A Framework of Resource Provisioning and Customized Energy-Efficiency Optimization in Virtualized Small Cell Networks

Guolin Sun[1], Addo Prince Clement[1], Gordon Owusu Boateng[1], Wei Jiang[2,3]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China
Chengdu, 611731, China

[2] German Research Center for Artificial Intelligence (DFKI GmbH), Kaiserslautern, Germany

[3] Department of Electrical and Information Technology (EIT), Technische University (TU) Kaiserslautern, Germany

[e-mail: guolin.sun@uestc.edu.cn]
*Corresponding author: Guolin Sun

Abstract

The continuous increase in the cost of energy production and concerns for environmental sustainability are leading research communities, governments and industries to amass efforts to reduce energy consumption and global CO_2 footprint. Players in the information and communication industry are keen on reducing the operational expenditures (OpEx) and maintaining the profitability of cellular networks. Meanwhile, network virtualization has been proposed in this regard as the main enabler for 5G mobile cellular networks. In this paper, we propose a generic framework of slice resource provisioning and customized physical resource allocation for energy-efficiency and quality of service optimization. In resource slicing, we consider user demand and population resources provisioning scheme aiming to satisfy quality of service (QoS). In customized physical resource allocation, we formulate this problem with an integer non-linear programming model, which is solved by a heuristic algorithm based on minimum vertex coverage. The proposed algorithm is compared with the existing approaches, without consideration of slice resource constraints via system-level simulations. From the perspective of infrastructure providers, traffic is scheduled over a limited number of active small-cell base stations (sc-BSs) that significantly reduce the system energy consumption and improve the system's spectral efficiency. From the perspective of virtual network operators and mobile users, the proposed approach can guarantee QoS for mobile users and improve user satisfaction.

1. Introduction

Research on the fifth-generation (5G) mobile cellular communication technology indicates that the traffic density in crowded cities or hotspot areas will reach 20~Tbps/Km^2 in the near future. It is expected that by 2020, Mobile Internet will need to be delivering 1GB of personalized data per user per day. Furthermore, traffic by 2030 is predicted to be up to 10,000 times greater than in 2010 and 100 Mbps end-user services will have to be supported. To be able to support such demand, future mobile cellular networks are expected to be deployed in a very dense and multi-layered way. Ultra-dense small cell network (UDN) is considered as one of the most promising methods to meet the traffic capacity requirement of 5G [1]. The realization of this is simply done by the dense deployment of small cells in the hotspots, where immense traffic is generated [2]. However, this triggers a proportional consumption on energy. From the perspective of network operators, the increasing energy costs cannot sustain future network operations. From the environmental point of view, "greenness" can be more meaningful with a comprehensive evaluation that includes both energy savings and network performance, which is the basis for energy efficiency (EE) metrics [3].

Network virtualization has been proposed and regarded as a promising concept in the information and communications technology (ICT) area, where physical infrastructure can be simultaneously shared by several independent virtual operators (VOs) of mobile networks and service providers (SPs), who do not need to own network infrastructure to provide mobile internet services to users. This will reduce the overall capital cost and operational cost, and simplify network management. In order to provide the differentiated services in 5G, network virtualization is proposed as a main enabler in 5G radio access networks (RAN) [4]. A pragmatic EE and QoS optimization is required in 5G by infrastructure providers (InPs), VOs of mobile networks and user equipment (UEs). To address this, we propose a generic framework of slice resource provisioning and customized physical resource allocation for EE and QoS optimization. Our motivation is to improve EE and QoS to satisfy InPs and multiple identified VOs via flexible user association and cell activation techniques in dense long-term evolution (LTE) compliant small cells. Actually, energy saving with sc-BS activation has a lot of available work. However, this topic is the first time to be investigated together with network virtualization, because of new constraint for slice resource on each base station, which is the output of resource provisioning stage. This leads to a new optimization problem and needs a new algorithm to find optimal solutions. In this paper, our main contributions can be summarized as.

- We propose a generic framework of slice resource provisioning and customized physical resource allocation for EE and QoS optimization. On resource slicing and provisioning, we consider a resource provisioning scheme with resource demand and user population aiming to satisfy QoS.
- For customized physical resource allocation, we formulate the EE-QoS optimization problem as an integer non-linear programming model with the aim of satisfying diverse QoS requirements of UEs with the minimum amount of active sc-BSs. Because the formulated model is NP-hard, a low-complexity heuristic algorithm for minimum vertex coverage problem is proposed.

The rest of the paper is organized as follows. Section II presents the related works. Section III presents the system model segmented into business and virtualization model, user association model, network model and utility model. Section IV provides the problem formulation and our proposed heuristic algorithm. Simulation results and analysis are

discussed in Section V. The conclusion is drawn in section VI.

2. Related Works

Wireless network virtualization is considered as a significant entity in future 5G technologies. As such, a number of researches in both industry and academia have been geared towards this area. Dynamic slicing via flexible scheduling in LTE wireless network virtualization was presented in [5]. The resource allocation scheme considered fairness requirements for different SPs. Authors in [6] proposed network virtualization substrate (NVS), a substrate for virtualizing wireless resources in cellular networks, specifically, WiMAX. They emphasized on three key requirements namely: customization, isolation and efficient resource utilization. They proposed flow scheduling and slice scheduling schemes on the basis of priority and data rate per slice. In [7], *Mahindra et al* proposed NetShare, which considers radio resource management framework of the whole network in RAN. NetShare outlines a proportion between resource demand of each base station and the minimum resource allocation that is guaranteed to each entity across the network to achieve maximum utilization of resources. AppRAN, an application-oriented framework for RAN sharing in mobile networks was proposed by *He and Song* in [8]. AppRAN separates radio resources from mobile operators by providing services that satisfy QoS of users. To abstract only the radio resources and not the entire base station, Hap-SliceR was presented in [9]. Unlike NVS, NetShare and AppRAN, Hap-SliceR slices radio resources in a dynamic manner and causes little change to LTE standards and protocols for its implementation. The authors in [10] proposed a novel heuristic based admission control mechanism that is able to allocate network resources dynamically to different slices. Effective network slicing management and prioritization in 5G systems provides higher QoS in individual slices and increases utilization of network resources. *Vikas Jumba* proposed an optimal algorithm for resource provisioning to simultaneously satisfy the rate and resource requirements of slices in [11]. Resource provisioning means making network resources available to be utilized by users. To demonstrate the ability of admission control algorithm to efficiently adjust the minimum slice requirements based on priority, the authors in [12] presented a joint resource provisioning and admission control policy in virtualized wireless networks. The same authors proposed a robust resource provisioning policy in [13] to maximize the EE of virtualized networks while satisfying QoS of slices.

Several studies in recent times also suggested a scheme, known as multiple base station scheduling (MBSS) [14]. MBSS leads to the substantial EE gains in LTE-compliant mobile cellular networks. However, the short-coming with MBSS is its computational complexity. With the increasing number of ultra-dense nodes, computational complexity will affect the practical implementation of the energy-saving algorithms. A flexible flow scheduling based on MBSS was analyzed considering user activities and traffic dynamics in [15]. Unlike our work, all aforementioned approaches either do not take the QoS constraints into account or considered improved QoS at the expense of EE. In trying to improve EE, the QoS of users must not be sacrificed unnecessarily. The authors in [16] also argue that these two different effects cancel each other in UDN. In an attempt to reduce the cost of energy, these methods only considered switching off sc-BSs and ignored users and their QoS. Trade-offs between

QoS and EE for users with different traffics was presented in [17]. The authors defined voice, best effort and QoS traffic as different traffic with transmission power and network bandwidth as different power consumption metrics. An energy efficient algorithm was proposed for the downlink of HetNets using user association in [18]. However, this scheme only refers to the users with video content. The goal of minimizing the system energy consumption and also maximizing the ratio of the peak-signal-to-noise-ratio was considered in [19]. Joint optimization by cell activation or cell coverage adjustment, user association, and sub-carrier allocation has been investigated in [20] [21]. This was done under the constraints of maintaining an average sum rate and rate fairness. They argued that energy consumption is dependent on both the spatio-temporal variations of traffic demands and the internal hardware components of sc-BS. Some other works considered an optimization problem for EE with a guaranteed service rate for mobile terminals to improve their QoS. A heuristic algorithm to reduce energy consumption, taking advantage of user association, and maintaining certain QoS constraint was proposed by [22].

