LTE Virtualization: from Theoretical Gain to Practical Solution

Liang Zhao*, Ming Li†, Yasir Zaki*, Andreas Timm-Giel† and Carmelita Görg*
*ComNets/TZI
University of Bremen, Otto-Hahn-Allee NW1, 28213 Bremen, Germany
Email: {zhaol, yzaki, cg}@comnets.uni-bremen.de
†Communication Networks (E-4)
Hamburg University of Technology, Schwarzenbergstrasse 95E, 21073 Hamburg
Email: {ming.li, timm-giel}@tuhh.de

***Abstract*—Virtualization of wireless networks has received more and more research attention. As a means to reduce the investment of mobile network operators and to improve the network performance, LTE (Long Term Evolution) virtualization is one important study cases in the scope of wireless virtualization. To investigate the advantages of virtualization on the LTE air interface, this work starts with an analytical model of FTP (File Transfer Protocol) transmissions in virtualized LTE systems, and an obvious multiplexing gain can be observed. Evaluations with the considerations of realistic simulation models and mixed services of FTP and VoIP (Voice over IP) traffic also validate the analytical analysis. Further, an extended multi-party spectrum sharing model is proposed. As one of the key issues, spectrum budget estimation is further analyzed based on the characteristics of real time services, i.e. VoIP, and it is validated by further simulations that the proposed mechanism can give a very close estimation on the instant spectrum requirements in one cell. Some of the other investigation possibilities in practice are also discussed at the end.**

***Keywords*–LTE virtualization, spectrum sharing, load estimation.**

I. Introduction

Network Virtualization is an attractive technique that has received more and more research attention, and it is seen as one of the key technologies and features in the future network development. For example, 4WARD [1] in Europe, GENI [2] in the U.S. and AKARI [3] in Asia, have put their effort on the network virtualization. Wireless virtualization is a crucial component of focusing on sharing wireless resources among multiple virtual networks, and e.g. in [4] a general framework of wireless virtualization was discussed. In our previous work, [5] and [6], LTE virtualization as a case study was investigated, and the focus was to reveal the advantages of LTE virtualization and an initial framework was proposed.

The targets of LTE virtualization are in line with the generic network virtualization, e.g. physical infrastructure sharing, isolation, cost and energy saving and spectrum sharing. This means multiple virtual operators (VOs) are sharing one common infrastructure from the air interface to the core network. This consists of passive elements (like sites, tower, antenna, feeder and power supply), active elements (like radio base station, radio network controller and the transport links), the components of the core network and the frequency bands. Some previous concepts of network sharing exists in 2G and 3G systems like Mobile Virtual Network Operator (MVNO, e.g. [7] for further reading), but LTE virtualization is doing more: the main radio resource, the frequency, can also be dynamically isolated and shared among different VOs. Each VO can enforce its own policies and algorithms in the individual virtual mobile system to achieve its commercial expectation.

Some of the other advantages of LTE virtualization to motivate the investigations of this paper are: Nowadays big companies play multiple roles in the market like infrastructure providers, network operators and service providers. In order to compete for users, they have to keep improving their infrastructure to ensure the coverage and QoS (Quality of Service) of the network. With virtualization, the infrastructure provider can concentrate on the maintenance of the physical equipment and save manpower for operating the networks. The duplication of deploying the infrastructure can be avoided, when multiple virtual mobile networks are running on a common infrastructure. The huge investment on the hardware and fundamental construction are saved as well. This also enables small companies (like "Mini VO" discussed in Section IV) to get into the market by renting small parts of the physical resources to compose their own mobile networks to provide specific services. The deployment, maintenance, migration and upgrade of the virtual mobile systems are more flexible thanks to online re-configurability of network virtualization. Multiple operators sharing one infrastructure can also save energy as the amount of the physical nodes is reduced.

In reality, the distribution of the user data traffic in the coverage area of a mobile system is irregular, from cell to cell and from time to time. Normally, the networks are deployed to handle the worst case, meaning the resources are overprovisioned for the situation of low-to-medium traffic load. This indicates that the spectrum is not always fully utilized. Virtualization can introduce the multiplexing gain by exploiting the multi-VO diversity through spectrum sharing, as discussed in our previous work [5]. In this work we answer the questions like how much gain we get and to which extent we can still obtain sufficient gain when a large number of users with different services presents in the individual VO's networks. In [6], a spectrum sharing framework was proposed by us that is based

978-0-9836283-0-9

This paper was peer reviewed by subject matter experts for publication in the Proceedings of ITC 2011

on different contracts signed between the spectrum owner (infrastructure provider), and the spectrum customers (VOs). Here the framework is generalized by proposing a multi-party sharing model, in which the spectrum sharing behavior of each VO is defined by their business policies. In our scope the presented solution is doing the spectrum sharing in a centralized way on licensed frequency, instead of the sensing-to-allocation way of the cognitive radio approach. Thus, one of the preconditions (and difficulties) to apply this framework into practice is the spectrum requirement estimation based on the current traffic load of each VO in a long time interval. One mechanism proposed is to address this issue by observing the statistical characteristics of realtime traffic, i.e. VoIP. Based on this, the accurate spectrum requirement estimation for each VO can easily be deduced from the number of active VoIP calls.

