User Access Control and Bandwidth Allocation for Slice-based 5G-and-beyond Radio Access Networks

Yao Sun*†, Gang Feng* *Senior Member, IEEE*, Lei Zhang†, Mu Yan*, Shuang Qin* *Member, IEEE*, Muhammad Ali Imran† *Senior Member, IEEE*
* National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, China
† School of Engineering, University of Glasgow, Glasgow
Email: fenggang@uestc.edu.cn

***Abstract*—In this paper, we investigate the resource management for radio access network slicing from user access control and wireless bandwidth allocation perspectives. First, to guarantee users' QoS, we propose two admission control (AC) policies to select admissible users from the perspective of optimizing the QoS and the number of serving users respectively. Then, to optimize the bandwidth utilization for the selected admissible users, we investigate the slice association and bandwidth allocation (SABA) problem and propose network centric and UE centric SABA policies respectively. Numerical results show that in typical scenarios, our proposed AC and SABA policies can significantly outperform traditional policies in terms of wireless bandwidth utilization and number of admissible users.**

I. Introduction

It is envisioned in 5G-and-beyond systems, networks will be further abstracted into network slicing (NS), which enables design, deployment, customization, and optimization of isolated virtual sub-networks/slices on a common physical network infrastructure [1]–[3]. This NS-based new architecture can dramatically improve network capabilities in terms of capacity, delay, transmission rate, etc. by providing tailored service to meet users' specific quality of service (QoS) demands. Recently, some research work starts to focus on NS function virtualization and softwarizaion for core networks and radio access networks (RAN) [4]–[7], where optimization of resource configuration between multiple NSs as well as some NS deployment problems, such as NS structure, cooperation between control and data planes are investigated. In the meanwhile, very little attention is paid to user equipment (UE) access control and wireless resource allocation (ACRA) in radio access network (RAN) slicing. Indeed, ACRA is one of the most important procedures for RAN slicing and thus determines both users and overall system performance [8].

In RAN slicing, the ACRA problem is fundamentally different from that in conventional mobile networks. First, from the network architecture perspective, the NSs are logically virtualized and isolated over shared physical networks. Hence, both physical and virtual resource constraints need to be considered to form a complete function chain for the specific service. Second, from the user association perspective, UEs should be associated with an NS via a specific physical access point (AP) such as base station (BS). Hence, a joint optimization of NS and BS selection should be addressed. Third, from the service perspective, NS-based networks provide guaranteed QoS for all serving UEs instead of the traditional "best effort" model [9]. Due to the aforementioned challenges, applying traditional ACRA mechanisms to RAN slicing may lead to low resource utilization, poor QoS provisioning, frequent NS re-configurations and etc. Therefore, designing new ACRA mechanisms dedicated for RAN slicing to optimize network performance becomes an essential yet challenging issue.

In this paper, we study user ACRA problem in RAN slicing for 5G-and-beyond mobile communication systems. First, to guarantee the QoS of UEs we investigate user admission control (AC), and propose two AC policies to select admissible users from the perspective of optimizing QoS and the number of users respectively. We then optimize the bandwidth utilization for the selected admissible users by solving the NS association and bandwidth allocation (SABA) problem and propose network centric and UE centric SABA policies respectively. Numerical results show that in typical scenarios, our proposed AC and SABA policies can significantly outperform the traditional policies in terms of the number of serving UEs and the bandwidth consumption.

In the rest of the paper, we present system model and problem formulation in Section II and III respectively. In Section IV, admission control policies are proposed to select the admissible users, and then the SABA problem for the admitted users is solved in Section V. We present numerical results in Section VI and conclude this paper in Section VII.

II. System Model

Consider a multi-slice and multi-AP communication scenario, where the BSs associated with multiple slices are deployed in the area. Each BS supports several NSs with different provisioned QoS, and each NS may also cover several BSs. Multiple UEs are randomly distributed in this area with different QoS requirements. Let $\mathcal{B}$, $\mathcal{S}$ and $\mathcal{U}$ denote the set of

This work was supported by the National Science Foundation of China under Grant number 61631004 and 61471089, the U.K. Engineering and Physical Sciences Research Council under Grant EP/S02476X/1, and the Basic Research on Innovation Projects under Grant number IFN2018206.

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

BSs, NSs and UEs, respectively. For a specific BS, say BS k, we use $\mathcal{S}_k$ to denote the set of NSs supported by BS k.

We identify a specific NS, say NS j, by transmission rate, delay, and the resource allocation in core and access networks. Thus, besides the slice ID, four elements $(R_j, D_j, \Lambda_j, \vec{B}_j)$ are used to identify the j-th slice, where R_j and D_j denotes the minimum transmission rate and the maximum delay that NS j can provide to its serving UEs respectively, Λ_j denotes the bandwidth allocated to NS j in core network, and $\vec{B}_j$ is a vector denoting the wireless bandwidth allocation of NS j from all BSs. Let $b_j^{(k)}$ be the k-th element of vector $\vec{B}_j$ denoting the bandwidth of NS j allocated by BS k. $b_j^{(k)} = 0$ when BS k is not in the coverage of NS j.

