entity referred to as capacity broker within the MNO infrastructure. The role of the capacity broker is to assist the MNO owning a shared RAN, to take full advantage out of the unused capacity. To accomplish an efficient resource allocation, the capacity broker needs to have (i) a global view of the network resource utilization, and (ii) knowledge of the expected traffic volumes, a challenging task due to lack of periodicity in short-term scales.

The contributions of this paper concentrate on facilitating network sharing integrating the capacity broker in the 3GPP network management architecture [3] and reusing the existing interfaces with a minimum set of enhancements. In addition, a resource management framework for on-demand capacity allocation is introduced considering two types of traffic: (i) Guaranteed with resources locked for explicit use of a Mobile Virtual Network Operator (MVNO) and (ii) Best-Effort (BE) in where resources are pooled and shared by all participants. To accomplish this we take a two-step approach: (i) analyze short-term forecasting algorithms that facilitate the capacity broker with accurate information regarding the expected traffic and (ii) identify the limit to slice the available resources into these two types of traffic classes, depending on the forecasting and its associated Confidence Degree (CD).

The remaining of the paper is organized as follows: Section II presents the related work, while Section III presents how the capacity broker is integrated in 3GPP LTE-A. Section IV elaborates the Multi-tenant resource Slicing (MuSli) framework. Section V analyzes the simulation set-up and the evaluation results. Finally, Section VI provides the conclusions.

II. State of the Art

The initial adoption of network sharing in the 3GPP concentrated on passive solutions, wherein MNOs share base station sites, masts, antennas, etc. Active sharing that followed enabled operators to share network resources for long term periods according to contractual agreements. For active network sharing 3GPP has specified two architectures in [4]: (i) the Multi-Operator Core Network (MOCN) and (ii) the Gateway Core Network (GWCN). In the former, each operator is sharing eNBs connected to core network elements belonging to each MNO using a separate S1 interface. In the latter, operators share additionally the Mobility Management Entity (MME). Our capacity broker proposal is compatible with both 3GPP network sharing architectures, while introducing on-demand resource allocation via the means of signaling extensions of the 3GPP network sharing management [3].

In [5], the authors introduce a preliminary approach for virtualizing an eNB. They detail notion of hypervisor that performs resource sharing among MNOs considering radio conditions, sharing contracts and traffic load. In advancing the basic eNB virtualization, [6] introduces the concept of Network Virtualization Substrate (NVS) that operates closely to the MAC scheduler. A tailored mixture of reserved and shared resources with respect to the NVS component is proposed in [7] with the objective to flexibly allocate shared resources modifying the MAC scheduler to reflect MVNO's traffic demands and priorities. In this work we adopt such NVS two-step process, but instead of concentrating on the MAC scheduler for performing resource differentiation, we introduce the capacity broker to provide different resource slices based on the expected traffic volume.

The notion of capacity broker is introduced in 3GPP from a business perspective in [8], considering the case where the capacity broker resides inside the network owner. A study adopting such capacity broker paradigm in LTE is detailed in [9] regarding a range of capacity and spectrum sharing options. Unlike such an approach that introduces a new control plane interface to coordinate the sharing agreements, our proposal is backwards compatible with the existing 3GPP network management architecture, reusing current interfaces, while introducing a minimum set of enhancements. In particular, the proposed capacity broker extends the deployment design study in [10], adopting the centralized alternative, while considering a framework that allocates resources into two different serving classes based on short-term load forecasting.

The accuracy of short-term load forecasts can significantly affect the capacity broker decisions for resource slicing. A wide range of solutions for short-term load forecasting have been reported in the literature [11], which can be distinguished in two categories. The first one employs the characteristics of traffic loads, such as spatial and temporal relevance or self-similarity [12]. The second category employs techniques, such as exponential smoothing to study the intrinsic dimensionality [13], Kalman filtering to capture the evolution of traffic loads [14] or modern signal processing techniques such as compressive sensing [15]. In this paper we investigate which of the aforementioned methods fits best the capacity broker paradigm and we provide a set of enhancements, to compensate the lack of periodicity and non-uniformities of a short-term prediction.