With the above mentioned motivation, our contributions in this paper are summarized as:

- The multiplexing gain introduced by the spectrum sharing of LTE virtualization is revealed from both analytical and simulation perspectives.
- An generalized multi-party model is proposed as a framework to enable the centralized spectrum sharing among different VOs based on their business policies.
- To address the key issue of an effective spectrum sharing, a mechanism of spectrum budget estimation based on real-time services is also proposed for practical applications of LTE virtualization.

The rest of this paper is organized as follows: Section II introduces the assumptions and system models employed by this work, and Section III starts with a simple analytical model of FTP transmissions to reveal a large multiplexing gain based on LTE virtualization. The results of this analytical model is validated by comparing it with the outcome from a system level LTE simulator. In order to clarify how much multiplexing gain can be obtained, extensive simulations are run with considering large numbers of VoIP users in the system. Motivated by the simulation results, an extended model of multi-party spectrum sharing is proposed in Section IV with some detailed discussions. In order to address the key issue to apply LTE virtualization into real systems, in Section V we further propose a mechanism of load/spectrum estimation based on the statistical characteristics of real-time services, e.g VoIP, to support the multi-party spectrum sharing model. At the end, Section VI draws the conclusions of this paper and gives some outlook for our future work.

II. Assumptions and System Models

A. Scenario

For the analytical investigation on the gain obtained from spectrum sharing, the assumption is that each VO o belonging to a set of operators $\mathcal{O} = \{1, 2 \text{ and } 3\}$ owns the same bandwidth $B = 5$ MHz, which is equivalent to $N = 25$ PRB (Physical Resource Blocks) on the LTE air interface. In the presented work, we only consider the downlink transmissions. Since the uplink transmission in LTE is coordinated by the eNodeB (eNB), the results onward are also valid for the uplink to some extent if additional considerations on the localized spectrum allocation of SC-FDMA (Single Carrier-FDMA) are taken into account.

Mutiple *virtualized* eNBs of different VOs are co-located in the *physical* eNBs. In our assumed scenario, the virtual network of each VO and the physical network have an identical topology, which means in each *physical* eNB each of the VOs has one virtual eNB. Ideally, the coverage of all VOs are perfectly overlapped, therefore the investigation can concentrate on one cell only because of the similar situation in adjacent cells. Each VO can share his spare spectrum to any other VOs in the same geographic region through the control of a central management unit named "hypervisor". Practically, the hypervisor can be located in the physical eNB.

B. Traffic Models

Two traffic models, FTP download and VoIP, are employed for the theoretical investigation and simulation. FTP transmission is a typical best effort (BE) service, which is delay tolerant, and in principle a huge bandwidth can be occupied by a few FTP transmissions if the channel is good enough and there is sufficient data in the buffer. The FTP download time is used in this work to benchmark the comparisons between different scenarios, and its setting follows Table I, where the files to be downloaded have the mean file size $\bar{\chi}$ of 5 MB following an exponential distribution. The inter-arrival time (IAT) defines the interval between two arriving downloads, where λ is then the Poisson-distributed arrival rate. IAT is also exponentially distributed and can be set with different mean values to scale the traffic load, meaning the lower the IAT the more the files to be downloaded.

In contrast, the VoIP service is using low bandwidth in the talking period but is highly delay sensitive for guaranteeing the quality of real-time conversations. Typically, the maximal delay budget of VoIP transmission on the air interface is 50 ms [8]. One simplified way of modeling VoIP traffic used in this work is given in Table II.

The activity factor of 0.5 means during a call the probability of talking and listening is the same, and both the talking time and silence (listening) time have the mean value of 3 seconds following an exponential distribution. Following the VoIP codec and additional overhead from different protocol layers [9], the data packet size with header compression based on ROHC (Robust Header Compression) [10] is 40 bytes at the beginning of each 20 ms on the air interface. Without header compression, the packet size is around 100 bytes or even more, but these kind of packets are only shortly used for carrying more signaling information, e.g. at the beginning of each call. SID (Silence Insertion Descriptor) packets are used during the silence time with very low payload. Thus, in the

TABLE I
FTP DOWNLOADING

Parameter	Assumption
average file size $\bar{\chi}$	exponential distribution with mean of 5 [MB]
average IAT $1/\lambda$	exponential distribution

TABLE II
VOIP TRAFFIC MODEL

Parameter	Assumption
activity factor	0.5
talking time	exponential distribution with mean of 3 [s]
silence time	exponential distribution with mean of 3 [s]
call duration	complete run time
data packet size	40 [byte] every 20 [ms] during talking time
SID packet	20 [byte] every 160 [ms] during silence

following evaluations, only the packets with 40 bytes are taken into account.

C. Simulation Model

A MATLAB®-based system level simulator is developed for proof of concept. The simulation area consists of 19 *physical* eNBs with a typical hexagonal layout, each of the eNBs has 3 sectors. The frequency reuse factor is 1, meaning each *virtual* sector has the resources of 25 PRBs. It has to be kept in mind that each of the *physical* eNBs can be virtualized into several *virtual* eNBs. Our focus is investigating the spectrum sharing of LTE virtualization. Related work on the hardware virtualization of eNB can be found as e.g. [11].