For a specific UE, say UE n with q_n volume data to be transmitted, the QoS can be described by two metrics: transmission rate $\bar{r}_n$ and delay $\bar{d}_n$ [10]. NS j is admissible for UE n when $R_j \geq \bar{r}_n$ and $D_j \leq \bar{d}_n$. We now study the two QoS metrics respectively. Let $r_n^{j,k}$ be the transmission rate of UE n served by NS j via BS k. For simplicity, we use Shannon theory to define the transmission rate, i.e., $r_n^{j,k} = w_n^{j,k} log_2(1 + SINR_n^k)$, where $w_n^{j,k}$ is the wireless bandwidth that BS k allocates to the UE n which is served by NS j, and $SINR_n^k$ is the signal-to-noise-ratio between UE n and BS k. We use $d_n^{j,k} = q_n/r_n^{j,k}$ to denote the delay in RAN of UE n served by NS j via BS k. Thus, the end-to-end delay can be approximately calculated as $d_n^{j,k} + D_j$.

III. Problem Formulation

The optimization problem in this work can be described as: minimizing the bandwidth consumption subject to QoS, NS and BS resource constraints through joint access control and bandwidth allocation for UEs. Before formulating this problem, we define a binary variable $x_n^{j,k} \in 0, 1, \forall (n, j, k) \in \mathcal{U} \times \mathcal{S} \times \mathcal{B}$, where $x_n^{j,k} = 1$ indicates that UE n is served by NS j via BS k. We can formulate the access control and bandwidth allocation problem **P1** as:

$$\mathbf{P1}: \min \sum_{n\in\mathcal{U}} \sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{B}} x_n^{j,k} w_n^{j,k}, \tag{1}$$

$$\text{s.t.} \sum_{n\in\mathcal{U}} \sum_{k\in\mathcal{B}} x_n^{j,k} r_n^{j,k} \leq \Lambda_j, \ \forall j \in \mathcal{S} \tag{1-1}$$

$$\sum_{n\in\mathcal{U}} x_n^{j,k} w_n^{j,k} \leq b_j^{(k)}, \ \forall j \in \mathcal{S}, \forall k \in \mathcal{B} \tag{1-2}$$

$$\sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{B}} x_n^{j,k} r_n^{j,k} \geq \bar{r}_n, \ \forall n \in \mathcal{U} \tag{1-3}$$

$$\sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{B}} x_n^{j,k} R_j \geq \bar{r}_n, \ \forall n \in \mathcal{U} \tag{1-4}$$

$$\sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{B}} x_n^{j,k} (d_n^{j,k} + D_j) \leq \bar{d}_n, \ \forall n \in \mathcal{U} \tag{1-5}$$

$$\sum_{j\in\mathcal{S}} \sum_{k\in\mathcal{B}} x_n^{j,k} = 1, \ \forall n \in \mathcal{U} \tag{1-6}$$

$$x_n^{j,k} \in 0, 1, \ \forall (n, j, k) \in \mathcal{U} \times \mathcal{S} \times \mathcal{B} \tag{1-7}$$

where $x_n^{j,k}$ and $w_n^{j,k}$ are the optimization variables. Constraint (1-1) is referred to as the wired link resource constraint to guarantee that the total transmission rate offered by an NS does not exceed the link budget of that NS. Constraint (1-2) states the wireless bandwidth constraint. It ensures that the total bandwidth allocated to UEs by NS j via BS k does not exceed the wireless bandwidth budget of NS j deployed on BS k. Note that constraint (1-2) also ensures that UEs cannot access an NS via the BS that cannot provide such a service. Constraints (1-3)-(1-5) guarantee the QoS (rate and delay) of UEs can be satisfied by its' serving BS and NS. Constraints (1-6) and (1-7) ensure that each UE can only access one NS via one BS at a time.

Since **P1** requires to guarantee the QoS of all the UEs with limited resources, there may be no feasible solution in the case of dense UE distribution and/or high QoS requirement. Therefore, some UEs cannot be served in the resources limited RAN slicing. In the following, we first need to conduct admission control to select suitable serving UEs for the network.

IV. Admission Control

In this section, we design admission control (AC) schemes for the network to find a set of admissible UEs whose QoS can be satisfied simultaneously. We assume that network slice allocates minimal required wireless bandwidth to UEs to satisfy the UEs' minimum QoS requirements. Thus, the original constraints (1-3)-(1-5) about UEs QoS become equality constraints. Moreover, as we use this assumption, the allocated bandwidth $w_n^{j,k}$ is not the optimization variable for AC design. We will optimize $w_n^{j,k}$ in the next section by solving **P1** for those admissible UEs.

***Definition** 1: Subset $\mathcal{A}$ is an admissible UE set (AUS) if problem **P1** is feasible when $\mathcal{U}$ is replaced by $\mathcal{A}$.*

Definition 1 describes the feasibility of a UE subset. Hence, for a specific AUS, the slice-based network can simultaneously guarantee the QoS of all the UEs in this AUS. However, it is not a sufficient condition to achieve good overall network performance. For example, it is meaningless to choose an AUS which contains only one UE. Therefore, we need to design an approach to find a UE subset which is not only feasible for problem **P1** but also achieves good overall network performance. In the following of this section, we will develop two AC schemes under Assumption 1 to optimize the QoS and number of admissible UEs respectively.