III. 3GPP Network Sharing Architecture

An overview of the 3GPP network sharing management architecture [3] considering the capacity broker is depicted in Fig. 1. The Master Operator - Network Manager (MO-NM) monitors and controls the shared network via the master operator shared RAN domain manager or network element manager on shared eNBs using the Inf-N interface. The capacity broker is placed on the MO-NM to facilitate capacity allocation and admission control to MVNOs by interacting with the Sharing Operator Network Manager (SO-NM) via Type 5 Interface. The main benefit of co-locating the capacity broker at the MO-NM, is the rapid access to network monitoring information such as UL/DL load and performance measurements, as well as access to network planning information, including the network topology.

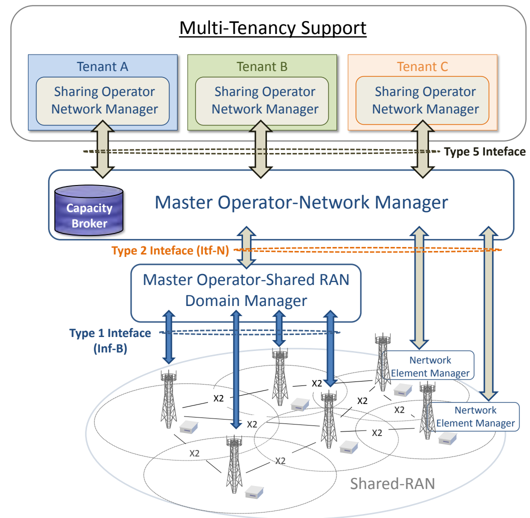

Fig. 1. Capacity Broker in 3GPP Network Sharing Management Architecture.

Our proposal requires extensions to Inf-N, to accommodate the PLMN-id, resource allocation e.g. Resource Blocks (RBs), start time and time duration. Inf-N should support resource measurements and performance monitoring per MVNO and to accomplish this we introduce the PLMD-id within each corresponding packet. The Inf-B, which is the interface between the master operator-shared RAN domain manager and the eNB should also support the same set of extensions as Inf-N. Type 5 interface between the MO-NM and the SO-NM it is established typically upon an agreement. Type 5 interface should be extended to accommodate dynamic resource requests from MVNOs, containing the PLMN id, the list of the MNO cells involved, the required resources (e.g. RBs), start time and time duration. Similarly to Inf-N, Type 5 interface should introduce the PLMN-id in each corresponding resource measurement and performance monitoring packet. For the portion of pooled resources, monitoring information should be shared among all participant's SO-NM systems.

IV. Multi-tenant Resource Slicing Framework

The main idea behind Multi-tenant Slicing (MuSli) is illustrated in Fig. 2, which shows how different types of resources (i.e., for guaranteed capacity and BE) are sliced, depending on the forecasted traffic volume (i.e. solid blue curve) at the time when a MVNO request reaches the capacity broker. For guaranteed capacity requests, the capacity broker allocates an isolated slice of resources locked for explicit use by the MVNO for the requested time. On the other hand, BE traffic uses the resources that cannot be allocated for guaranteed traffic, i.e., RBs that are not predicted as free for the entire duration of a request but are partially used by the MNO. These resources are offered as a resource pool that can be shared by all participant MVNOs including the MNO. In this way, more efficient resource utilization can be achieved. In our scenario, the incoming requests are divided into the two aforementioned

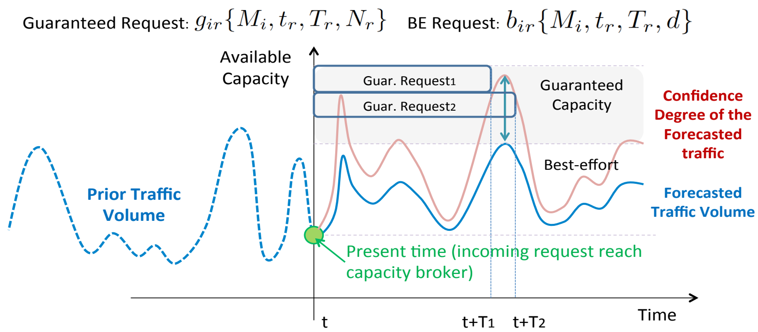

Fig. 2. Multi-tenant Slicing (MuSli) into different traffic classes considering the forecasted traffic.