Although some works have considered energy efficiency and resource virtualization separately, few work has been done on joint optimization. Recently, *Z. Xu et al* studied the power efficiency of different network function virtualization (NFV) implementations in [23]. EE versus delay tradeoff in wireless networks virtualization was presented by *Q. Shi et al* in [24]. Since the implementation of a power saving strategy should consider multiple conflicting objectives, this work extends the results presented in traditional methods by introducing a multi-objective optimization framework, designed to manage the multiple network slices over dynamic traffic demands with a posteriori decision making support. The proposed joint dynamic resource provisioning and physical resource allocation approach can further reduce energy consumption and satisfy QoS requirements of users by serving dozens of users with minimal slicing resources through segregated functions of InPs and VOs.

3. System Model

The proposed joint resource provisioning, cell activation and resource allocation scheme aiming at network power minimization and QoS satisfaction introduces a centralized logical software defined networking (SDN) controller in a multi-tenant RAN. In our scenarios, there is one Macro-cell BS (MBS) and a set of sc-BSs. The MBS only operates at the control plane and the sc-BSs operate at the data plane. With this system architecture, it is very effective to turn off some sc-BSs without any adverse effect on the overall system. The data flow scheduling among sc-BSs is implemented by SDN controller. The system model is presented with business model, user association and energy model, network model and utility model.

3.1 System architecture

In this paper, we adopt a 3-level hierarchical system architecture, which includes InP, VOs and UEs, as shown in **Fig. 1**. Level 1 comprises the InPs who own and manage the network infrastructure. They have oversight responsibility over energy and radio resource management. Level 2 embodies the VOs. They have dedicated radio resources from the InPs and do not need to own the physical infrastructure. The UEs are in level 3. They do not deal directly with the InPs. Their interest is in the services provided by the VOs. Basically the InP partitions the physical radio resource it owns, into slices which are allocated to the VOs the

InP caters for. Each slice has sc-BSs with which UEs that belong to the particular VO associate. A slice is made up of individual UEs who the VOs provide services to. In a slice, there is an association between a UE and a sc-BS. Before a UE can be able to communicate with other UEs, it has to be associated with a sc-BS in the slice it belongs to. In result, VOs care about the satisfaction of users whiles InP targets minimizing embedding cost and efficient resource utilization.
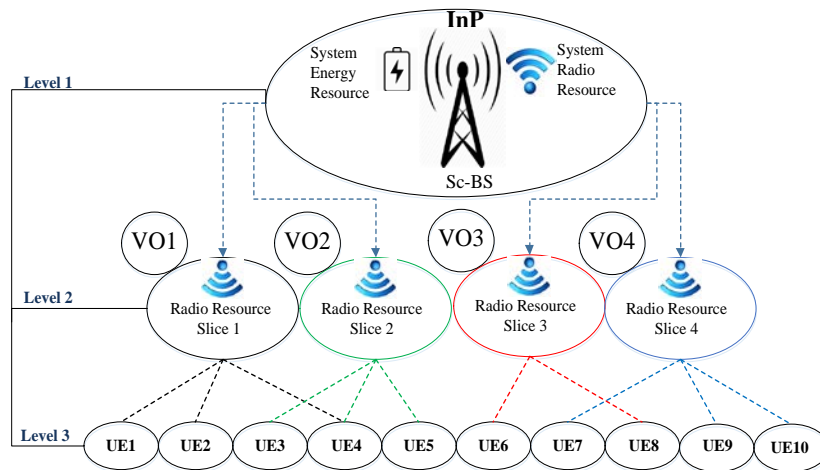

Fig. 1. Illustration of system architecture for virtualized cellular network

As a key part of the relationship between the actors in **Fig. 1**, InPs lease resources to VOs on contract base, which is known as service level agreement (SLA). SLA is a key characteristic of every level of resource virtualization and resource management, which includes deadlines, client QoS concerns and meeting the overall desired system objectives. A major problem is the reduction of operational governance control, i.e., VOs and UEs have less control over the actual service level being offered by the InPs, since the former do not own the infrastructure. As a result, the QoS is integrated in the SLA [25]. SLA is an important issue for both InPs and VOs, who require efficient SLA management from the complete SLA lifecycle perspective [26]. In this regard, authors propose a QoS-aware framework in [27] that provides automated runtime support for service discovery, negotiation, monitoring, and service provider rating, which help consumers handle recovery from SLA violations, service failures and runtime environment limitations by renegotiating and substituting problematic services. In our framework, QoS, resource occupation, service failure and recovery are the key SLA between the aforementioned actors. Confidentiality capabilities and trustworthiness are additional aspects of SLA between VOs and UEs.

Virtual resources are created by splitting physical resources into multiple virtual slices. These radio resources are allocated to the independent VOs with an example, shown in **Fig. 1.** In view of resource utilization, the system energy is shared among virtual operators while the system radio resource is sliced into multiple slices based on the user demands. Sharing the limited energy among all of the VOs in the system fulfills the EE dimension of virtualization. The radio resource is shared among VOs who serve UEs directly based on their QoS demands. User association is granted based on slice priorities. Each VO serves its UEs independently. From this decoupled model, the VOs can provide tailored services that will improve user QoS. The InPs can focus on infrastructure and physical resource management. The UEs, based on their demands can subscribe to the VO with the best offer.

3.2 User association and EE model

In our scenario, an association between sc-BS set and UE set is discussed. The sc-BSs have two states, active and inactive. An active state denotes a sc-BS with an associated UE and inactive if otherwise. We define four relationships between UEs and sc-BSs. These are association, no-association (idle), multi-association and handover. The "association" defines relationship where UE successfully associates with an active sc-BS. "No-association" defines the relationship where no UE is associated with a particular sc-BS. Two leading factors result into a "no-association" relationship. First, if there are no UEs in the cell coverage of a particular sc-BS and secondly, if the UE in the cell coverage of a particular sc-BS has other candidate sc-BSs with adequate radio resource. This leads to the third relationship called "handover" shown in **Fig. 2**. With this relationship, UEs are aggregated on minimum set of active sc-BS to conserve energy. Finally, we also have "multi-association" relationship. In trying to maximize the resource utilization of active sc-BSs, UEs are allowed to amass the resources on the same kind of slice on two different active sc-BSs. This further improves the QoS even for the edge UEs. Our objective is to satisfy UEs with minimal system energy as shown in **Fig. 2.** This is achieved by effective flow aggregation using handover.
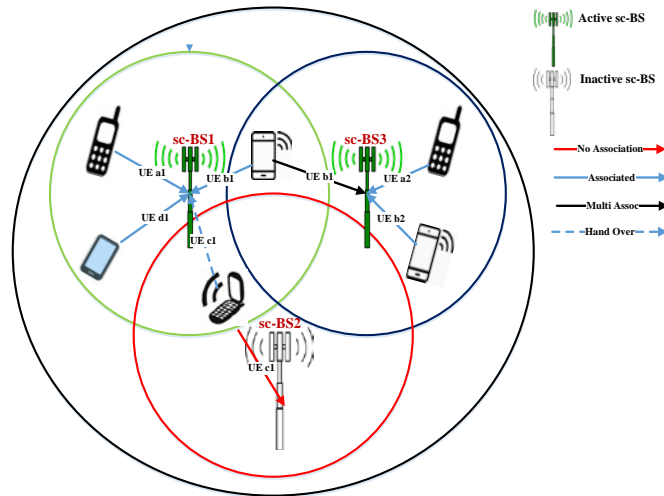

Fig. 2. Flexible user association model for EE

In this example, UEs (a, b, c, d) represent UEs from different VOs demanding radio resources from the sc-BS. Based on the distance, Signal-to-Interference-plus-Noise Ratio (SINR) and available resource, UEs a1, b1 and d1 associate with sc-BS 1. This creates 3 slices on sc-BS 1. UE b1 further associates with sc-BS 3 creating a "multi-association" with sc-BS 1 and sc-BS 3 to satisfy its QoS demand. UEs a2 and b2 associate with sc-BS 3 and create two slices since the UEs are of two different operators. When UE b1 associates with sc-BS 3, no new slice is created since UEs b1 and b2 are of same operator type. UE c1 originally associates with sc-BS 2. Because sc-BS 1 is already active and has resources to support UE c1, handover occurs and sc-BS 2 is successfully shut down. Because UE c1, from a different VO, associates with sc-BS 1, a fourth slice is created on sc-BS 1.