A. Optimal QoS AC Scheme

In this subsection, we first consider AC scheme from UE QoS perspective. Our main idea is to reject the UEs which have QoS gap between the requirement and availability. First, we introduce two elastic variables $\check{r}_n$ and $\check{d}_n$ for UE n to describe the rate and delay degradation respectively. We restrict that $0 \leq \check{r}_n \leq \bar{r}_n$ and $\check{d}_n \geq 0$. Therefore, the rate and delay requirement of UE n can be referred to as $\bar{r}_n - \check{r}_n$ and $\bar{d}_n + \check{d}_n$ respectively. The UE n is admissible when $\check{r}_n = 0$ and $\check{d}_n = 0$. Our strategy is to reject the UEs with minimal sum of QoS degradation. Thus, we have the following definition.

__Definition__ 2: Subset $\mathcal{A}$ is a QoS-admissible UE set (QoS-AUS) if $\mathcal{A}$ is an AUS, and with the minimum achievable value of $\sum_{n\in\mathcal{U}\setminus\mathcal{A}}\left(\frac{\check{r}_n}{\bar{r}_n}+\frac{\check{d}_n}{\bar{d}_n}\right)$.

Here we use the normalized degradation of transmission rate (i.e., $\check{r}_n/\bar{r}_n$) and delay ($\check{d}_n/\bar{d}_n$). This definition describes both the feasibility and the QoS performance of a UE subset. In the following, we design an AC scheme named QoS-AC to find the QoS-AUS. By introducing elastic variables $\check{r}_n$ and $\check{d}_n$, we formulate problem **P2** as follows.

$$\mathbf{P2}:\ \min \sum_{n\in\mathcal{U}}\left(\frac{\check{r}_n}{\bar{r}_n}+\frac{\check{d}_n}{\bar{d}_n}\right), \tag{2}$$

$$\text{s.t.} \sum_{n\in\mathcal{U}}\sum_{k\in\mathcal{B}} x_n^{j,k} r_n^{j,k} \le \Lambda_j,\ \forall j\in\mathcal{S} \tag{2-1}$$

$$\sum_{n\in\mathcal{U}} x_n^{j,k} w_n^{j,k} \le b_j^{(k)},\ \forall j\in\mathcal{S},\forall k\in\mathcal{B} \tag{2-2}$$

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}} x_n^{j,k} r_n^{j,k} = \bar{r}_n-\check{r}_n,\ \forall n\in\mathcal{U} \tag{2-3}$$

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}} x_n^{j,k} R_j = \bar{r}_n-\check{r}_n,\ \forall n\in\mathcal{U} \tag{2-4}$$

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}} x_n^{j,k}(d_n^{j,k}+D_j) = \bar{d}_n+-\check{d}_n,\ \forall n\in\mathcal{U} \tag{2-5}$$

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}} x_n^{j,k} = 1,\ \forall n\in\mathcal{U} \tag{2-6}$$

$$x_n^{j,k}\in 0,1,\ \forall(n,j,k)\in\mathcal{U}\times\mathcal{S}\times\mathcal{B} \tag{2-7}$$

In **P2**, the optimization objective is minimizing the normalized QoS degradation of all UEs. Compared with the constraints in **P1**, the only difference is using equalities in constraints (2-3)-(2-5) to replace the inequalities in (1-3)-(1-5) by introducing elastic variables. In **P2**, the optimization variables are binary indicators $x_n^{j,k}$ as well as the continuous elastic variables $\check{r}_n$ and $\check{d}_n$. Hence, **P2** is a mixed integer liner programming (MILP). As we introduce the elastic variables into the above MILP, the QoS of UEs can vary with the elastic variables, and thus problem **P2** is always feasible. Using Lagrange decomposition theory [11], we can obtain the optimal solution of **P2** denoted as $\{\check{\boldsymbol{r}}^*,\check{\boldsymbol{d}}^*,\boldsymbol{x}^*\}$.

Then we design QoS-AC policy based on the solution of **P2** to find the QoS-AUS in Definition 2. Let $\check{r}_n^*$ and $\check{d}_n^*$ be the optimal solution of UE n. If $\check{r}_n^*=0$ and $\check{d}_n^*=0$, it means that there is no QoS degradation of the UE. In other words, this UE can be admissible for the network. Hence, based on this observation, we design QoS-AC policy, where the UEs with $\check{r}_n^*=0$ and $\check{d}_n^*=0$ can be accepted by the network, and others are rejected due to limited resources. Hence, the admissible set of UE can be expressed as $\mathcal{A}_{Q-A}=\{n:\check{r}_n^*=0,\check{d}_n^*=0,n\in\mathcal{U}\}$.

We can prove that the QoS-AC policy can minimize the QoS degradation of the rejected UEs, i.e., it guarantees the QoS performance and the feasibility. Moreover, this AC policy also guides network operators to re-allocate bandwidth in the NS reconfiguration phase thus to satisfy QoS of all the UEs in set $\mathcal{U}$ with the minimum bandwidth consumption. However, network slice reconfiguration is beyond the scope of this work. QoS-AC policy is focused on QoS degradation, and the performance of the number of admissible UEs cannot be guaranteed. In the next subsection, we will design another AC scheme named Num-AC to maximize the number of admissible UEs.