Some assumptions have been employed without losing the generality of the results: downlink channel knowledge at the eNB is perfect. 2D correlated slow fading is recalculated in a large time interval $\gg$ TTI (Transmission Time Interval)$=$1 ms, and TTI-based fast fading is also considered in simulations. Users are fixed for a while (e.g. 10 sec $=10^4$ TTIs) and reallocated again in the next batch of TTIs. Since the results are obtained from Monte-Carlo evaluations and there is enough randomness introduced, then the lack of moving users is not affecting the validation of the results. Active users are only located into 3 sectors of the central eNB, and the surrounding eNBs are considered as the source of interference. The assumption here is that all of the PRBs from non-serving cells are fully utilized and there are no advanced interference avoidance and no adaptive power loading mechanisms applied on the PRBs. Then the SINR (signal to noise and interference ratio) $\gamma_{c,k,n}$ of user k on PRB n in serving cell c is defined as

$$\gamma_{c,k,n} = \frac{P^r_{c,k,n}}{\mathrm{N}_0 + \sum_{c' \neq c} P^r_{c',k,n}}, \tag{1}$$

where $P^r_{c,k,n}$ is the received power at the UE (User Equipment) antenna, and it is differentiated from cell to cell and PRB-specific due to frequency-selectivity. N_0 is the thermal noise on each PRB. By including the channel fading, there is

$$P^r_{c,k,n} = \frac{P^t_{tot}}{N \cdot PL_{c,k,n}}, \tag{2}$$

where P^t_{tot} is the total output power at the eNodeB antenna, which is homogeneously distributed on N PRBs and attenuated by the path gain PL including path loss, slow fading and fast fading. An overview of this system level simulation model is summarized in Table III.

In order to clarify the scope of the presented work, the generality and the limitations of our assumptions are discussed: first, for the simplicity of the simulations, only VoIP and FTP users are used to represent homogeneous users. If other services are used like HTTP, SMS and VoD (video on demand) that have different traffic characteristics, the individual simulation results will be affected and a new evaluation is required. Within our future work, more scheduling principles, e.g. Proportional Fair and Max C/I will be investigated, and the influences of such scheduling principles on the resource utilization will be studied. Apart from the "identical coverage of VOs" assumption, the topologies of the different VOs may be different due to e.g. cell breathing and beam-forming, especially at cell edges. Thus, the spectrum sharing at the "non-overlapping" area could be very difficult and severely limited by the inter-cell interference. Corresponding solutions should be proposed as well. Nevertheless, the intention of the presented work is to reveal the potential gain of the LTE visualization by employing simplified models, and the general tendency of the virtualization gain should still hold. In the following sections, simulations are performed using 50 runs with different seeds so as to achieve convergent results, if there is no additional prompt. Since users are gradually arriving to the system, the initial phases are excluded from the analysis so as to obtain accurate results.

TABLE III
SIMULATION MODELS

Parameter	Assumption
number of VO	1 to 3
eNBs in each VO	19 hexagonally in 3 tiers
sectors of each eNB	3
site-2-site distance	500 [m]
spectrum of each VO	5 MHz := 25 PRBs
total output power	20 [W] := 43 [dBm]
path loss model	COST 231
2D slow fading	20 m decorrelation distance
fast fading	Rosa Zheng model [12]
mobility model	reallocation of stationary users
traffic model	Table I and II
scheduling	VoIP-prioritized Round-Robin

III. ANALYTICAL ANALYSIS AND VALIDATION WITH SIMULATION

A. Simple Analytical Analysis

In order to reveal the multiplexing gain obtained by LTE virtualization, we consider an FTP-only scenario as the beginning of our analysis. In the design of the scheduler it is reasonable that all of the parallel FTP downloads are scheduled in a fair way. Since all of the flows (FTP downloads) can share the radio resources equally, a Processor Sharing (PS) model can be used to model the scheduler behavior. As a best effort service, the traffic of FTP connections is elastic and one FTP connection can theoretically occupy the whole capacity C of one cell, if only one UE stays in the cell with good enough channel and sufficient data in the buffer. Hence M/G/1-PS is a valid model for this situation, that each flow has the ability to fully utilize the whole capacity when no other flow is present in the system. "M" requires the flow arrival follows a Poisson process with arrival rate λ, meaning the inter-arrival time between two downloads follows an exponential distribution with mean $1/\lambda$. "G" means a generally distributed service time and "1" server applies the PS discipline. According to [13] the expected sojourn time (the average FTP downloading time τ) for a file with size χ in M/G/1-PS is given by :

$$\mathrm{E}_{\mathrm{M/G/1-PS}}\left\{\tau_1\right\} = \frac{\chi}{C} \cdot \Delta_1(\rho), \tag{3}$$

where ρ and Δ_1 denote the mean traffic load and delay factor on the air interface, respectivly. ρ is defined by the mean file size $\bar{\chi}$ and arrive rate λ as

$$\rho = \frac{\lambda \cdot \bar{\chi}}{C}, \tag{4}$$

and

$$\Delta_1(\rho) = 1 + \left.\frac{\mathrm{Er_C}(R, R\cdot\rho)}{R\cdot(1-\rho)}\right|_{R=1} = \frac{1}{1-\rho} \tag{5}$$

indicates the increase of the file downloading time when the arrival rate λ is higher and multiple parallel downloads are sharing the capacity. $\mathrm{Er_C}(\cdot)$ is the *Erlang C formula* (checking e.g. in [13]) meaning the waiting probability in a system.

In case of multiple VOs virtualization, an assumption has been made that all the radio resources from all VOs can be optimally shared to all VOs in each TTI. In other words, all the resources are scheduled to all VOs in a same way from one common sharing pool through one centralized scheduler. Therefore, the above mentioned sharing model (M/G/1-PS) can be further applied to this multiple VOs virtualization model.