classes according to their service requirements. CBR requests need a specific amount of RBs, while BE requests obtain an unspecified variable bit rate, depending on the portion of the resource pool and on the amount of competing requests. The CD of the forecasted traffic volume (i.e., red curve in Fig. 2), influences the resource provision accuracy. In principle, there is a trade-off between service quality assurance and number of served requests. On the safe side, the use of the upper bound CD of the predicted traffic to accommodate guaranteed requests, ensures service quality but results into accommodating fewer requests. Therefore, an MNO can tune the CD of the forecasting, to treat the requests, according to the desired level of certainty.

A. Capacity Forecasting

Due to the flexible, on-demand nature of MVNO's requests, we study methods to predict capacity in short-term basis. Short-term traffic is challenging to forecast, due to stronger variations that may be observed and lack of periodicity.

1) Improving short-term forecasting: To compensate for the inability to appropriately perceive the non-uniformities of user traffic, we analyze the traffic sub-components of the forecasting. Specifically, we extract periodical patterns using the Fast Fourier Transform (FFT) and analyze the predicted traffic into the time domain by transforming it in n distinct components into the frequency domain. To identify the spectral peaks in frequency we compute $X(m\omega_s) = \sum_{k=0}^{N-1} x_k e^{-jm\omega_s k}$, $\omega_s = \frac{2\pi}{N}$, where N is the number of the frequency samples. We introduce a vector x of time domain traffic samples, and the expression returns a vector X consisting of samples $X(m\omega_s)$, $m = 0, 1, \cdots, N-1$.

Then, we use the Inverse Fast Fourier Transform (IFFT) to transform each periodical component into the equivalent time series, which provides as a more regular traffic pattern compared to the initial prediction. For every component in the frequency domain, $k = 1, .., C$ (where C is the total number of components), we calculate $x_k = \frac{1}{N} \cdot \sum_{m=0}^{N-1} X_N(m\omega_s) \cdot e^{jmw_s k}$. In this way we manage to decompose the prior traffic, into k periodic components that can be predicted more accurately compared to the initial non-uniform traffic.

2) Forecasting Confidence Degree: The estimated value of utilized RBs at time t, can be described as a function of the actual value (i.e., $U(t)$) and an error $\epsilon(t)$, i.e., $\hat{U}(t) = U(t) + \epsilon(t)$. If $\hat{U}(t) > U(t)$ holds, then the forecasting results in overestimated RBs (i.e., $\epsilon(t) > 0$). Otherwise, when $\hat{U}(t) < U(t)$, the forecasting gives underestimated RBs (i.e., $\epsilon(t) < 0$). For using the forecasting results in MuSli:

(i) We estimate $\epsilon(t)$, that varies according to the predicted load. In the 1000 experimental forecasts we conducted in differently loaded cells, we observe that $|\epsilon(t)|$ can be approximated by an exponential function, $\epsilon(t) \approx \exp(bt) + c$, where $b, c \in \mathbb{R}$.

(ii) We represented the experimentally estimated capacity values (i.e., $\hat{U}(t)$) in three different CD$_a$s by using the student's distribution [16], over the average value of the predicted capacity across 1000 simulations. We make the assumption that the sample size of the estimated capacity is large and its standard deviation is unknown. The desired CD$_a$ is $(1-a) \times 100\%$. For $a = 1$ $CD_1 = 99\%$, for $a = 5$ $CD_5 = 95\%$ etc.