We define three power consumption states as *sleep*, *idle* and *active*. The active state combines the sum of consumption in receiving and transmitting statuses. Power consumption

(kw/hr) at the sleep state is negligible and subsequently eliminated from our model. Given a
set of scan periods, $c \in \{1, 2, 3, \dots, C\}$, a set of active sc-BSs $z \in \{1,2,3, \dots, Z\}$ and
$y \in \{1,2,3, \dots, Y\}$ for idle sc-BSs, let $\mathcal{p}_{active}$ and $\mathcal{p}_{idle}$ denote energy consumption in active
and idle state respectively. We define the energy consumption ∂ in equation (1) as;

$$\partial = \sum_{c=1}^{C}\left(\sum_{z=1}^{Z} \mathcal{p}_{active} + \sum_{y=1}^{Y} \mathcal{p}_{idle}\right) , \qquad (1)$$

From the control plane, an energy manager controls sc-BS switching and UE associations.
The energy manager can turn on a sleeping sc-BS, if some flow is located inside its
coverage and turn off the sc-BS if otherwise or after a handover. In order to reduce energy
consumption and latency, the energy manager scans each sc-BS in a fixed period, e.g. half-
an-hour, and sends a switch-off command to the idle sc-BSs. However, a sc-BS is switched
on if a flow enters the coverage of a sleeping sc-BS and has no access to any other active sc-
BS. A sc-BS will enter sleep mode again, if there is no transmit or receive command from
its neighboring UEs. The scanning period can be configured by the controller via secure
sockets layer (SSL).

3.3 Network Model

Given a set of infrastructure nodes, in terms of sc-BS, $j \in \{1,2,3,\dots,J\}$ and a set of VOs,
in terms of slices, $k \in \{1,2,3,\dots,K\}$ located within a given area, the whole system has a set
of mobile UEs, in terms of UE $i \in \{1,2,3,\dots,I\}$. Let a_{ij} denote the binary association
indicator between UE i and sc-BS j, where $a_{ij} = 1$ indicates UE i associates with sc-BS j;
otherwise, $a_{ij} = 0$. Let o_{ik} denote the binary association indicator between UE i and slice k,
where $o_{ik}=1$ means that UE i associates with slice k; otherwise, $o_{ik}=0$. Practically, each UE
associates with only one slice on one sc-BS; thus;

$$\sum_{k=1}^{K} \sum_{j=1}^{J} o_{ik}.a_{ij} = 1 , \qquad (2)$$

$$a_{ij} = \begin{cases} 1, if\, UE\, i\, is\, associated\, with\, \mathrm{sc} - \mathrm{BS}\, j \\ 0, \qquad\qquad\qquad otherwise \end{cases} \qquad (3)$$

$\beta_{ij} \in [0,1]$ is used to denote the percentage of radio resource allocated by sc-BS j to UE i.
Certainly,

$$\sum_{i=1}^{I} a_{ij}.\beta_{ij} \leq 1 , \qquad (4)$$

which means, the percentage of radio resource allocated by sc-BS j to UE i cannot exceed
100%.

The radio resource allocated to sc-BS j is W_j measured in Hz. A fixed equal power
allocation mechanism is used; where the normalized transmit power on sc-BS j is P_j in
watts/Hz. Therefore, by using Shannon theory, the spectrum efficiency b_{ij} for the link
between UE i and sc-BS j with a successful association is expressed as:

$$b_{ij} = \log_2\left(1 + r_{ij}\right) , \qquad (5)$$

where $r_{ij} = g_{ij}P_j/\left(\sum_{l,l \neq j} g_{il}P_l + \delta\right)$ is the SINR between UE i and sc-BS j, considering the
interference from the neighboring sc-BS l. δ is the power spectrum density of additive white
Gaussian noise. g_{ij} is the large-scale channel gain that includes the path-loss and the
shadowing effect and P_l is the normalized power on sc-BS l. Therefore, the system
throughput without slicing is determined as:

$$T = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij}.\beta_{ij}.W_j.b_{ij} , \qquad (6)$$

Considering the minimum radio resource amount ρ_{ij} for satisfying QoS of UE i, we can deduce equation (7) from (6), where D_i is the aggregated data rate or throughput demand from UE i as.

$$\rho_{ij} = D_i/b_{ij} , \qquad (7)$$

In our scenario, the radio resource needed by UEs may be unlimited, so we can allocate resource considering fairness to all of the associated UEs. With a logarithmic utility function as the objective in order to maximize the total system capacity, an equal radio resource allocation is optimal for the fairness proportional scheduling [28]. Therefore, we ignore the radio resource constraint and consider β_{ij} as $1/N_j$, where N_j is the number of UEs on active sc-BS j and is equal to $\sum_{i=1}^{I} a_{ij}$. When some sc-BSs do not have any associated UE, they are shut down and N_j will be 0. To avoid N_j being zero, we replace $1/N_j$ with $1/(N_j + 1)$.

We consider the slice throughput achieved with minimum utilization of resources with respect to the result of resource provisioning, S_{kj}. S_{kj} refers to the available radio resource of the kth slice on sc-BS j .We can express the throughput of a slice as:

$$T_k = \sum_{i=1}^{I} \sum_{j=1}^{J} a_{ij} . o_{ik} . \beta_{ij} . S_{kj} . W_j . b_{ij} , \qquad (8)$$

where $\beta_{ij} \in [0,1]$ is used to denote the percentage of radio resource allocated by the jth sc-BS to user UE i. S_{kj} denotes the radio resource (S) provisioning of the kth slice on sc-BS j.

3.4 Utility model

In order to offer better QoS to UEs, the required transmission rate should be guaranteed. We consider measuring the satisfaction of a UE i with a sigmoid function, which can be expressed as [29]. The whole formula is used for defining the QoS of users as:

$$U_i(R_i) = \frac{1}{1+e^{-\tau_i(R_i - R_i^{min})}} , \qquad (9)$$

where R_i^{min} is the minimum throughput demand required by the UE i and τ_i is a constant deciding the steepness of the satisfactory curve. In addition, R_i is the average user throughput for UE i, which is determined by the network infrastructure, transmission power, noise, interference and many other related factors. It is easy to verify that: 1) $U_i(R_i)$ is a monotonic increasing function with respect to R_i, because individual UEs will feel more satisfied if they receive higher throughput above their minimum demand and vice versa; 2) $U_i(R_i)$ of each UE i is scaled between 0 and 1, *i.e.* $U_i(R_i) \in (0, 1)$.

4 Problem Formulation and Algorithm

4.1 Dynamic Resource Provisioning

In resource provisioning, slices are classified based on user requests. UE i prioritizes sc-BS j based on user demand (UD) or user population (UP). To prioritize slices from the highest to lowest, we define a weight Ω_k with UD and Φ_k with UP for slice k. Slices are allocated resources at their respective sc-BSs based on the network budget or weight assigned to the slice. Assuming data rate demand of UE i (D_i) of each slice is equal, a slice's weight based on UD and UP at a sc-BS can be defined respectively as follows:

$$\Omega_k = \sum_{i=1}^{I_k} D_i \qquad\qquad \forall i \in I_k, \qquad (10)$$

$$\Phi_k = I_k ,\tag{11}$$

This implies that, each slice can obtain its weight-based UD ratio ($\Omega_{rat_{kj}}$) and UP ratio ($\Phi_{rat_{kj}}$) at a particular sc-BS j considering the weights of other slices sharing the same BS. The weight-based UD and UP ratios of a slice k can be expressed as follows:

$$\Omega_{rat_{kj}} = \frac{\Omega_{kj}}{\sum_{k=1}^{K} \Omega_{kj}} , \qquad \forall\, k \in \{1,2, \dots, K\}\tag{12}$$

$$\Phi_{rat_{kj}} = \frac{\Phi_{kj}}{\sum_{k=1}^{K} \Phi_{kj}}, \qquad \forall\, k \in \{1,2, \dots, K\}\tag{13}$$

Proposition 1：*Considering the overall resource provisioning, the resource demand must be proportional to the weight-based user demand ratio (Ω_rat_{kj}) or weight-based user population ratio (Φ_rat_{kj}) of slice k to sc-BS j.*

To allocate radio resources to satisfy the resource demand of each slice at every sc-BS, the needed radio resource of slice k in sc-BS j is determined as follows:

$$S_{kj} = \frac{\Omega\_rat_{kj}}{\sum_{k=1}^{K} (\Omega\_rat_{kj})} \times W_j , \quad \forall\, k \in \{1,2, \dots, K\},$$

(14)

$$S_{kj} = \frac{\Phi\_rat_{kj}}{\sum_{k=1}^{K} (\Phi\_rat_{kj})} \times W_j, \quad \forall\, k \in \{1,2, \dots, K\},\tag{15}$$

where W_j denotes the total amount of resource allocated to sc-BS j and S_{kj} denotes the amount of resources needed by a slice k at sc-BS j. In calculating the spectral efficiency x_{kj} of the radio resources on each slice, the sum of individual UEs' gained throughput Th_i with respect to the radio resource allocated per slice, S_{kj} is given by:

$$x_{kj} = \sum_{i=1}^{I} Th_i / \sum_{j=1}^{J} S_{kj},\tag{16}$$

4.2 Customized physical resource allocation

In this section, we formulate an optimization problem for energy optimal user association and resource allocation. We aim to use cell activation, flexible user association and radio resource allocation to satisfy QoS of UEs while minimizing the energy consumption. Resources allocated to a slice are based on UD of UEs in the slice and the amount of resources needed by a slice to satisfy the QoS requirements of its users as expressed in equations (10) and (14) respectively. We assume that sc-BSs can change their on-off states by the energy manager on the SDN controller. Inactive sc-BSs can be switched on when they have some UEs ready for association. Active sc-BSs can be switched off when they have no associated UEs or when the associated UEs can re-associate with neighboring active sc-BSs. In our network slicing scenario, our objective is to minimize the total network power consumption as well as the number of active sc-BSs, while satisfying QoS requirements of VOs. The result is that, the UEs can be associated with their respective slices on at least one sc-BS, which can satisfy the demand D_i of UE i.