Suppose each VO has the same amount of spectrum and the same mean number of FTP users which corresponds to the same mean file size and arrival rate in the M/G/1-PS model. In case of O VOs, the arrival rate and system capacity increase by O times, meaning (4) and (5) keep unchanged. One important observation based on (3) is that the expected sojourn time in the system decreases to

$$\mathrm{E}_{\mathrm{M/G/1-PS}}\left\{\tau_O\right\} = \frac{\chi}{O\cdot C} \cdot \Delta_1(\rho) = \frac{1}{O}\mathrm{E}_{\mathrm{M/G/1-PS}}\left\{\tau_1\right\}. \tag{6}$$

This means that if 2 VOs are sharing their capacities, 50% of the downloading time can be saved, and further downloading time reductions of 66.7% and 75% can be expected for the situation of 3 VOs' and 4 VOs' capacity sharing respectively.

The discussion above is actually revealing the significant multiplexing gain because of the asynchronous traffic fluctuation in individual VOs. This means if the traffic load in one VO is low (less number of FTP downloads), the spared spectrum can be utilized by other high-load VOs to enhance the performance. However, the uncertainty on this conclusion is how much gain can be obtained in the real LTE systems. The transmission on the air interface is PRB-based, which means the system capacity can be shared only with corresponding limitations, i.e. the achievable rates of different users are highly depending on their instant channel conditions. Normally, in one cell there are many VoIP connections which need at least one PRB each and there might be only very limited free resources left in each VO for multiplexing. Thus, some comparisons and validations are done by system level simulations and numeric results are shown in the next section especially in Fig. 1 and Table. IV.

B. Validation by Simulation

In order to validate the conclusion drawn by the analytical analysis in the previous section and to clarify the above mentioned questions, different scenarios are set up based on the system level simulator introduced in Section II-C:

- The number VOs $O = \{1, 2 \text{ or } 3\}$.
- The number of VoIP users is ranging from 0 to 150 for each VO, the duration of the calls is as long as the simulation run.
- The mean IAT ($1/\lambda$) of FTP downloads is ranging from 10 to 30 seconds to scale the traffic load.
- The scheduling principle is using the VoIP-prioritized Round-Robin method. Firstly, all of the VoIP transmissions are satisfied. If there are resources left, the FTP transmissions are executed. All of the PRBs are evenly distributed within individual service categories.

According to (1) and Table II, the transmission of each VoIP packet of 40 bytes in 1 ms satisfies

$$N_{c,k} = \left\lceil \frac{1.8\times10^5 \cdot \zeta \cdot \log_2\left(1+\overline{\gamma}_{c,k}\right)}{40\times8\times10^3} \right\rceil, \tag{7}$$

where $N_{c,k}$ is the number of PRBs required by VoIP user k in serving cell c for sending one 40-Byte packet on the air interface, and $\lceil\cdot\rceil$ is the ceiling operator. In OFDMA of LTE, each of the PRB has the bandwidth of 180 kHz. Assuming perfect channel knowledge and advanced PHY procedures like AMC (Adaptive Modulation and Coding) and Turbo code, the usage of (7) actually follows the Shannon's theorem of channel capacity. $\overline{\gamma}_{c,k}$ is the mean SINR over $N_{c,k}$ PRBs, and parameter $\zeta \in (0, 1)$ excludes the signaling overhead and models minor packet loss.

Similarly, the FTP download in each TTI is represented as

$$\chi^r_{c,k}(t+1) = \chi^r_{c,k}(t) - 180\cdot\zeta\cdot N_{c,k}\cdot\log_2\left(1+\overline{\gamma}_{c,k}\right), \tag{8}$$

where $\chi^r_{c,k}(t+1)$ is the size of the remaining part of one FTP download in the next TTI, and $N_{c,k}$ is the number of PRBs allocated to the FTP user k by the Round-Robin scheduler in TTI t. When $\chi^r_{c,k}(t+1) \leq 0, \forall t$, this FTP download is completed.

Note that for simplification the index o of VOs is omitted in (7) and (8), but the operations hold for all VOs. Fully meshed simulations are intensively executed according to the combinations of different settings. Fig. 1 illustrates the 3-D results of the average FTP downloading time $\mathrm{E}(\tau)$. It is obvious that with an increased number of VOs in the system (total number of users are increased proportionally), the average FTP downloading time $\mathrm{E}(\tau)$ decreases significantly. Even with a large number of VoIP users and more frequent FTP transmissions, the virtualization gain is still considerable. With an increased number of VOs, the multiplexing gain is

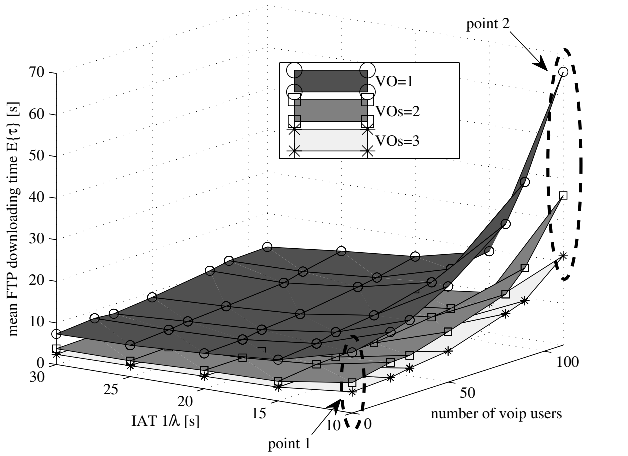


Fig. 1. Multiplexing gain from virtualization represented by decreasing FTP downloading time because of spectrum sharing (simulations)

exploited, and that is why the surfaces of VOs=2 and VOs=3 are getting close. In order to control the scale of the figure, some parts of the results like the points of VoIP users >100 are removed, but this doesn't disturb the tendency of the results.