B. System Model

Let us denote as $M = \{M_0, M_1, \cdots, M_V\}$ the set of MVNOs sharing the RAN, where $i = 0$ refers to the mobile network owner and $i = 1, \cdots, V$ to participant MVNOs. Requests from the former are prioritized over the MVNOs' ones. As depicted in Fig. 2, every M_i generates a set of $\mathcal{G}_{M_i}$ guaranteed requests in the format $g_{ir}\{M_i, t_r, T_r, N_r\}$, where $g_{ir} \in \mathcal{G}_{M_i}$, t_r is its starting time, T_r its duration and N_r the number of requested RBs. Similarly every M_i generates a set of BE requests (i.e., $\mathcal{B}_{M_i}$) in the format $b_{ir}\{M_i, t_r, T_r, d\}$, where $b_{ir} \in \mathcal{B}_{M_i}$ and d is the demanded data rate (bps). The total load of all guarantee requests in time t is $|\mathcal{M}| \times |\mathcal{G}_{M_i}| \times N_r$ and the corresponding of BE requests equals $|\mathcal{M}| \times |\mathcal{B}_{M_i}| \times d$.

Let us denote the total capacity of a cell as C, which is fixed in terms of RBs. Therefore, the available capacity for MuSli (described in Algorithm 1) is defined as $A(t) = C - U(t)$. The estimated available capacity for a cell at time t, is equal to $\hat{A}(t) = C - \hat{U}(t) = C - U(t) - \epsilon(t) = A(t) - \epsilon(t)$. In the following, we define the necessary parameters for treating a request. We define as γ the average amount of resources unavailable along T_r:

$$\gamma = \frac{1}{T_r}\int_t^{t+T_r} (N_r - A(t))\mathbb{I}_E(t)dt, \quad (1)$$
$$= \frac{1}{T_r}\int_t^{t+T_r} (N_r - C + \hat{U}(t))\mathbb{I}_E(t)dt - \frac{1}{T_r}\int_t^{t+T_r} \epsilon(t)\mathbb{I}_E(t)dt$$
$$= \frac{1}{T_r}\int_t^{t+T_r} \hat{x}(t)\mathbb{I}_E(t)dt - \frac{1}{T_r}\int_t^{t+T_r} \epsilon(t)\mathbb{I}_E(t)dt,$$

where E is the event $E = \{N_r > A(t)\}$ and $\mathbb{I}_E(t)$ is 1 if the event E is true and 0 otherwise. Moreover, $E = \{\epsilon(t) < N_r - C + \hat{U}(t)$, where $\hat{x}(t) = N_r - C + \hat{U}(t)\}$ is known upon the arrival of a new guaranteed request to the capacity broker. Correspondingly the SLA violation of a guaranteed request equals $T_r \times N_r$ (i.e., the total number of RBs of a request lacking resources at some point of its duration).

BE requests with no strict capacity bounds, are accepted provided that certain delay bound is satisfied; therefore, we define δ as the time violation criterion for accepting such kind of requests. In [17], a recommendation limit is defined where a file is dropped if its transfer is not completed within a maximum transfer time equal to T_{drop}. This serves as the SLA violation threshold for accepting BE requests in our scenario. Furthermore, we define the mean number of unused RBs while treating a request (guaranteed or BE), ϕ, as:

$$\phi = \frac{1}{T_r}\int_t^{t+T_r} (A(t) - N_r)(1 - \mathbb{I}_E(t))dt \quad (2)$$
$$= \frac{1}{T_r}\int_t^{t+T_r} (\epsilon(t) - \hat{x}(t))(1 - \mathbb{I}_E(t))dt.$$

C. MuSli: Algorithm for Multitenant Slicing of Capacity

Given the estimated capacity for T time and the characterization of its CD_a, in the following we present an heuristic algorithm for resource slicing among the two types of requests based on the defined parameters. We aim to improve the RB utilization provided that (i) the admitted requests are increased and (ii) the SLA is satisfied per operator and per service type. MuSli, as described in Algorithm 1, prioritizes M_0 (i.e., the mobile network owner) traffic and then accommodates others MVNOs requests in first-in-first-served basis.