With the network model in equation (7) and (8), we can define our objective function as:

$$Min \sum_{i=1}^{I} \sum_{k=1}^{K} \sum_{j=1}^{J} \frac{o_{ik} . a_{ij}}{N_{kj}+1} . S_{kj} \, , \tag{17}$$

subject to

$$\vartheta_i = \frac{S_{kj} . o_{ik} . a_{ij}}{N_{kj}+1} log_2\left(1 + r_{ij}\right) \geq D_i \quad \forall i \in I \, , \tag{18}$$

$$\sum_{j=1}^{J} a_{ij} \geq 1 \quad \forall i \in I \, , \tag{19}$$

$$N_{kj} = \sum_{i=1}^{I_k} a_{ij} \, , \tag{20}$$

From constraint (18), we deduce that, for any successful QoS-based user association, the achieved throughput of UE i cannot be less than D_i of UE i, where D_i denotes the QoS requirement of UE i. Constraint (19) defines that each single UE can associate with at least one sc-BS, but when D_i of UE i cannot be met by the radio resource of a single sc-BS, UE i can be associated with two sc-BSs simultaneously satisfying its QoS requirement. In this case, there is no need to turn on additional sc-BSs. In constraint (20), N_{jk} is the number of UEs associated with slice k on active sc-BS j. In order to evaluate the system performance, we define the system resource utilization φ as

$$\varphi = \frac{\sum_{k=1}^{K} \sum_{j=1}^{J} S_{kj} . u_j}{\sum_{j}^{J} W_j . u_j} \, , \tag{21}$$

where S_{kj} is the total allocated radio resource of slice k on sc-BS j and W_j is the total radio resource of sc-BS j . u_j is a binary indicator showing the status of sc-BS j, where $u_j = 1$ means that sc-BS j is active; otherwise, $u_j = 0$. Furthermore, the average throughput (ψ_k) of UEs on slice k is

$$\psi_k = \frac{\sum_{i=1}^{I} \vartheta_i}{I_k} \, , \tag{22}$$

where ϑ_i represents the actual throughput gain from equation (18) and I_k is the number of UEs for slice k. The whole formula reflects the average QoE gain by UEs and the satisfaction provided by the VO.

The objective function in (17) shows an integer non-linear programming (INLP) problem. With parameters of algorithms defined, attention can be turned to describing common algorithms for solving INLPs. The algorithm shares many general characteristics with the well-known branch-and-bound or branch-and-cut methods for solving MILPs. Branch-and-bound is a divide-and-conquer method where the dividing (branching) is done by partitioning a set of feasible solutions into smaller subsets. A subset is maintained if an optimal solution is in the subsets and a subset is discarded if its bound indicates that it cannot contain an optimal solution. Computational complexity of the branch-and-bound methods has exponential growth with growing scale, which solves our optimization problem very slowly. Therefore, we transform it into the minimum vertex coverage (MVC) problem in bi-partite graph.

4.3 The proposed heuristic algorithm

Definition 1: *Coverage bi-partite graph*

The coverage bi-partite graph is based on the weights on the edges, which is defined in terms of the SINR between a sc-BS and a UE. In the coverage bi-partite graph, the lines connecting the sc-BSs and the UEs are referred to as the edges. We denote the SINR-based weights by

r_{ij}, where i is the particular UE and j is the particular sc-BS. If $r_{ij} > UE\ sensitivity\ (dBm)$, it means an edge exists between the sc-Bs and the UE; otherwise, an edge does not exist between them.

We construct bi-partite coverage graph-based signal propagation, as shown in **Fig. 3(G1)**. Given an undirected coverage graph *G1 (V1, V2, E, r)*, *V1* is the set of all active sc-BSs. *V2* is the set of mobile UEs. *E* is a set of edges each connecting UE and sc-BS. The weight r_{ij} on each edge is defined as the SINR between UE i and sc-BS j. Furthermore, we need to check whether the resource demands of UEs are satisfied or not. Therefore, we transform the bi-partite coverage graph *G1* to the effective radio resource demand graph *G2*. We provide an example to explain the transformation from *G1* to *G2* in **Fig. 3.**

Definition 2: *Demand bi-partite graph*

The demand bi-partite graph is also based on the weights on the edges, which is defined in terms of the required radio resource needed on a sc-BS to satisfy the QoS of a UE. We define the resource-based weights by ρ_{ij}, where i is the particular UE and j is the particular sc-BS. If $\rho_{ij} < \frac{D_i}{b_{ij}}$, it means an edge exists between the sc-Bs and the UE; otherwise, an edge does not exist between them.
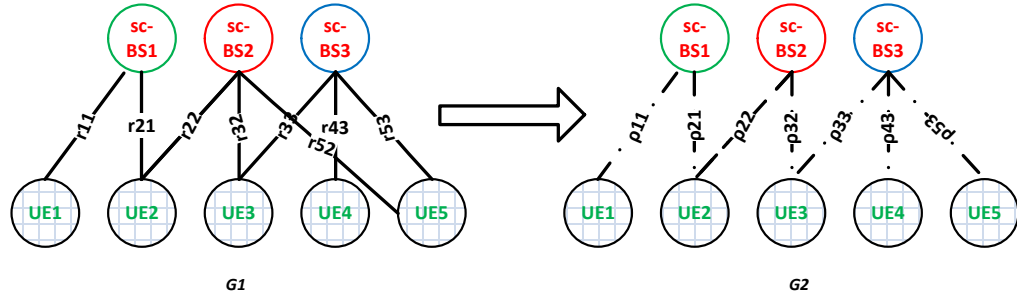

Fig. 3. Problem Transform with constraint bi-partite graph coverage

Proposition 2: *Minimized number of active sc-BSs can be achieved by finding a minimum vertex coverage on demand bi-partite graph to satisfy most users step by step.*

Based on graph theory, we know minimum vertex coverage problem is NP-hard. We consider a heuristic algorithm based on greedy idea to solve this problem. As an example, shown in **Fig. 4**, we update the original resource demand graph in **Fig. 4a** to a new graph in **Fig. 4b** without UE4 and UE5 since they have already found their association. The same iteration is repeated in the new sub-graph until all of five UEs find their associated sc-BSs.

The proposed detailed algorithm is summarized in **Fig. 5**. During each iteration, we find a sc-BS which has maximum degree, so we need to sort the sc-BS vertex-based degree in $O(J * log\ J)$ complexity time. In a worst case, the entire sc-BS vertices are included to cover UE vertex, hence the algorithm's main loop iterates *J* number of times. Because we need to find the associated sc-BSs for all the UEs, inside every *J* loop, there is *I* number of computations. Based on the above analysis, the total time complexity of the proposed heuristic algorithm is $O(I * J^2 \log J)$.
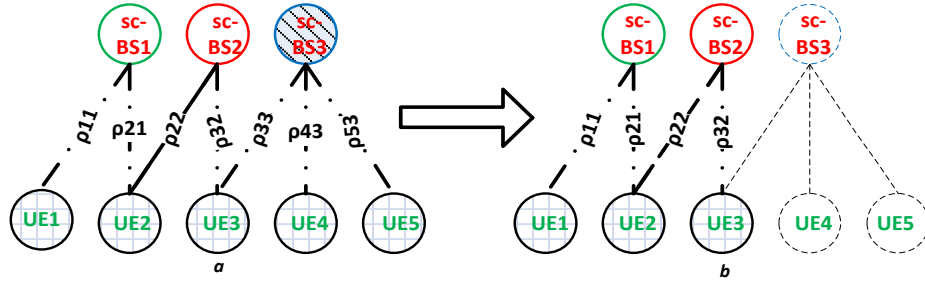
Fig. 4. Minimum Vertex Coverage (MVC) model for cell activation