B. Num-AC Scheme

In the proposed QoS-AC, we find that some UEs with only rate or delay degradation (i.e. $\check{r}_n^*>0,\check{d}_n^*=0$ or $\check{d}_n^*>0,\check{r}_n^*=0$) should be rejected. This means that some unviolated constraints are deleted in **P1**, which implies that the network may have some spare resources to accept more UEs. Hence, from the number of admissible UEs perspective, the performance of QoS-AC policy may not be good. Moreover, the number of admissible UEs is also one of the key performance measures of UE admission control policy. Therefore, we propose Num-AC policy based on the solution of **P2** to optimize the number of admissible UEs.

By analyzing the optimal solution of **P2**, we find that the smaller $\check{r}_n^*$ or $\check{d}_n^*$ is, the more likely the rate or delay of UE n can be satisfied. Based on this observation, we develop Num-AC policy to find the admissible UE subset with respect to the number of admissible UEs, and we denote by $\mathcal{A}_{N-A}$ the corresponding subset. The basic idea of this policy is trying to add the UEs with small $\check{r}_n^*$ and $\check{d}_n^*$ into set $\mathcal{A}_{N-A}$.

First of all, the UEs with $\check{r}_n^*=0$ and $\check{d}_n^*=0$ are definitely admissible for the network. Hence, $\mathcal{A}_{Q-A}\subseteq\mathcal{A}_{N-A}$. We then try to find more admissible UEs from set $\mathcal{U}\setminus\mathcal{A}_{Q-A}$, and add these UEs into $\mathcal{A}_{N-A}$. The details of Num-AC policy are summarized as Algorithm 1.

Algorithm 1 : Algorithm of Num-AC policy.

Input: problem **P2** formulated by all UEs.
Output: set of admissible UEs $\mathcal{A}_{N-A}$.

Initialization Stage:
1: $\mathcal{A}_{N-A}=\emptyset$, $\mathcal{A}_{temp}=\emptyset$
2: obtain the optimal solution $\check{\boldsymbol{r}}^*$, $\check{\boldsymbol{d}}^*$ and $\boldsymbol{x}^*$ by solving **P2**
3: add all UEs with $\check{r}_n^*=0$ and $\check{d}_n^*=0$ into $\mathcal{A}_{N-A}$
Search Stage:
4: find the UE i respect to $\min_{i\in\mathcal{U}\setminus\mathcal{A}_{N-A}}\left(\frac{\check{r}_n^*}{\bar{r}_n}+\frac{\check{d}_n^*}{\bar{d}_n}\right)$
5: $\mathcal{A}_{temp}=\{\mathcal{A}_{N-A}\cup \text{UE } i\}$
6: obtain the optimal solution $\check{\boldsymbol{r}}^*$, $\check{\boldsymbol{d}}^*$ and $\boldsymbol{x}^*$ by solving **P2** with respect to $\mathcal{A}_{temp}$
7: **if** $\sum_{n\in\mathcal{A}_{temp}}\left(\frac{\check{r}_n^*}{\bar{r}_n}+\frac{\check{d}_n^*}{\bar{d}_n}\right)=0$ **then**
8: $\quad\mathcal{A}_{N-A}=\mathcal{A}_{temp}$
9: $\quad$Go back to line 4
10: **else**
11: $\quad$break
12: **end if**
13: **output** $\mathcal{A}_{N-A}$

In the initialization stage, we add the definite admissible UEs. Then in the search stage, we check the feasibility of other UEs one by one. The smaller the value of $\left(\frac{\check{r}_n^*}{\bar{r}_n}+\frac{\check{d}_n^*}{d_n}\right)$ is, the more likely the UE is admissible. Hence, we check UEs with the smallest value of $\left(\frac{\check{r}_n^*}{\bar{r}_n}+\frac{\check{d}_n^*}{d_n}\right)$ first. To reduce the computational complexity, once a UE is unfeasible for the network, we stop the check, and then obtain the set $\mathcal{A}_{N-A}$. Therefore, this policy needs to solve **P2** at most $|\mathcal{U}|$ times in the worst case.

V. Network Slice Association and Bandwidth Allocation Policy

We now focus on the network slice association and bandwidth allocation (SABA) problem formulated in **P1** for the admissible UE subsets. In this section, we develop two NS association and bandwidth allocation policies, i.e., Net-SABA and UE-SABA from network and UE perspective respectively. Net-SABA policy determines NS association and bandwidth allocation for UEs to minimize the overall bandwidth consumption. UE-SABA policy with low computational complexity tries to reduce the individual UE bandwidth consumption from a feasible solution.