The criterion for accepting a request is the predicted capacity. In the first phase of the algorithm (lines 1-10) we treat guaranteed requests generated by M_i. Guaranteed requests are accepted only if there is available capacity alongside their duration without encountering an SLA violation. Every time that we accept a request the available capacity is updated. In the second phase of the algorithm (lines 11-20), we accept a BE request if the time violation criterion is fulfilled during the lasting time of a request.

Algorithm 1 MuSli

Require: $g_{ir} \in \mathcal{G}_{M_i}$ or $b_{ir} \in \mathcal{B}_{M_i}$
Ensure: CD_a

1: **if** $A(t) > 0$ **then**
2:   **while** $g_{ir} \in \mathcal{G}_{M_i}$ **do**
3:     **if** $\phi > 0$ and $\gamma = 0$ **then**
4:       ACCEPT j
5:       UPDATE $A(t)$
6:     **else**
7:       REJECT
8:     **end if**
9:   **end while**
10: **end if**
11: **while** $b_{ir} \in \mathcal{B}_{M_i}$ **do**
12:   **for** t_r **to** T_r **do**
13:     **if** $\delta < T_{drop}$ **then**
14:       ACCEPT
15:       UPDATE $A(t)$
16:     **else**
17:       REJECT
18:     **end if**
19:   **end for**
20: **end while**

V. Performance Evaluation

A. Scenario and Parameters

We consider an Urban Micro-cell scenario with 19 BSs in a hexagonal cell layout with 3 sector antennas based on the IMT-Advanced evaluation guidelines [18]. Detailed system parameters are included in Table I. Users are allocated and move in the network following SLAW model [19], which is a human walk mobility model considering users moving in confined gravity areas.

TABLE I
Basic System Parameters used in the Simulation

Parameters	Settings/Assumptions
PLMNs	3 (i.e., 1 MNO and 2 MVNOs)
Inter-site distance	200 m (ISD)
Bandwidth	20 MHz (100 RBs) 2.5 GHz
Path loss Model	36.7log10(d[m]) + 22.7+26log10(fc[GHz])

With regard to the forecasting, we collected the prior data traffic records from 57 microcell sectors with coverage 2000 m^2. The information that data records contain are *Time*, *Sector ID* and *Resource Volume*. For our simulations, we used two traffic models to represent guaranteed Constant Bit Rate (CBR) and BE traffic following parameters in [17]. The users generate guaranteed CBR VoIP traffic with 64 kbps in DL, as well as BE traffic FTP requests with size fileS=0.5 Mbyte every 60 seconds ($T_{drop} = 8$ seconds). The inter-arrival rate follows a Poisson distribution and the service rate is exponentially distributed.

B. Forecasting Evaluation

For our study we consider the following short-term capacity forecasting methods: ARIMA [12], compressive sensing-based method [15], Kalman filter [14], and Holt-Winters [13]. To identify the most suitable method for the capacity broker, we generated data that spanned in a two-hour prior time period using the SLAW mobility model [19] and we obtained a 20 minute forecast. According to that model, the generated data capture the spatial non-uniformities due to the users' mobility. To compare the performance of the aforementioned methods we choose to use a Random Forest Regressor [20] considering a set of network instances with different load conditions (i.e., low, medium and highly loaded cells). We use the Root Mean Square Error (RMSE), to compare the studied methods. RMSE represents the sample standard deviation of the difference between predicted and observed values. From the observations of the Regressor, we conclude that the most accurate forecasting under different scenarios is the Holt-Winters technique. As shown in Table II, results obtained without using FFT can capture the actual traffic load introducing an error. The outcome of RMSE improves all cases when applying FFT, with the greatest gain achieved in methods that leverage the seasonality of the input data for deriving their results (i.e., Holt-Winters and Kalman filter).