Algorithm 1 Minimum Vertex Coverage Based Cell Activation

1	**Initialize** I, K, J, E, D_i;
2	Slice allocation:
3	Allocate resources to slice k based on Eqn. (10) and Eqn. (14)
4	**Calculate** the priority of each network slice;
5	pque = priority_queue(1,2,3,···,K);
6	**For** each sc-BS $j \in V_1$, the set of active sc-BSs
7	j.user = {};
8	**While** pque.empty() == false
9	k = pque.pop();
10	Update V_2, the set of users for slice k;
11	Build a new bi-partite graph G = (V_1, V_2, E, ρ);
12	**While** $V_2 \neq \emptyset$
13	Choose a vertex sc-BS $j \in V_1$ of maximum degree in graph G
14	Find an associated user UE $i \in$ UE ij of sc-BS j with minimum degree;
15	**While** the available resource on sc-BS j can satisfy user UE i or UE $ij \neq \emptyset$;
16	j.user = j.user $\cup$ {UE i};
17	V_2 = V_2 - {UE i};
18	Update the available resource of sc-BS j;
19	Find another associated user UE $i \in$ UE ij with minimum degree;
20	**End**
21	**If** available resource on sc-BS j is exhausted
22	j.user == $\emptyset$, switch off the sc-BS j;
23	V_1 = V_1-{sc-BS j};
24	**End**
25	**If** V_2 does not change
26	**While** any UE $i \in V_2$ can be satisfied by sc-BS j_1,sc-BS $j_2 \in V_1$
27	Update available resources on sc-BS j_1 and sc-BS j_2;
28	j_1.user = j_1.user $\cup$ {UE i};
29	j_2.user = j_2.user $\cup$ {UE i};
30	V_2 = V_2 - {UE i};
31	Update V_1;
32	**End**
33	**If** $V_2 \neq \emptyset$
34	Switch on inactive sc-BS j nearest to user UE i;
35	add sc-BS j to V_1;
36	**End**
37	**End**
38	**End**
39	**End**
40	**End**
41	**Output** all of association set j.user and sc-BS status

Fig. 5. The proposed heuristic algorithm

   The priority based slice scheduling scheme is used. In Step 1, a set of parameters about the network are initialized. These parameters are; the number of sc-BSs J, the number of slices K, the number of UEs I, the set of edges connecting UEs and sc-BSs E and the user demand of one user D_i. In step 2, we schedule slices based on the predefined priority in equations (10) and (11) and update available radio resources and user proportion in each

scheduled period. Then we build a bi-partite graph. In Step 3, we provision resources to the individual slices in the network based on equations (10) and (14). In Step 4, in line 11~35, we can find a minimum vertex coverage with the proposed heuristic algorithm for each slice, in which the scenario of multi-association is considered. Firstly, we consider only one sc-BS satisfying the QoS of UEs in line 11~19. However, in line 20~23, we switch off the sc-BSs with an association between UE and sc-BS j set as $\emptyset$ .We continue from line 24~31 by considering the case where one sc-BS is not enough to satisfy the demand of a user. The use of two sc-BSs is employed and multi-association is discussed. Finally, we switch on nearby inactive sc-BSs and add them to the set of active sc-BSs when two sc-BSs cannot still satisfy the demand of UEs. The output of the algorithm is the association user set and the status of the sc-BSs.

5. Performance Evaluation

5.1 Scenario configuration

To evaluate the performance of our proposed algorithm, we perform the numerical simulations using MATLAB. The simulation parameters follow the 3GPP Typical Urban environment and are listed in **Table 1**, unless otherwise specified **[30]**.

Table 1. Simulation Parameters

Parameter and units	Values
Type of small-cell BS	3GPP LTE sc-BS
Number of Macro-cell BS	1
Number of small-cell BSs, J	19 hexagonal grid, wrap-around
Number of UEs, I	25-300
Number of slices, K	4
Power consumption (Idle state), $\wp_{idle}$	9.2dBm
Power consumption (active state), $\wp_{active}$	19.3dBm
User demand per slice (Mbps), Ω_k	1; 5; 1.5; 2
Distance of neighbor sc-BSs(m), d	100
Transmit power per sc-BS (Watt), P_j	20
System Bandwidth (MHz), W_j	20
Noise PSD, δ	-174 dBm/Hz
Carrier Frequency band(GHz)	2.0
Shadowing effects(dB)	8
Number of interfaces per UE	2
UE sensitivity (dBm)	-120;
PathLoss model	15.3 + 37.6log(d) (3GPP Typical Urban)
Slice priority	UP and UD

With the specified actors in the defined system models, the controller takes over the scheduling, control and oversight of monitoring traffic dynamics. There are 4 virtual operators according to which we have 4 slices of the radio resource. The radio resource requirements of all the UEs in one slice are equal. During the simulation, we do not consider the MBS with the assumption that all sc-BSs are attached to it. As specified in network models, we configure 19 LTE sc-BS in a hexagonal grid wrap around with a distance of 100 meters apart. The number of UEs ranges from 25 to 300 and iterated incrementally with 25

UEs in each scan period (c). To be more explicit in our work, we adopted the 3GPP LTE sc-BS for our simulation. The bandwidth of BSs is set at 20MHz. The SINR threshold of UE sensitivity is set at -120dBm for edge UEs. As specified in association model, each UE has a maximum of two interfaces. The energy consumption largely depends on the transceiver power settings, traffic load and the active duration of sc-BS. As specified in the energy consumption model, the transmitter power of sc-BS is set as 19.3dBm, and 9.2dBm respectively for the active, and idle states. The transmitter power in sleep state is negligible, so we ignore it in our model. In addition, two utility functions are adopted in terms of resource utilization in equation (21) and QoS satisfaction in equation (9).

From here, a typical UE distribution model is presented in **Fig. 6.** The different colors show UEs of different slices. From a random distribution of UEs with the associated priority based on UD, a closer look at **Fig. 6** reveals a significant number of sc-BSs with no associated UEs and are therefore shut down to conserve energy. In order to evaluate our proposed model and algorithm, we define five different use cases. The network topologies in our simulation include overlapping sc-BS set per slice with uniformly distributed UEs. We define Case I as simple on-off cell activation with the simple nearest-sc-BS association without slicing, which is also called the simple on-off scheme. We define Case II as the cell activation with the branch and cut algorithm. However, the computational complexity makes it difficult to work when user population is large. In Case III, we configure equal resource (static slicing) for all slices on a common sc-BS and MVC based cell activation algorithm is used, identified as MVC with static slicing (MVC-SS). In Case IV, we configure radio resource on a common sc-BS for multiple slices based on UD per slice in a dynamic way, which is identified as MVC with UD-based dynamic slicing (MVC-UD). Case V is similar to Case IV, but the dynamic slicing here is based on UE proportion (UP) per slice, which is identified as MVC with UP-based dynamic slicing (MVC-UP).
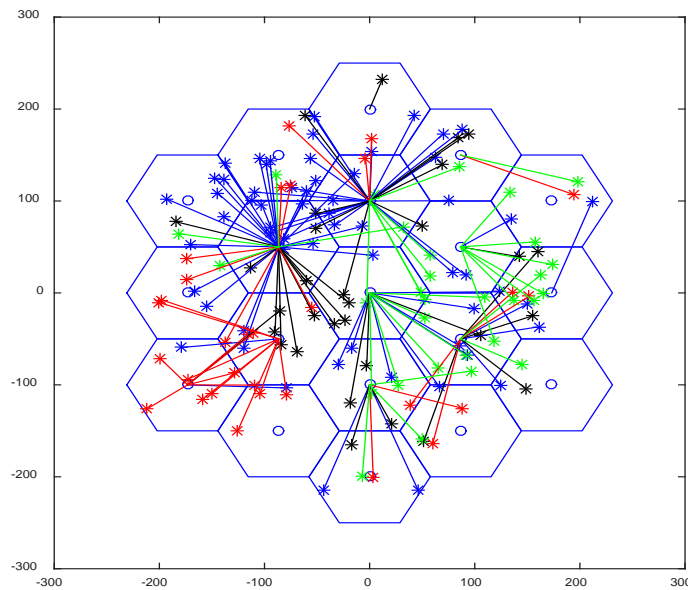

Fig. 6. Topology and UE distribution models
(Different colors show UEs of different slices)

Simulation results are classified under the different actors in the joint cell activation, user association and virtual resource allocation for EE and QoS optimization in virtualized small cell network. The InP is concerned with energy cost minimization while VOs and UEs focus on user satisfaction maximization. From the perspective of InPs, we present results in terms of the number of active sc-BSs, energy consumption, resource utilization view on each sc-BS, and spectral efficiency. From the perspective of the VOs and the UEs, we focus on handover and user satisfaction.