A. Network Centric Policy Net-SABA

We first develop the network centric SABA policy named Net-SABA. For convenience, we use $y_n^{j,k} = w_n^{j,k}/b_j^{(k)}$ to replace $w_n^{j,k}$, and thus $y_n^{j,k} \in [0,1]$. In addition, we define a mapping function $\phi(n,j,k)$ to determine a unique integer between 1 and $|\boldsymbol{x}|$ when n,j,k is given, where $|\boldsymbol{x}|$ is the number of elements of all $x_n^{j,k}$. Then, we transform variables $x_n^{j,k}$ and $y_n^{j,k}$ into $\tilde{x}_{\phi(n,j,k)}$ and $\tilde{y}_{\phi(n,j,k)}$, i.e., $x_n^{j,k} = \tilde{x}_{\phi(n,j,k)}$ and $y_n^{j,k} = \tilde{y}_{\phi(n,j,k)}$. Let $\phi^{-1}_{(n)}$, $\phi^{-1}_{(j)}$ and $\phi^{-1}_{(k)}$ be the inverse function of n, j and k, respectively. Let $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ be the set of $\tilde{x}_{\phi(n,j,k)}$ and $\tilde{y}_{\phi(n,j,k)}$ respectively. In the following, we solve **P1** with respect to $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$.

Note that it is hard to directly find the optimal solution for **P1** due to the binarity of $\widetilde{\boldsymbol{x}}$. Here, we first relax the feasible region of $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$ to a convex set, and then solve **P1** subject to the relaxed convex feasible region. Let Z be the feasible region of **P1**, and thus $Z = \{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}) : s.t.\ \text{constraints}\ (1-1)-(1-7)\}$. The *convex hull* of a set Z, denoted by $conv(Z)$ is the smallest convex set that contains Z [12]. Using the similar idea of [13], we give $conv(Z)$ in the following.

Define the polynomial factors of degree d as $F_d(J_1,J_2) = [\Pi_{i\in J_1}\tilde{x}_i][\Pi_{j\in J_2}(1-\tilde{x}_j)]$, where $J_1, J_2 \subseteq \{1,2,,\ldots,|\boldsymbol{x}|\} \equiv \mathcal{J}$, $J_1 \cap J_2 = \emptyset$ and $|J_1 \cap J_2| = d$. To linearize the cross-product terms of $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{y}}$, we define $u_J = \Pi_{i\in J}\tilde{x}_i$ and $v_{J,m} = \tilde{y}_m\Pi_{i\in J}\tilde{x}_i$ for $m = 1,\cdots,|\boldsymbol{x}|$, where $u_\emptyset = 1$ and $v_{\emptyset,m} = \tilde{y}_m$, for $m = 1,\cdots,|\boldsymbol{x}|$. We denote by $f_d(J_1,J_2)$ and $f_d^m(J_1,J_2)$ the linearized forms of polynomial expressions $F_d(J_1,J_2)$ and $\tilde{y}_m F_d(J_1,J_2)$ respectively. For convenience, let $\tilde{b}_n^{j,k} \equiv b_j^{(k)} log_2(1+SINR_n^k)$, and $\phi \equiv \phi(n,j,k)$.

For constraints (1-1)-(1-3), and (1-5), we then use constraints (3)-(6) to relax them, and for constraints (1-4), (1-6), $\tilde{x}_\phi \in \{0,1\}$ and $\tilde{y}_\phi \in [0,1]$, we give (7)-(10) to relax them,

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}}\tilde{x}_\phi R_j - \bar{r}_n \geq 0,\ \forall n\in\mathcal{U} \tag{7}$$

$$\sum_{j\in\mathcal{S}}\sum_{k\in\mathcal{B}}\tilde{x}_\phi - 1 \geq 0,\ \forall n\in\mathcal{U} \tag{8}$$

$$f_{D_1}(J_1,J_2) \geq 0, for\ (J_1,J_2)\ of\ order\ D_1 = min\{d,|\boldsymbol{x}|\} \tag{9}$$

$$\begin{aligned} &f_{D_2}(J_1,J_2) \geq f_{D_2}^m(J_1,J_2) \geq 0, for\ m = 1,\cdots,|\boldsymbol{x}|,\\ &and\ (J_1,J_2)\ of\ order\ D_2 = min\{d+1,|\boldsymbol{x}|\} \end{aligned} \tag{10}$$

By using these relaxed constraints, we obtain a convex relaxation of original feasible region Z. Let $Z_d = \{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}},\boldsymbol{u},\boldsymbol{v}) : s.t.\ \text{constraints}\ (3)-(10)\}$. Then the d-degree convex relaxation of Z can be expressed as $Z_{Pd} = \{(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}}) : s.t.(\widetilde{\boldsymbol{x}},\widetilde{\boldsymbol{y}},\boldsymbol{u},\boldsymbol{v}) \in Z_d\}$. Actually, for all degrees $0 \leq d \leq |\boldsymbol{x}|$, Z_{Pd} is a convex relation of the feasible region Z. The larger the degree d is, the tighter the relaxation Z_{Pd} is, and the higher computational complexity is needed [13].

***Theorem** 1: $Z_{P|\boldsymbol{x}|}$ is the convex hull of **P1**, i.e., $Z_{P|\boldsymbol{x}|} = conv(Z)$*

Proof: By using Theorem 3.5, Extension 1 and Extension 2 of [13], we can easily obtain Theorem 1. ∎

After the relaxation, the feasible region of **P1** becomes a convex set with linear constraints. Hence, **P1** becomes a linear programming which can be solved easily. Based on the solution of the relaxed **P1**, we design Net-SABA policy. In Net-SABA policy, we associate UE with the NS and the BS according to $\widetilde{\boldsymbol{x}}^*$, and allocate bandwidth according to $\widetilde{\boldsymbol{y}}^*$, where $\{\widetilde{\boldsymbol{x}}^*,\widetilde{\boldsymbol{y}}^*\}$ is the optimal solution to the relaxed problem. Net-SABA policy requires global network information (the association and bandwidth allocation of all UEs) and high computational complexity. In the following, we will design a UE centric SABA policy to reduce individual UE bandwidth consumption with low computational complexity.