TABLE II
RMSE OF THE STUDIED FORECASTING METHODS

	HW	Kalman	Comp.Bas.Sens.	Arima
Without FFT	4.18	5.25	7.1	9.9
With FFT	2.46	3.97	5.96	7.43

C. MuSli Simulation Analysis

In this section we study the performance of MuSli for varying forecasting CD_as (i.e., 90%, 95% and 99%), in two scenarios. In the first, MVNOs generate only VoIP requests requiring guaranteed capacity. In the second scenario, MVNOs generate both guaranteed and BE requests, with a traffic mix ratio 20% - 80%. Both scenarios are studied for the time duration of the prediction (i.e., 20 minutes), while augmenting the offered load. To evaluate the performance of MuSli we compared it with the baseline scenario, where admission for an MVNO request is based on resource availability at the arrival moment of a request. We conducted Monte-Carlo event-based simulations in MATLAB® with 1000 iterations to achieve statistical validity for each forecasting step.

1) Scenario 1 - Guaranteed VoIP Traffic: In Fig. 3, we study the (a) RB utilization provided that the offered load consists of requests without SLA violation, (b) SLA violation, versus the offered load associated with MVNO requests.

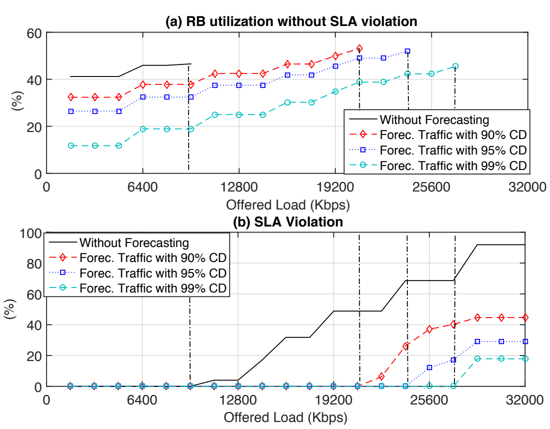

Fig. 3. (a) RB utilization and (b) SLA violation for VoIP Traffic.

Our first observation is that when MuSli is applied, the utilization without SLA violation, is increased far beyond the baseline approach as we introduce more offered load. In Fig. 3(a), the curves of RB utilization stop at a certain offered load limit beyond which we accept no more requests due to SLA violation (i.e., 9600 Kbps for the baseline scenario, 20800 Kbps for 90% CD, 24000 Kbps for 95% CD and 27200 Kbps for 99% CD). The vertical dashed lines, point out this limit (i.e., successfully served requests' load without violating SLAs). The maximum amount of offered traffic without an SLA violation for 90% CD, results in 23% higher utilization compared to the baseline scenario, whereas this percentage for 95% CD is 17.27%. Despite the fact that MuSli using 99% CD results into 42.3% (i.e., lower than the 44.2% achieved by the baseline scheme), it can accommodate higher offered load associated with MVNO requests (i.e., 27200 Kbps). Fig. 3(b) shows the SLA violation for the different approaches. When MuSli with 99% has the maximum offered load of requests without SLA violation (i.e., 27200 Kbps), the baseline scenario results in 68.63% violations. For the same offered traffic, MuSli with 90% CD and with 95% result in 52.71% and 19.3% SLA violations respectively.

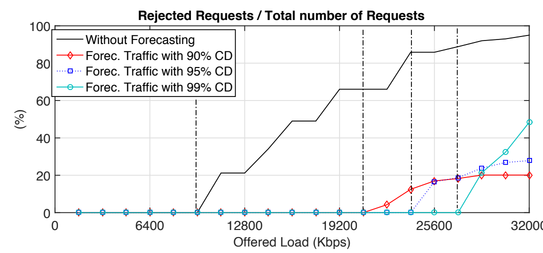

Fig. 4. Rejected Requests for VoIP Traffic.