Table IV shows the comparison between the estimation based on the M/G/1-PS model and simulation results, and those points of simulation results are picked from Fig. 1 as indicated. On point 1 there is no VoIP user in the system, and the results in terms of mean FTP downloading time reduction from M/G/1-PS and simulation are quite close to each other, i.e. 50% reduction of $\mathrm{E}_{\mathrm{M/G/1-PS}}\{\tau\}$ compared to 51.4% reduction of $\mathrm{E}\{\tau\}$ when 2 VOs are sharing their capacities against to VO=1, and 66.6% reduction estimated by M/G/1-PS compares to 67.1% reduction simulated with 3 sharing VOs. For the situation with 100 VoIP users in the system, the simulation results are still close to the ones theoretically estimated with minor increase of the differences.

Two remarks can be made to explain the small relative errors between analytical estimations and simulations: first, the cell capacity C in (3) is no longer a constant in the simulation, because on the LTE air interface the instant cell throughput is depending on the physical distribution of the users, the channel fading and the interference. Multiple VOs are actually sharing the spectrum (PRBs), which can be indirectly translated into instant cell capacity. Second, when VoIP users are coexisting with FTP downloads in the system (point 2), there is a reduction of C for FTP downloads because of the VoIP-prioritized scheduling. Besides the uncertainty of instant C, the reduction of C is also fluctuating due to the bursty nature of the VoIP traffic. The simulation results reflect a long-term expectation which is scenario-specific and nearly approaching the theoretical estimation (with minor simulation errors).

TABLE IV
COMPARISON BETWEEN M/G/1-PS MODEL AND SIMULATIONS: THE RELATIVE GAIN FROM CAPACITY SHARING

Points	VOs	M/G/1-PS	Simulation
point 1 FTP only	$O = 1 \rightarrow 2$	50%	1-5.91s/11.91s = 51.4%
	$O = 1 \rightarrow 3$	66.7%	1-3.92s/11.91s = 67.1%
point 2 FTP and VoIP=100	$O = 1 \rightarrow 2$	50%	1-35.9s/66.71s = 46.2%
	$O = 1 \rightarrow 3$	66.7%	1-21.5s/66.71s = 67.8%

IV. EXTENDED MODEL OF MULTI-PARTY SPECTRUM SHARING

Both theoretical analysis and simulation results in the previous section reveal the absolute multiplexing gain of LTE virtualization. Motivated by this, a multi-party spectrum sharing model is proposed and discussed in this section.

In [14] a capacity sharing model between two operators was proposed with different strategies depending on how much capacity will be reserved or shared by each individual operator. Actually, four sub-strategies therein, "Fully split", "Fully pooled", "Partial reservation" and "Unbalanced", can be summarized into a single picture for spectrum sharing as seen in Fig. 2(a), where 2 operators have the spectrum of $2 \times B$ all together and both of them have a resource reservation of $B_{r1}, B_{r2} \in [0, B]$, repectively. The rest of the spectrum $2B - B_{r1} - B_{r2}$ is shared by VO_1 and VO_2. The dashed line sliding in Fig. 2(a) indicates exactly how much resources are being utilized by which operator after sharing.

The drawback of this graphical model is that it is difficult to illustrate the concept of spectrum sharing if more than 2 operators join this game. In [6] we already proposed a contract based spectrum sharing model, and by combining it we can have an extended diagram of multi-party spectrum sharing as in Fig. 2(b). As an example shown in the picture there are 4 VOs sharing the spectrum, and each of them has different strategies. From a logical point of view, the reserved spectrum is kept in the individual pool of each VO and the shared portion of spectrum is put into and taken from the "Sharing Pool" in the middle.