5.2 InP EE perspective

The result presented in **Fig. 7** is the number of active sc-BSs against the traffic volume in terms of UE population. The scenario is configured with an overlapped sc-BS set with uniformly distributed UEs. The traffic changes lead to the status of sc-BS changing among active, idle and sleep. The results of the number of active sc-BS, as illustrated in **Fig. 7**, show that the number of active sc-BS correlated positively with the traffic volume.
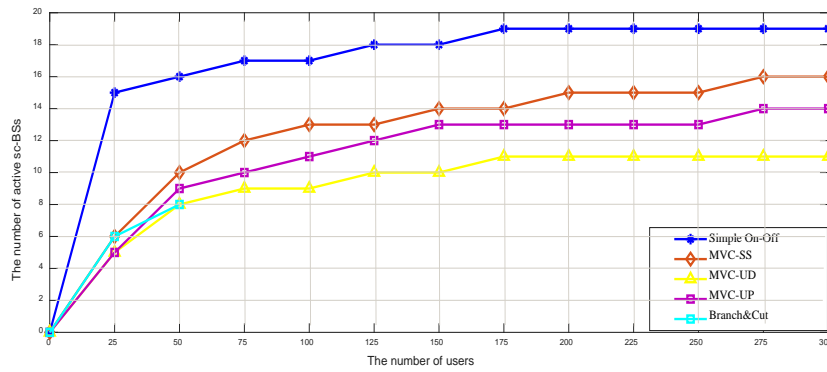

Fig. 7. Number of active sc-BSs with increasing user population

In all 5 cases, an increase in traffic resulted in relatively proportional increase in the number of active sc-BS. All other four cases outperform Case I. The gain is even more in UE-sc-BS association based on the resource demand per slice (MVC-UD). While the simple on-off increases active sc-BS up to 19, MVC-UP increased to 14 with the same traffic profile. The gain in MVC-UD is much gloomy with as low as 11 active sc-BS to serve the same traffic served by 19 active sc-BS with the simple on-off scheme and 16 active sc-BS with the MVC-SS scheme.

To illustrate the achieved energy consumption of the proposed algorithm, a simulation is done with three slicing schemes, MVC-SS, MVC-UD, and MVC-UP compared with the existing branch and cut scheme against simple on-off scheme. The total energy consumption is compared in w/hr. The result is illustrated in **Fig. 8**. The increasing rate in active sc-BS set as illustrated in **Fig. 7** manifested in the related energy consumption. The proposed MVC-UD algorithm outperforms the others and closely near to the branch &cut scheme, however immune to complexity problem. Because UEs were satisfied with significantly less active sc-BS in MVC-UD, the system energy consumption was equally significantly less. MVC-UD recorded up to 140w/hr, whiles simple on-off recorded upto 220w/hr. With low traffic, there is no significant difference in the cases under our proposed algorithm. The difference of gain is shown with increasing traffic.
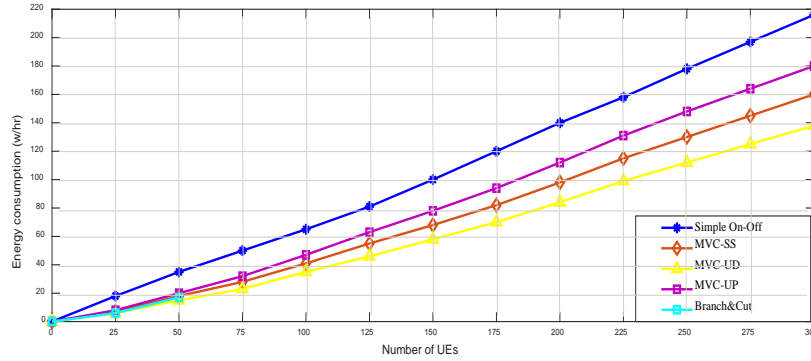
Fig. 8. Energy Consumption with increasing user population

In this simulation, we consider the gains of radio resource virtualization on resource utilization per slice, spectrum efficiency and user satisfaction. These metrics are compared for dynamic UD and static resource allocation respectively based on our algorithm. We set the number of sc-BS for each slice (S_{kj}) as 19.

In **Fig. 9a** and **Fig. 9b**, we can see the radio resource occupied on each sc-BS by the four slices. With the resource constraint for each slice, MVC-SS shown in **Fig. 9a** cannot achieve better resource utilization. With 16 out of 19 active sc-BSs, only one sc-BS's radio resource is nearly-fully utilized with some other active sc-BS with as low as 50% utilization. The MVC-UD shown in **Fig. 9b** maximizes the utilization of available resources on fewer active sc-BS, since each slice has the flexibility with the use of available resources on active sc-BS. These two scenarios on the resource view are shown in **Fig. 9a** and **Fig. 9b** respectively. In this simulation, we presented two kinds of slice prioritization. Prioritization based on sum of total system UD and based on UP. These are indicated as MVC-UD and MVC-UP, respectively. From the results of active sc-BS set, system energy consumption and resource utilization, it is obvious MVC-UD outperforms MVC-UP. We therefore focus on MVC-UD for two topologies against MVC-SS in the following result.
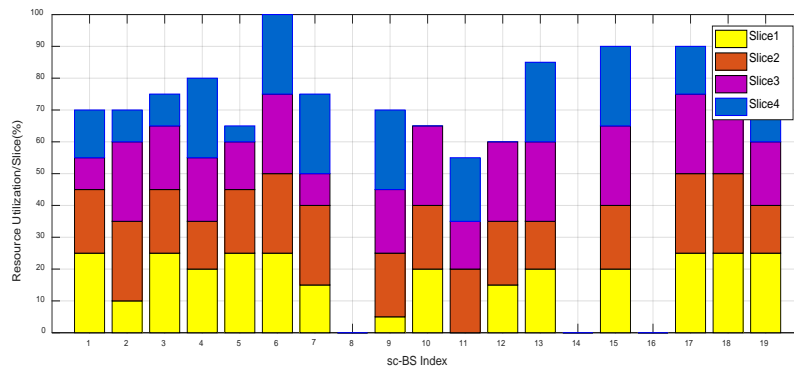

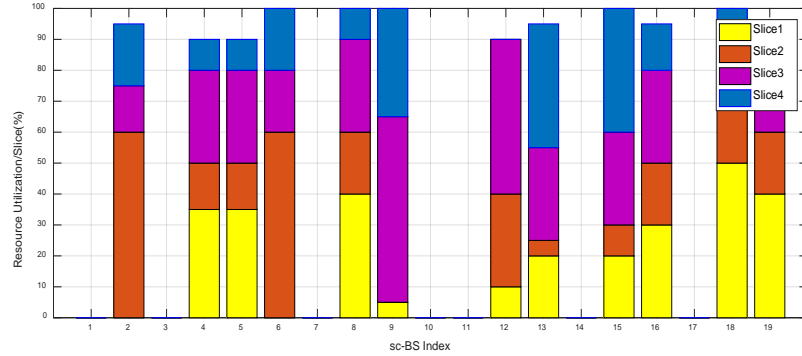**Fig. 9a.** Static bandwidth allocation MVC-SS

Fig. 9b. Dynamic bandwidth allocation MVC-UD

The efficiency of two different topologies of overlapped sc-BS (clustered and uniformly distribution UEs) is analyzed for MVC-UD and MVC-SS schemes shown in **Fig. 10.** MVC-UD is compared with MVC-SS under Clustered UEs (C) and for uniformly distributed UEs (U). The comparison gave a clear understanding on which of the two topologies best supports the different schemes and with what amount of traffic. The system resource utilization is rapidly improved with MVC-UD under clustered UEs configuration in low traffic. This is evident that, with a higher concentration of UEs in a limited area, the possibilities of optimizing the available resources are higher. The uniformly distributed configuration shows a much steeper slope. It is clear that, this configuration is not efficient enough for low traffic zones. It is observed that, at the peak of traffic, dynamic slicing attains a significantly effective utilization level of system resource. We can therefore, conclude that the dynamic resource allocation based on UD can achieve better resource utilization and therefore better efficiency from the perspectives of InPs with increasing traffic. For a more definite conviction, we still study the user satisfaction shown in **Fig. 12**.
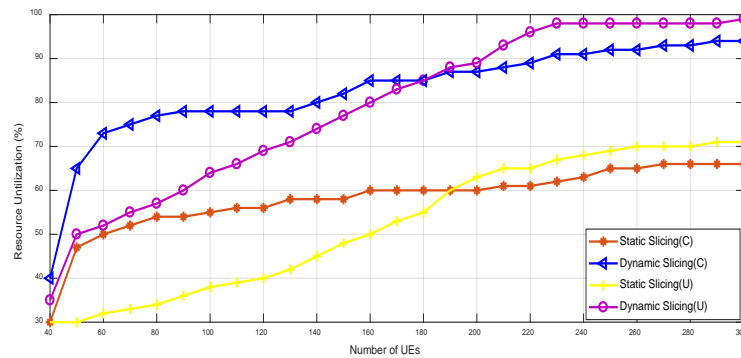