B. UE Centric Policy UE-SABA

In this subsection, we develop an efficient UE-SABA policy to reduce individual UE bandwidth consumption. UE-SABA policy has two steps, obtaining initial solution and searching better solution. In the first step, let $\boldsymbol{x}^{(0)}$ and $\boldsymbol{w}^{(0)}$ be the optimal solution obtained by Num-AC scheme. Hence, $\boldsymbol{x}^{(0)}$ and $\boldsymbol{w}^{(0)}$ is also the feasible solution to **P1**. We use $\boldsymbol{x}^{(0)}$ and $\boldsymbol{w}^{(0)}$ as the initial solution.

Then in the second step, we try to find a better solution for each UE with the fixed associations of other UEs. In details, let $\boldsymbol{x}^{(s)}$ and $\boldsymbol{w}^{(s)}$ respectively be the NS association and the corresponding bandwidth allocation after s searching steps. In the $(s+1)$-th searching step, for a specific UE $n \in \mathcal{A}_{N-A}$, we first fix the associations and bandwidth allocations of other UEs, i.e., $\boldsymbol{x}^{(s+1)} = \boldsymbol{x}^{(s)}$ and $\boldsymbol{w}^{(s+1)} = \boldsymbol{w}^{(s)}$ except for the n-th element. Then we optimize the n-th element $x_n^{j,k}(s+1)$ and $w_n^{j,k}(s+1)$. We find the set $\mathcal{H}_n = \{(x_n^{j,k}, w_n^{j,k}) : s.t.\ \text{constraints}\ (1-3)-(1-5), \text{and}\ w_n^{j,k} \leq w_n^{j,k}(s) - \epsilon\}$, where $w_n^{j,k}(s)$ is the bandwidth allocation of UE n at the s-th step, and ϵ is an any positive parameter. If $\mathcal{H}_n = \emptyset$, we

$$\Lambda_j f_d(J_1,J_2) - \sum_{\phi\in\mathcal{J}-(J_1\cup J_2)} \tilde{b}_n^{j,k} f_{d+1}^{\phi^{-1}_{(n,k)}}(J_1+\phi,J_2) - sum_{\phi\in J_1}\tilde{b}_n^{j,k} f_d^{\phi^{-1}_{(n,k)}}(J_1,J_2) \geq 0, \forall j\in\mathcal{S}, and\ (J_1,J_2)\ of\ order\ d, \quad (3)$$

$$f_d(J_1,J_2) - \sum_{\phi\in\mathcal{J}-(J_1\cup J_2)} f_{d+1}^{\phi^{-1}_{(n)}}(J_1+\phi,J_2) - \sum_{\phi\in J_1} f_d^{\phi^{-1}_{(n)}}(J_1,J_2) \geq 0, \forall j\in\mathcal{S}, k\in\mathcal{B}, and\ (J_1,J_2)\ of\ order\ d, \quad (4)$$

$$\sum_{\phi\in J_1} \tilde{b}_n^{j,k} f_d^{\phi^{-1}_{(n,k)}}(J_1,J_2) - \sum_{\phi\in\mathcal{J}-(J_1\cup J_2)} \tilde{b}_n^{j,k} f_{d+1}^{\phi^{-1}_{(n,k)}}(J_1+\phi,J_2) - \bar{r}_n f_d(J_1,J_2) \geq 0, \forall n\in\mathcal{U}, and\ (J_1,J_2)\ of\ order\ d, \quad (5)$$

$$\sum_{\phi\in J_1}\left[q_{\phi^{-1}_{(n)}} f_d(J_1,J_2) + D_{\phi^{-1}_{(j)}}\tilde{b}_n^{j,k} f_d^{\phi^{-1}_{(n)}}(J_1,J_2)\right] - \sum_{\phi\in\mathcal{J}-(J_1\cup J_2)}\left[q_{\phi^{-1}_{(n)}} f_{d+1}(J_1+\phi,J_2) + D_{\phi^{-1}_{(j)}}\tilde{b}_n^{j,k} f_{d+1}^{\phi_{(n)}{}^{-1}}(J_1+\phi,J_2)\right] +$$
$$\sum_{\phi}\left[\bar{d}_{\phi^{-1}_{(n)}}\tilde{b}_n^{j,k} f_d^{\phi^{-1}_{(n)}}(J_1,J_2)\right] \geq 0, \forall n\in\mathcal{U}, and\ (J_1,J_2)\ of\ order\ d, \quad (6)$$