Fig. 4 presents the percentage of rejected VoIP requests with the offered load. Apparently, more CBR requests are accepted when MuSli is applied, compared to the baseline scenario. When MuSli with 99% CD serves the maximum offered load without SLA violation (i.e., 27200 Kbps), the baseline scenario rejects 89.8% of the MVNO requests. On the same time, MuSli with 95% and 90% reject 19.25% of them. Since MuSli is facilitated with traffic prediction, it selects the MVNO request that can be served without an SLA violation, considering their starting time and duration. Hence, MuSli results in a more efficient RB utilization compared to the instantaneous decision. We observe that lower CD results in accepting more requests but with an SLA violation for lower offered load.

2) Scenario 2 - Guaranteed VoIP and BE FTP Traffic: In Fig. 5 we study the same metrics depicted in Fig.3, considering two types of requests (i.e., guaranteed and BE in ratio 20%-80%). In this scenario, because of the traffic ratio and different types of requests, we double the offered load. The vertical dashed lines denote the limit of successfully served requests without SLA violation (i.e., 9600 Kbps for the baseline scenario, 38400 Kbps for 90% CD, 44800 Kbps for 95% CD and 51200 Kbps for 99% CD).

By introducing two types of capacity slicing classes, the MNO can make the best out of the available RBs achieving higher utilization than in Scenario 1. The highest utilization

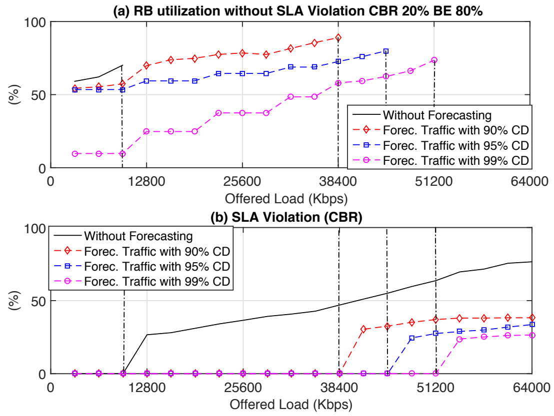

Fig. 5. (a) RB utilization and (b) SLA violation for 20% VoIP and 80% FTP.

without SLA violation (i.e., 89.9%) is achieved for MuSli with 90% CD, for 38400 Kbps. When the maximum offered load for MuSli with 99% CD is introduced (i.e., 51200 Kbps), 73.54% of RBs are utilized, whereas for 95% the corresponding percentage is 79.97%. Although this approach is the most conservative in slicing resources among MVNO requests based on SLA, it still outer-performs the baseline scenario. In Fig.5(b) we study the SLA violation for CBR requests. When Musli with 90% accommodates the maximum amount of load (i.e., 38400 Kbps), the baseline scenario results in 46.96% SLA violation. Although MuSli with 99% CD results in the lower utilization, it achieves the lowest SLA violation for the highest offered load of request (i.e., 51200 Kbps). Therefore, we confirm the trade-off among service quality and CD of the forecasted capacity.

In Fig.6 we study CBR and BE rejected requests. We observe that MuSli with 99%, after accepting the maximum load without SLA violation (i.e., 51200 Kbps), rejects both CBR and BE requests exponentially. At the same time, for 51200 Kbps, MuSli with 90% rejects 19.63% CBR and 20.47% BE traffic. BE requests are dropped with lower rate compared to the CBR ones when available capacity exists due to their delay tolerance. The same tendency appears for MuSli with 95% CD. In this Scenario, MuSli achieves higher utilization by accepting first CBR requests according to their starting time and duration and then BE based on the available capacity.

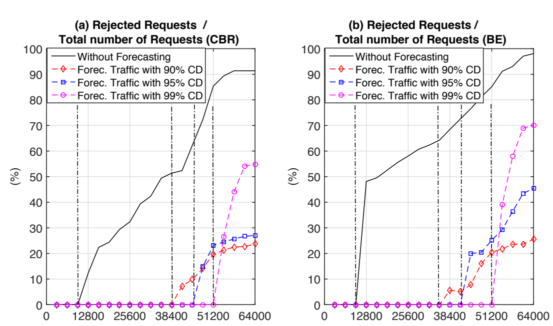

Fig. 6. Rejected Requests for 20% VoIP and 80% FTP Traffic.