Fig. 10. Spectrum Efficiency of System

5.3 VO-UE QoS Perspective

In order to aggregate as much UEs per slice on the least feasible amount of active sc-BSs, flow handover could happen in each scanning period. This result is illustrated in **Fig. 11**. The total number of flow handover is similar and proportional to the traffic dynamics with the exception of the simple on-off scheme, where virtually no handovers were noticed. An increased duration or the number of handover increases the latency and has negative impacts

on user QoS. Within our proposed algorithm, there is not very significant difference in the three cases. However, MVC-UD shows prospects over other cases. It is therefore deemed a better approach in granting user QoS with less latency period.
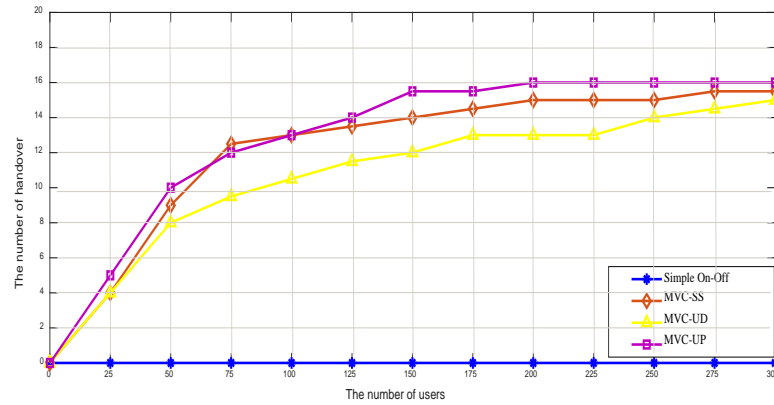

Fig. 11. Flow handover with increasing traffic

In order to check the throughput gain per slice against our QoS guarantee constraint in equation (18), we illustrate the average throughput gain per slice and the user satisfaction respectively in **Fig. 12a** and **12b**.
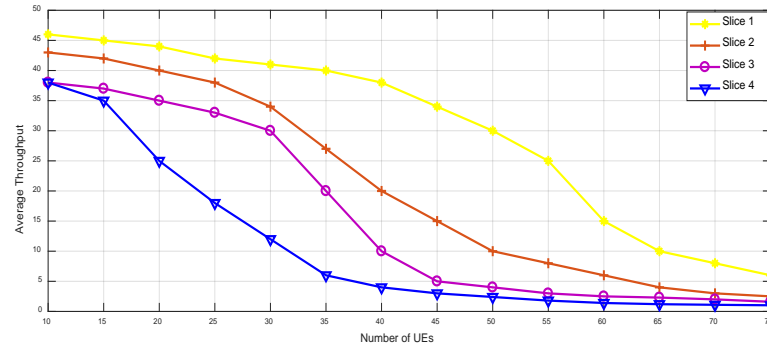

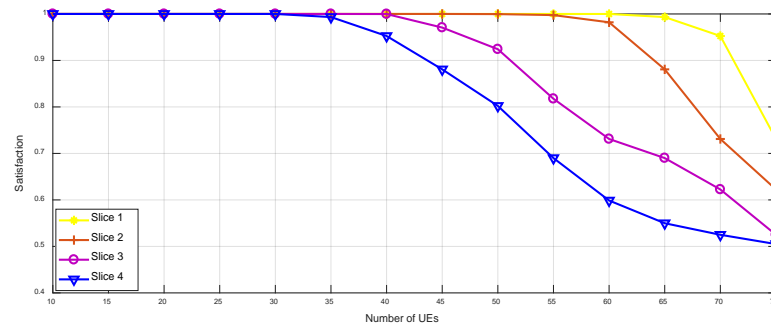**Fig. 12a.** Average Throughput Gain Per Slice

Fig. 12b. User Satisfaction per Slice

In our simulation, we study the benefits of resource virtualization on these two metrics: average throughput and user satisfaction. The traffic volume has significant impact on the

average user throughput gain. The traffic-QoS relationship is inversely proportional. With the number of UEs increasing from 10, 50 to 75 for each of the four slices, we can see the average throughput and satisfaction decrease because of user congestion. This is as a result of increasing ratio of UEs against limited radio resource. With resource virtualization, our algorithm meets the minimum QoS that satisfies each UE at the peak of traffic. Since the priority is based on UD, higher demanding slices are first served. Slice 1 (highly demanding UEs) therefore received higher satisfaction followed by 2, 3 and 4. From **Fig. 12b,** each UE is satisfied with at least its minimum QoS. We can therefore conclude that, the algorithm is sufficient in granting minimum satisfaction even to edge UEs. Our algorithm therefore succeeded in guaranteeing user QoS over minimal active sc-BS as well as minimizing the system energy consumption.

6. Conclusion

In this paper, we proposed a dynamic resource provisioning and customized physical resource allocation scheme for EE and QoS optimization in virtualized wireless networks. In the resource provisioning scheme, we conclude that prioritizing slices based on UD performs better than user proportion based slicing scheme. In the customized physical resource allocation, we proposed a heuristic algorithm based on minimum vertex coverage of bi-partite graph. The low-complexity algorithm was compared with the simple on-off scheme and the optimal branch and cut algorithm. Simulation results showed that, the proposed UD based scheme outperform other existing schemes in terms of energy consumption and user satisfaction on QoS. This is because serving users with the minimum available active sc-BSs minimized the energy consumption and maximized spectral efficiency and resource utilization.

Acknowledgment

References

[1] Nokia White paper, "Ultra Dense Network (UDN) White Paper," in *Nokia Solutions and Network*s, pp 4-26, 2016. Article (CrossRef Link)

[2] N. Bhushan et al. "Network densification: The dominant theme for wireless evolution into 5G," in *IEEE Commun. Mag*. vol. 52 no. 2 pp. 82-89 Feb. 2014. Article (CrossRef Link)

[3] I. Hwang B. Song S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," in *IEEE Commun. Mag*., vol. 51 no. 6 pp. 20-27 Jun. 2013. Article (CrossRef Link)

[4] D. Sabella, A. Vaillant, P. Kuure, U Rauschenbach, and F. Gius, "Mobile-Edge Computing Architecture: The role of MEC in the Internet of Things, " in *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 84-91, Oct. 2016. Article (CrossRef Link)

[5] M.I. Kamel, L. Bao Le, and A. Girard, "LTE Wireless Network Virtualization: Dynamic Slicing via Flexible Scheduling", *in 2014 IEEE 80th Vehicular Technology conference* (VTC), pp. 1-5, 2014. Article (CrossRef Link)

[6]   R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "NVS: A Substrate for Virtualizing Wireless Resources in Cellular Networks", *in IEEE/ACM Transactions on Networking* (TNET), vol. 20, no.5, 2012, pp.1333-1346. Article (CrossRef Link)

[7]   R. Mahindra, M. Khojastepour, H. Zhang and S. Rangarajan,"Radio Access Network Sharing in Cellular Networks" *in Proc. IEEE Int. Conf. Netw. Protocols,* pp.1-10, Oct. 2013. Article (CrossRef Link)

[8]   J. He and W. Song, "AppRAN: Application-Oriented Radio Access Network Sharing in Mobile Networks" *in Proc. IEEE Int. Conf. Commun.*, pp. 3788-3794, Jun 2015. Article (CrossRef Link)

[9]   A. Aijaz, "Hap-sliceR: A Radio Resource Slicing Framework for 5G Networks with Haptic Communications", *IEEE System Journal*, vol. PP, issue. 99, pp. 1-12, 2017. Article (CrossRef Link)

[10] M. Jiang, M. Condoluci and T. Mahmoodi, "Network slicing management & prioritization in 5G mobile systems," in *Proc. of 2016 IEEE 22nd European Wireless Conf.*, pp.1-6, 2016. Article (CrossRef Link)

[11] V. Jumba, "QoS-aware Resource Provisioning in Virtualized Wireless Networks" *Master Thesis, McGill University, Canada* (April 2016). Article (CrossRef Link)