---

obtain $x_n^{j,k}(s+1) = x_n^{j,k}(s)$ and $w_n^{j,k}(s+1) = w_n^{j,k}(s)$. If $\mathcal{H}_n \neq \emptyset$, we find the pair $\left(x_n^{j,k}(s+1), w_n^{j,k}(s+1)\right)$ in $\mathcal{H}_n$ that satisfies 1) **P1** is feasible respect to $\boldsymbol{x}^{(s+1)}$ and $\boldsymbol{w}^{(s+1)}$, and 2) $w_n^{j,k}(s+1)$ is the smallest one among all the pairs which satisfy condition 1). If all the pairs in $\mathcal{H}_n$ are infeasible for **P1**, we have $x_n^{j,k}(s+1) = x_n^{j,k}(s)$ and $w_n^{j,k}(s+1) = w_n^{j,k}(s)$. In this way, we obtain $x_n^{j,k}(s+1)$ and $w_n^{j,k}(s+1)$, and thus $\boldsymbol{x}^{(s+1)}$ and $\boldsymbol{w}^{(s+1)}$. Therefore, the $(s+1)$-th searching step is finished. The searching termination criteria is set as the association and bandwidth allocation of all the UEs are unchanged.

VI. Simulation Results

In this section, we evaluate the performance of the proposed two AC policies (QoS-AC and Num-AC) as well as two SABA policies (Net-SABA and UE-SABA). We compare them with two traditional mechanisms: (1) NS prior association (NSA) and (2) BS prior association (BSA). In detail, for a specific UE, NSA mechanism first finds the NS that satisfies the QoS requirement of the UE, and then finds the BS covered by this NS with sufficient bandwidth. BSA mechanism first finds the BS with the maximum SINR for the UE, and then finds the NS deployed in this BS with satisfied QoS guarantee. In both NSA and BSA mechanisms, if such a pair of NS and BS is found, the UE is admissible and associated with the NS and BS. The bandwidth allocation policy for NSA and BSA is to allocate the minimal required bandwidth to UEs to satisfy the QoS requirement. Hence, NSA and BSA mechanisms contain both AC and SABA policies.

We consider a network which consists of a macro BS (MBS) located at the central of a circular area with a radius of 500m and varying number of pico BSs (PBS), femto BSs (FBS), NSs and UEs. The transmit power of MBS, PBS and FBS is set to 46dBm, 30dBm and 20dBm, respectively. We use $L(d) = 34+40log(d)$ and $L(d) = 37+30log(d)$ to model the pass loss for the MBS/PBSs and FBSs respectively [14]. All the BSs share 20MHz bandwidth. Each NS randomly covers 4 BSs, and provides different rate and delay performance. UEs are randomly distributed in this area with different rate and delay requirements. In the following, we examine the performance of proposed AC and SABA policies respectively.

In the first experiment, we compare the number of admissible UEs of the four AC policies QoS-AC, Num-AC, NSA and BSA. In this experiment, we fix the number of NSs and BSs to 20 and 21 (including one MBS) respectively. Fig. 1 shows the number of admissible UEs for the four AC policies with different UE distributions. From this figure, we can see that the number of admissible UEs of QoS-AC and Num-AC are always higher than that of the other two traditional policies which do not consider the characteristics of NS-based RAN. Specifically, when the number of UEs is 200, the admissible number of UEs for Num-AC, QoS-AC, BSA and NSA is 173, 142, 118 and 92, respectively. These results show that the proposed Num-AC policy can serve 47% and 88% more UEs when compared with NSA and BSA.

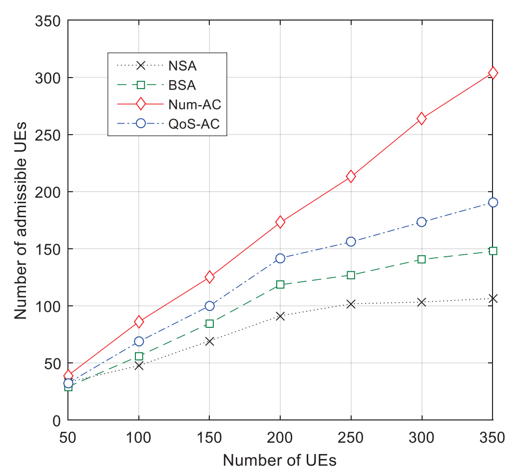

Fig. 1. The number of admissible UEs for the four AC policies.

In the second experiment, we compare the bandwidth consumption of the four SABA policies (Net-SABA, UE-SABA, NSA and BSA) with the same parameters as the first experiment. Fig. 2 shows the bandwidth consumption for the four policies as a function of number of UEs. From Fig. 2,

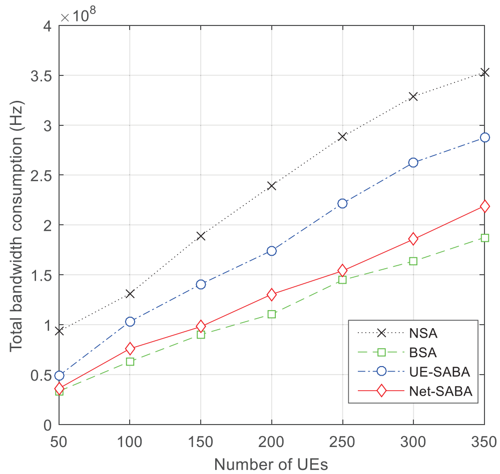

Fig. 2. Relationship between bandwidth consumption the number of UEs.

we can see that the bandwidth consumption of the traditional policy BSA is always the smallest. This is because that UEs always access the BS with the maximum SINR value in BSA policy. However, from the first experiment, we can see that the number of admissible UEs is much smaller than that of the two proposed policies. Moreover, we find that the difference of bandwidth consumption between Net-SABA and BSA is relatively small (for example, 7% for 150 UEs), implying that much more UEs can be served with a small compromise on bandwidth consumption.