VI. Conclusion

In this paper we proposed a capacity broker architecture enabling on-demand resource slicing for multiple tenants. The current 3GPP LTE-A network sharing architecture is extended to introduce the capacity broker element along with the required interface extensions. In addition, this work improved a capacity short-term traffic forecasting algorithm to capture the traffic non-uniformities. We also proposed MuSli, a capacity slicing algorithm for two types of requests (i.e., guaranteed and BE) taking into account the CD of the forecasting. Finally, the proposed framework has been evaluated in terms of mobile network resource utilization improvements and SLA violations cost by simulation. Our results show, that for guaranteed services without SLA violation, our capacity broker architecture and algorithm would allow a resource block utilization gain in the range of 15-25% (while regarding 95-90% CD of the forecasting and a traffic mix of requests). Future work concentrates on analyzing incoming/service rates considering a dynamic capacity limit, with varying link conditions.

Acknowledgement

This work has been funded by the MITN Project CROSSFIRE (PITN-GA-2012-317126).

References

[1] GSMA, "Network Infrastructure Sharing," Sep. 2012.
[2] K. Larsen, "Network Sharing Fundamentals," Jul. 2012.
[3] *3GPP TS 32.130 Telecommunication management; Network Sharing; Concepts and requirements, Rel.12*, Dec. 2014.
[4] *3GPP TS 23.251, Network Sharing; Architecture and Functional Description*, Rel.13, Mar. 2015.
[5] Y. Zaki *et al.*, "LTE Wireless Virtualization and Spectrum Management," in *IFIP WMNC, Budapest*, Oct. 2010.
[6] X. Costa-Perez *et al.*, "Radio Access Network Virtualization for Future Mobile Carrier Networks," *IEEE Comm. Mag.*, vol. 51, no. 7, Jul. 2013.
[7] T. Guo and R. Arnott, "Active LTE RAN Sharing with Partial Resource Reservation," in *IEEE VTC Fall, Las Vegas*, Sep. 2013.
[8] *3GPP TR 22.852, Study on RAN Sharing enhancements, Rel.12*, Sept. 2014.
[9] J. Panchal, R. Yates, and M. Buddhikot, "Mobile network resource sharing options: Performance comparisons," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, Sep. 2013.
[10] P. Calallero *et al.*, "RMSC: A Cell Slicing Controller for multi-tenant Mobile Network," in *IEEE VTC-Spring, Glasgow*, May 2015.
[11] W. Jiewu *et al.*, "User traffic collection and prediction in cellular networks: Architecture, platform and case study," in *IEEE IC-NIDC, Beijing*, Sep. 2014.
[12] Y. Shu *et al.*, "Wireless traffic modeling and prediction using seasonal arima models," in *IEEE ICC, Anchorage*, vol. 3, May 2003.
[13] D. Tikunov and T. Nishimura, "Traffic prediction for mobile network using Holt-Winter's exponential smoothing," in *15th SoftCOM, Dubrovnik*, Sep. 2007.
[14] A. Yadav *et al.*, "A constant gain kalman filter approach to target tracking in wireless sensor networks," in *IEEE ICIIS, Chennai*, Aug. 2012.
[15] R. Li *et al.*, "Energy savings scheme in radio access networks via compressive sensing-based traffic load prediction," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 4, Apr. 2014.
[16] J. F. Lawless and M. Fredette, "Frequentist prediction intervals and predictive distributions," *Biometrika*, vol. 92, no. 3, 2005.
[17] *3GPP TR 36.814 Further advancements for E-UTRA physical layer aspects*, Rel. 9, Mar. 2010.
[18] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," Report ITU-R M.2135-1, Dec. 2009.
[19] K. Lee *et al.*, "SLAW: Self-similar Least-action Human Walk," *IEEE/ACM Trans. Netw.*, vol. 20, no. 2, Apr. 2012.
[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, 2001.