[12] S. Parsaeefard, V. Jumba, M. Deraksshani and T. Le-Ngoc, "Joint Resource Provisioning and Admission Control in Wireless Virtualized Networks" in *Proc. IEEE Wireless Commun. and Netw. Conf.* (WCNC), pp. 2020-2025, March 2015. Article (CrossRef Link)

[13] V. Jumba, S. Parsaeefard, M. Deraksshani and T. Le-Ngoc, "Energy Efficiency Robust Resource Provisioning in Wireless Virtualized Networks" in *Proc. of 2015 IEEE Intl. Conf. on Ubiquitous Wireless Broadband, (ICUWB) Workshop on Next Generation ICT and 5G Networking (GreeNets)*, pp. 1-5, Oct. 2015. Article (CrossRef Link)

[14] Koudouridis, G.P., Hui Gao, Legg, P., "A centralised approach to power on-off optimization for heterogeneous networks," in *2012 IEEE Vehicular Technology Conference* (VTC Fall), vol. 1, no.5, pp. 3-6, Sept. 2012. Article (CrossRef Link)

[15] Sun, G., Addo, P. C., Wang, G., and Liu, G. 2016. "Energy efficient cell management by flow scheduling in Ultra-dense networks," in *KSII Transactions on Internet and Information Systems,* vol. 10, no. 9, pp 4108-4122 Sept. 2016. Article (CrossRef Link)

[16] J.G. Andrews, F. Baccelli, R.K. Ganti, "A tractable approach to coverage and rate in cellular networks," in *IEEE Trans. Commun.,* vol. 59, no. 9, pp. 3122-3134, Nov. 2011. Article (CrossRef Link)

[17] X. Zhang, J. Zhang, Y. Huang, W. Wang, "On the study of fundamental trade-offs between QoE and energy efficiency in wireless networks," *Transactions on Emerging Telecommunications Technologies*, vol. 24, no. 3, pp. 259-265, April. 2013. Article (CrossRef Link)

[18] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient context-aware user association for outdoor small cell heterogeneous networks," in *Proc. of 2014 IEEE Int. Conf. on Commun. (ICC)*, pp. 1614–1619, Jun. 2014. Article (CrossRef Link)

[19] Y. Xu, R. Hu, L. Wei, and G. Wu, "QoE-aware mobile association and resource allocation over wireless heterogeneous networks," in *Proc. of IEEE Global Commun. Conf. (GLOBECOM)*, pp. 4695–4701, Dec. 2014. Article (CrossRef Link)

[20] L. Su, C. Yang, Z. Xu, and A. Molisch, "Energy-efficient downlink transmission with base station closing in small cell networks," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 4784–4788, May 2013. Article (CrossRef Link)

[21] E. Chavarria Reyes, I. Akyildiz, and E. Fadel, "Energy consumption analysis and minimization in multi-layer heterogeneous wireless systems*," in IEEE Trans.on Mobile Computing.,* vol. 14, no. 12, pp. 2474–2487, Dec. 2015. Article (CrossRef Link)

[22] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy efficient user association in cognitive heterogeneous networks," in *IEEE Commun. Mag.,* vol. 52, no. 7, pp. 22–29, Jul. 2014. Article (CrossRef Link)

[23] Z. Xu, F. Liu, T. Wang, H. Xu "Demystifying the Energy Efficiency of Network Function Virtualization" *in IEEE/ACM 24th International Symposium on Quality of Service (IWQoS),* pp.

1-10, 2016. Article (CrossRef Link)

[24] Q. Shi, L. Zhao, Y. Zhang, G. Zheng, F. Yu, H. Chen "Energy Efficiency versus Delay Tradeoff in Wireless Networks Virtualization" *in IEEE Transactions on Vehicular Technology*, vol. 67, Issue 1, pp. 837-841, Jan. 2018 Article (CrossRef Link)

[25] G. Adam, M. Shikharesh "Energy Aware Resource Management for MapReduce Jobs with Service Level Agreements in Cloud Data Centers" in *2016 IEEE International Conference on Computer and Information Technology*, pp.568-577, Dec. 2016. Article (CrossRef Link)

[26] European Commission, "Cloud Computing Service Level Agreements-Exploitation of Research Results" in *Communications networks, Content and Technology*, *unit E2-Brusseles*, June 2013. Article (CrossRef Link)

[27] G. Asanka, P. Indika, M. Dulani, "Managing Service Level Agreements in Service Oriented Product Lines" *in 16th International Conference on Advances in ICT for Emerging Regions (ICTer)*, pp. 274 -280, 2016. Article (CrossRef Link)

[28] Q Ye, B Rong, Y chen, M AI-Shalash, "User association for load balancing in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706-2716, 2012. Article (CrossRef Link)

[29] C. Xu, T.Li, M. Sheng, et al. "Self-organized dynamic caching space sharing in virtualized wireless networks," in *IEEE Globecom Workshops (GC Wkshps)*, pp.1-6, Dec. 2016. Article (CrossRef Link)

[30] 3GPP TR 36.815, v.9.0.0, Further advancements for E-UTRA physical layer aspects, Mar. 2010. Article (CrossRef Link)

Guolin Sun received his B.S., M.S. and Ph.D. degrees all in Comm. and Info. System from the University of Electronic Sci.&Tech. of China (UESTC), Chengdu, China, in 2000, 2003 and 2005 respectively. After Ph.D. graduation in 2005, Dr. Guolin has got eight years industrial work experiences on wireless research and development for LTE, Wi-Fi, Internet of Things (ZIGBEE and RFID, etc.), Cognitive radio, Localization and navigation. Before he joins the School of Computer Science and Engineering, University of Electronic Sci.&Tech. of China, as an Associate Professor on Aug. 2012, he worked in Huawei Technologies Sweden. Dr. Guolin Sun has filed over 30 patents, and published over 30 scientific conference and journal papers, acts as TPC member of conferences. Currently, he serves as a vice-chair of the 5G oriented cognitive radio SIG of the IEEE (Technical Committee on Cognitive Networks (TCCN) of the IEEE Communication Society. His general research interest is software defined networks, network function virtualization, radio resource management.

Prince Clement Addo received his Bachelor in Information Technology from University Education, (College of Technology) Kumasi-Ghana, West Africa, in 2014 and MSc. Computer Science and Technology from the University of Electronic Science and Technology of China (UESTC) in 2017. He is currently a PhD student in the school of Management Science and Engineering (UESTC). From 2005 to 2010, he worked as an Instructor and research assistance in Ghana Education Service and Center for Competency Based Training and Research (CCBTR) University Education Winneba-Kumasi respectively. He is also a member of the Mobile Cloud-Net Research Team – UESTC. His interests include Mobile/Cloud Computing, Data mining, network resource virtualization and management.

Gordon Owusu Boateng received his Bachelor in Telecommunications Engineering from the Kwame Nkrumah University of Science and Technology, Kumasi-Ghana, West Africa, in 2014. He is currently studying MSc. Computer Science and Technology in University of Electronic Science and Technology of China (UESTC). From 2014 to 2016, he worked under sub-contracts for Ericsson (Ghana) and TIGO (Ghana). He is also a member of the Mobile Cloud-Net Research Team – UESTC. His interests include Mobile/Cloud Computing, 5G Wireless Networks, Data Mining and SDN.

Wei Jiang received his Ph.D degree from Beijing University of Posts and Telecommunications (BUPT) in 2008. Since Mar. 2008, he has been worked 4 years in Central Research Institute of Huawei Technologies, in the field of wireless communications and 3GPP standardization. In Sept. 2012, he joined the Institute of Digital Signal Processing, University of Duisburg-Essen, Germany, where he was a Postdoctoral researcher and worked for EU FP7 ABSOLUTE project and H2020 5G-PPP COHERENT project. Since Oct. 2015, he joined the Intelligent Networking Group, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany, as a senior researcher and works for H2020 5G-PPP SELFNET project. Meanwhile, he also works for the Department of Electrical and Information Technology (EIT), Technische University (TU) Kaiserslautern, Germany, as a senior lecturer. He served as a vice Chair of IEEE TCCN special interest group (SIG) "Cognitive Radio in 5G". He is the author of more than 30 papers in top international journals and conference proceedings, and has 27 patent applications in wireless communications, most of which have already been authorized in China, Europe, United States or Japan. He wrote a chapter "From OFDM to FBMC: Principles and Comparisons" for the book "Signal Processing for 5G: Algorithms and Implementations" (Wiley, 2016). His present research interests are in digital signal processing, multi-antenna technology, cooperative communications, 5G, and machine learning.