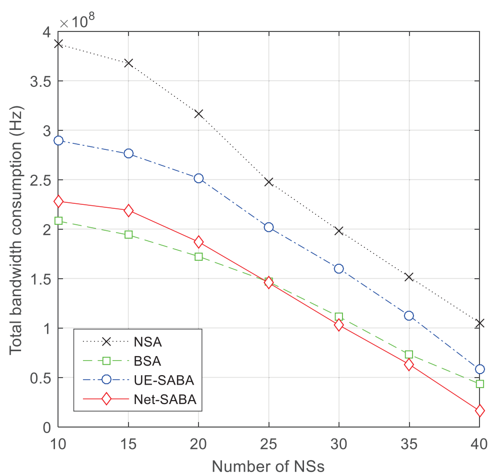

Fig. 3. Relationship between bandwidth consumption the number of NSs.

Next, we examine the bandwidth consumption of the four SABA policies for varying number of NSs while using fixed number of UEs 200. Fig. 3 shows the bandwidth consumption for the four policies with different number of NSs. We can see that the bandwidth consumption of all four policies monotonically increases with the number of NSs. The more NSs deployed, the better association choice for UEs, and thus the less bandwidth consumption. When the number of NSs equals 40, all the four policies can achieve low bandwidth consumption due to the sufficient available NSs. Similar to that in the second experiment, the average UE bandwidth consumption of Net-SABA is the lowest, although the total bandwidth consumption of Net-SABA is slightly higher than that of BSA.

VII. Conclusions

In this paper, we have investigated the user access control and bandwidth allocation in radio access network slicing for 5G-and-beyond systems. From the QoS and the number of users perspectives respectively, we have proposed two user admission control policies to select the admissible users. Then, we have studied the NS association and bandwidth allocation problem for those admissible users. We have proposed two policies Net-SABA and UE-SABA from network and UE perspective. Numerical results show that in typical scenarios, our proposed AC and SABA policies can significantly outperform the traditional policies in terms of the number of admissible UEs and bandwidth consumption.

References

[1] Ericsson, "5G systems," *Ericsson White Paper*, no. January, 2015.

[2] L. Zhang, A. Ijaz, J. Mao, P. Xiao, and R. Tafazolli, "Multi-service Signal Multiplexing and Isolation for Physical-Layer Network Slicing (PNS)," in *VTC2017-Fall Proceedings*, 2017.

[3] L. Zhang, A. Ijaz, P. Xiao, A. Quddus, and R. Tafazolli, "Subband filtered multi-carrier systems for multi-service wireless communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1893–1907, 2017.

[4] O. Sallent, J. Pérez-Romero, R. Ferrús, and R. Agustí, "On Radio Access Network Slicing From A Radio Rersource Management Perpective," *IEEE Wireless Communications*, vol. 24, no. 5, pp. 166–174, 2017.

[5] A. Nakao, P. Du, K. Yaoshiaki, G. Fabrizio, T. Anteneh Atumo, Gebremariam, A Tarik, and B. Miloud, "End-to-end Network Slicing for 5G Mobile Networks," *Journal of Information Processing*, vol. 25, pp. 153–163, 2017.

[6] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "CellSlice : Cellular Wireless Resource Slicing for Active RAN Sharing," in *Communication Systems and Networks (COMSNETS)*, 2013, pp. 1–10.

[7] H. M. Soliman and A. Leon-Garcia, "QoS-aware frequency-space network slicing and admission control for virtual wireless networks," in *2016 IEEE Global Communications Conference, GLOBECOM 2016 - Proceedings*, 2016, pp. 1–6.

[8] X. Chen, Z. Han, H. Zhang, G. Xue, Y. Xiao, and M. Bennis, "Wireless Resource Scheduling in Virtualized Radio Access Networks Using Stochastic Learning," *IEEE Transactions on Mobile Computing*, vol. 17, no. 4, pp. 961–974, 2018. [Online]. Available: http://ieeexplore.ieee.org/document/8014506/

[9] X. An, C. Zhou, R. Trivisonno, R. Guerzoni, A. Kaloxylos, D. Soldani, and A. Hecker, "On end to end network slicing for 5G communication systems," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 4, 2016.

[10] C. Liang and F. R. Yu, "Wireless Network Virtualization: A Survey, Some Research Issues and Challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6887287

[11] Y. Sun, G. Feng, S. Qin, Y.-C. Liang, and T.-S. P. Yum, "The SMART Handoff Policy for Millimeter Wave Heterogeneous Cellular Networks." *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1456–1468, 2018.

[12] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.

[13] H. D. Sherali and W. P. Adams, "A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems," *Discrete Applied Mathematics*, vol. 52, no. 1, pp. 83–106, 1994.

[14] Q. Ye, B. Rong, Y. Chen, M. Al-shalash, C. Caramanis, and J. G. Andrews, "User Association for Load Balancing in Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.