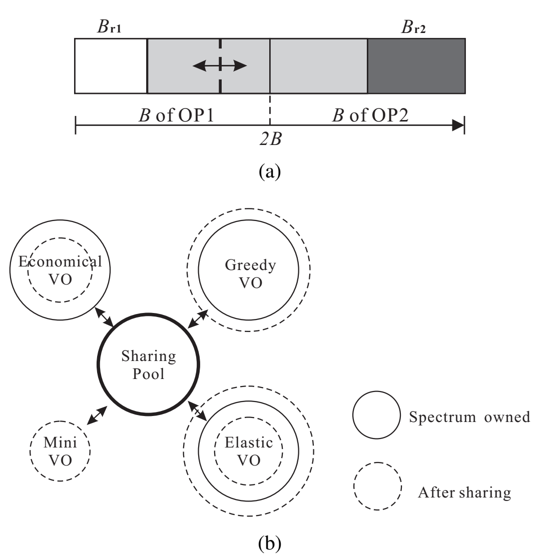


Fig. 2. Extension of spectrum sharing model

The sharing strategies are based on the services provided by the different VOs. The target of one operator is saving the cost on the spectrum utilization, then his strategy should be conservative as "Economical VO" meaning a minimum bandwidth is reserved to meet the stable transmissions of his users. One VO can also be a "Greedy VO" with resource budgets $\geq B$ meaning he is not sharing his own spectrum but can borrow the spectrum from others to provide better performance to his users. Actually, most of the VOs should be "Elastic VOs" to adjust their spectrum reservation adapting to the traffic load and the mixture of the traffic. Additionally, one special type of VO, "Mini VO" can exist thanks to the spectrum sharing of LTE virtualization. He doesn't own any spectrum, but can utilize the spared spectrum from other VOs to provide specific services for a certain area and customers, e.g. in the working area of a company or a factory. Note that the difference between this model and the proposal in [6] is that the assumption here is each VO has his own spectrum (except Mini VO) and shares some of this spectrum with the other VOs. In [6], the infrastructure provider owns the whole spectrum and allocates the spectrum to VOs according to their instant requirements and priorities, which is actually one special case of "Fully pooled". In each sharing interval, VOs report their spectrum requirements to a cental unit, hypervisor, which makes the decision of spectrum sharing. When competition happens, meaning the sum of the requirements exceeds the resources available, the hypervisor will try to fullfill the resource limitation according to different policies. For example, hypervisor will evenly cut down the requirements of all VOs or gradually reduce the biggest requirement.

Generally, the formulation of the sharing policies is a trade-off between performance and operation expenditure of VOs. It is very important to VOs that they can draw a baseline to estimate how much resources are required for supporting their traffic. On the other hand, due to the reason of processing complexity and signaling overhead, spectrum sharing among VOs cannot be processed on TTI-level, and meanwhile the traffic load can fluctuate. Therefore, a reliable mechanism of load estimation is crucial for a practical application of spectrum sharing in LTE virtualization, especially on a large time scale $\gg 1$ ms.

V. Load Estimation based on Real-time Services

From the description in Section IV the success of the spectrum sharing is highly depending on the accuracy of the load estimation from different VOs for guaranteeing their minimum performance. From the resource efficiency point of view, the budget of resource requirement should tightly follow the traffic load, and at the same time some margin should also be left for tolerating occasional traffic peaks. Normally, as in [6] the moving average method is used for the spectrum requirement estimation, and the limitation of it is that its performance is quite sensitive to the weight factor of the traffic load history and to the interval of the resource re-allocation. To overcome this problem, in this section one new load estimation mechanism is proposed based on the statistical characteristics of the real time-services, e.g. VoIP. VoIP is one of the most important services for operators. It needs a low but constant requirement of bandwidth to sustain voice calls and it is also very delay sensitive because the conversation must be in real time.

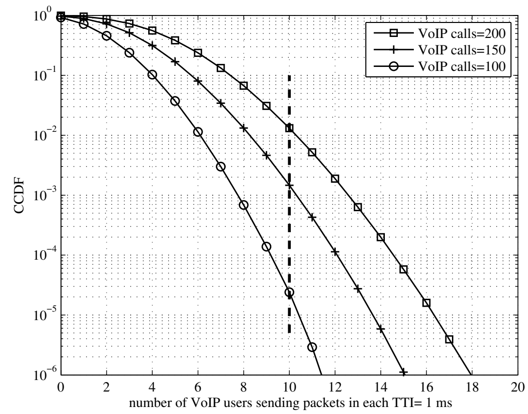


Fig. 3. CCDF of the number of sending users in each 1 ms

A. Load Estimation based on VoIP Traffic Model

By recalling the traffic model of VoIP in Section II-B, it can be noticed that during a call on average less than 5% (every 20 ms in ON period) of the time the VoIP traffic really has data packets to be transmitted over the air interface. Then a natural assumption is that even with a large number of active VoIP calls ($100 \sim 200$) in one cell the actual required spectrum to sustain all of them is quite limited, because the data packets from different VoIP calls are transmitted in different time intervals (TTIs). The worst case scenario, e.g. > 100 parallel VoIP packets in the same TTI on the air interface, has a very low probability. This can be validated by Fig. 3, which is obtained after a long simulation run (50 hours system time) based on the traffic model in Table II. With more details in the low probability area on the logarithmic y-axis, the CCDF curves indicate the probability of the average number of active users in each TTI$=$ 1 ms. It can be seen that even with 200 active VoIP calls in one cell, e.g. the probability of the parallel active users greater than 13 is less than 0.1%. The load estimation of 10 active user in each TTI can already cover 99% of the user level load, which means 20 PRBs are enough for supporting 200 VoIP calls if each standard VoIP packet of 40 bytes occupies 2 PRBs on average.

According to the VoIP traffic model, in the talking time of one user the packets appear in every 20 ms. By checking the statistical features of VoIP traffic, Fig. 4 proves that even if 200 VoIP users have calls in parallel, a very strong autocorrelation of their overall traffic can still be observed every 20 ms. With increased lag up to 500 ms, the autocorrelation drops slightly. In a short time interval (e.g. 500 ms), the transmission sequence of one talking user on each TTI can be seen as deterministic, meaning one packet arrives every 20 ms, and all of

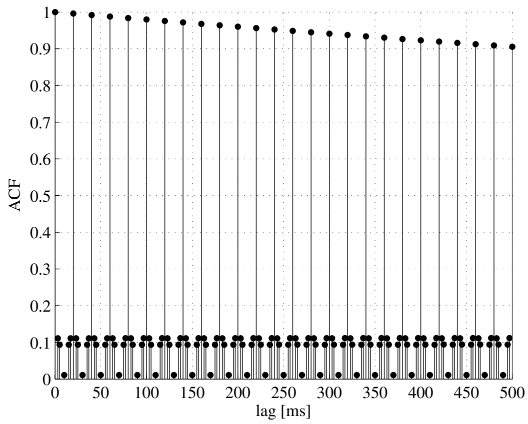
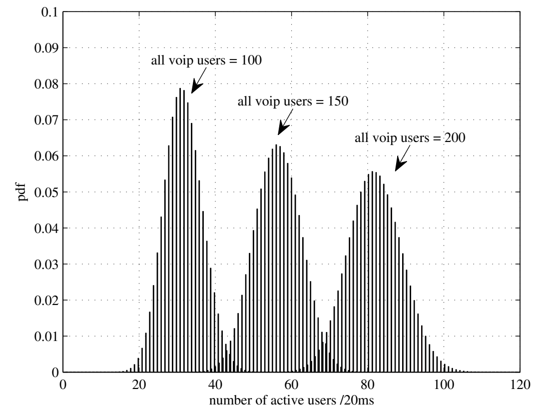
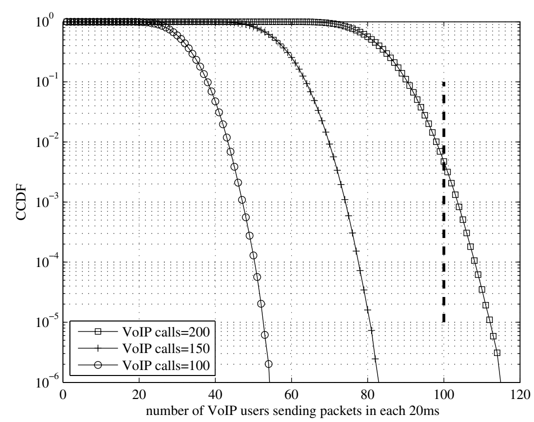


Fig. 4. Autocorrelation of user activities for the case of 200 VoIP calls



(a)



(b)

Fig. 5. Statistics of of the number of sending users in each 20 ms

the talking users may transmit their packets on different TTIs. The overall transmission sequence is the aggregation of all user traffic, and it is then nearly deterministic. Minor randomness appears from the fact that several users are switching from/to the talking to/from the silence state, TTI by TTI, but the state of major talking users are stable within the 20-ms-repeat. This indicates that the load estimation can be done to cover the total traffic in 20 ms other than every 1 ms. This means the system is still stable, if the data packets in the buffer can be transmitted before the peak of the next 20 ms.

Fig. 5(a) and 5(b) show the PDF and CCDF of active users summarized every 20 ms, respectively. From the PDF figure, it can be seen that the distribution is nearly normal distributed due to the central limit theorem, and the slobes are shifted to the right with an increasing number of VoIP calls. More importantly, on the CCDF figure, if the load estimation of 100 sending users/20 ms (meaning 5 sending users/ms in average) is given, it can already support the traffic from 200 VoIP calls with a probability greater than 99% ($\approx 1 - 0.5\%$)! With the consideration of PRB consumption per packet, e.g. 2 PRB/packet, the traffic of 200 VoIP calls can be supported by the spectrum estimation of 100 sending users/20ms corresponding to 5 sending users/1ms, which requires 10 PRBs/1ms, and this is only half of the 1ms-based estimation of 20 PRBs/ms.

In order to validate this conclusion, several simulations are performed with different resource budgets, ranging from $N = 8$ to N=20 PRBs estimation given by one VO. In order to show the influence The benchmark is selected as the packet delay of 200 parallel VoIP calls, because it is the most important QoS requirement of real time services. If the resources are sufficient, the performance should be good enough to satisfy the delay restriction of VoIP service of 50 ms [8] latency on air interface. The curves on Fig. 6 show the CCDF of the packet delay in logarithmic scale. It is clear that the number of PRBs $N >= 15$ is overprovisioned and with a budget of N=14 PRBs the VoIP transmissions are still very stable, meaning the delay performance of all of the packets are strictly within the 50 ms limitation. $N = 13$ is the critical value in this simulation scenario, where most VoIP packets ($\sim 99\%$) are having a delay less than 50ms and 1% of the packets have a delay ranging from 50 to 130 ms. The curves of the budgets < 12 PRBs are not shown, because the delay is monotonically increasing along with the simulation time. The reason is that in this simulation scenario, an average of 2$\sim$3 PRBs are required to support the transmission of one VoIP packet, which is greater than the example of 2 PRB/packet assumed in the above discussion. By employing advanced opportunistic scheduling combined with PHY techniques like MIMO (Multiple Input and Multiple Output), the resource requirement can be further reduced, and the 20ms-based load estimation proposed is much closer to the real traffic load than the 1ms-based estimation.

B. Discussions

Since all of the VoIP traffic use the same codec, the load contribution from VoIP traffic is proportional to the number of VoIP users. This means the mapping from the numbers of active VoIP users in one cell to the load estimation is

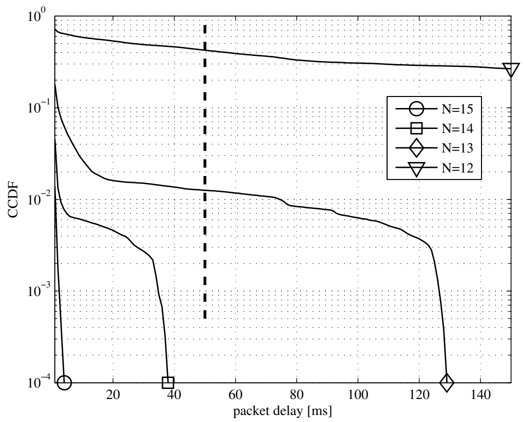


Fig. 6. Validation on resource budget based on 20 ms estimation: packet delay performance of 200 VoIP calls, *N*: number of PRBs reserved

straightforward. Some lookup tables of different number of active uses can be built up for the mapping. One argument could be whether there are enough resources left *in* the spectrum estimation for other services like FTP download, because in reality the system load is the mixture of the traffic from different services. To answer that, again the example in the previous subsection can be taken: 100 sending users/20ms is equivalent to 5 sending users/1ms. This setting is actually overprovisioned to cover more than 99% of the cases, which means some resources are left over for other services due to the fluctuation of VoIP traffic. Further from a scheduling point of view, best effort services like FTP can have very adaptive data rates with low sensitivity to the latency, depending on the channel quality and system load. Therefore, for designing a reasonable scheduler allocating the radio resources to users, VoIP traffic should have higher priority than FTP traffic. If the resource budget is defined by estimating the VoIP traffic, it guarantees the performance of the real time services and by adding proper margins other best effort services can be accommodated as well. The size of the margin depends on the business strategies of VOs as discussed in Section IV: the tradeoff of performance and operation expenditure. Nevertheless, this proposed mechanism for load estimation is purely based on the number of users having real-time services, not sensitive to the parametrization and not limited by the estimation interval, and it gives the baseline for the minimum requirements.

Further, more detailed VoIP traffic models will improve the statistics, especially the packets without header compression mentioned in Section II-B which require more resources per transmission and also interfere with the autocorrelation feature. Other realtime services, like video streaming requiring more bandwidth, can also be taken into account for load estimation in a similar way. One thing that is also worth mentioning here is that the required PRB per VoIP packet is an average value and can be extracted from the simulations for different scenarios.

VI. Conclusion

After the discussion in the previous sections, we can conclude that a large multiplexing gain can be obtained via spectrum sharing in LTE virtualization. This is validated both from analytical and simulative perspectives. Motivated by this, we proposed an extended model of multi-party spectrum sharing, for which the key issue of a practical application is the accurate estimation of the spectrum requirement to exploit the multiplexing gain. A new mechanism of spectrum estimation is then further proposed, and it is based on the traffic model of realtime services and can be easily applied with little limitations. The numeric results also proved the feasibility of this mechanism.

Some of the topics that we intend to investigate within our future research includes: spectrum sharing of LTE virtualization in multi-cell scenarios, especially for a highly mobile and interfered environment, the impact of using more detailed and heterogeneous traffic models as well as different practical scheduling algorithms, and finally analytical traffic load estimation following some classical models like [15], rather than the one obtained from the simulations as shown in Section V.

References

[1] S. Baucke and C. Görg. Virtualisation as a Co-existence Tool in a Future Internet. In *ICT Mobile Summit - 4WARD Workshop*, Stockholm, Sweden, June 2008.

[2] S. Paul and S. Seshan. GENI Technical Document on Wireless Virtualization. http://groups.geni.net, September 2006.

[3] AKARI architecture conceptual design for new generation network. http://akari-project.nict.go.jp, 2008.

[4] J. Sachs and S. Baucke. Virtual Radio-A Framework for Configurable Radio Networks. In *WICON'08*, Hawaii, USA, Nov. 2008.

[5] Y. Zaki, L. Zhao, A. Timm-Giel and C. Görg. A Novel LTE Wireless Virtualization Framework. In *Second International ICST Conference on Mobile Networks And Management (Monami)*, pages 1–13 CD publication), Santander, Spain, September 2010.

[6] Y. Zaki, L. Zhao, A. Timm-Giel and C. Görg. LTE Wireless Virtualization and Spectrum Management. In *Third Joint IFIP Wireless and Mobile Networking Conference (WMNC)*, Budapest, Hungary, October 2010.

[7] MVNO Directory. http://www.mvnodirectory.com, 2010.

[8] F. Khan. *LTE for 4G Mobile Broadband*. CAMBRIDGE UNIVERSITY PRESS, 2009.

[9] E. R. Vale, M. A. Grivet, and J. C. Brandão. Reducing the VoIP Packet Overhead in UMTS Air Interface. In *2009 2nd International Conference on Adaptive Science & Technology*, Accra, Ghana, December 2009.

[10] ROHC IETF working group. Robust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed, July 2007.

[11] VANU. http://www.vanu.com, 2009.

[12] Y. R. Zheng and C. Xiao. Simulation Models With Correct Statistical Properties for Rayleigh Fading Channels. In *IEEE TRANSACTIONS ON COMMUNICATIONS, VOL. 51, NO. 6*, June 2003.

[13] K. Lindberger. Balancing Quality of Service, Pricing and Utilisation in Multiservice Networks with Stream and Elastic Traffic. In *In Proc. of the International Teletraffic Congress (ITC 16)*, Edinburgh, Scotland, 1999.

[14] Network Sharing in LTE, Opportunity & Solutions. In *Technology White Paper, Alcatel-Lucent*, 2010.

[15] J. Digle and J. Langford. Models for Analysis of Packet Voice Communication Systems. In *IEEE JSAC, Vol. 4, No. 6*, 1986.