

Received December 13, 2020, accepted December 21, 2020, date of publication December 29, 2020, date of current version January 19, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048062

# Slice Merging/Splitting Operations and Tenant Profit Optimization Across 5G Base Stations

XIAO FU<sup>1</sup>, (Graduate Student Member, IEEE), QINGGUO SHEN<sup>1,2</sup>, (Member, IEEE), WENJIN WANG<sup>1</sup>, (Member, IEEE), HONGWEI HOU<sup>1</sup>, AND XIQI GAO<sup>1</sup>, (Fellow, IEEE)

<sup>1</sup>National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

<sup>2</sup>Purple Mountain Laboratories, Nanjing 211111, China

Corresponding authors: Xiqi Gao (xqgao@seu.edu.cn) and Qingguo Shen (shenqg2@163.com)

This work was supported by the National Key R&D Program of China under Grant 2018YFB1801103, the Jiangsu Province Basic Research Project under Grant SBK2019050020, the National Natural Science Foundation of China under Grants 61901110, 61761136016, and 61801114, and the Huawei Cooperation Project.

**ABSTRACT** Network slicing is a promising technology that can tailor the network to support diverse types of services over a common physical network infrastructure, and tenants can lease the customized slices from network providers to provide customized services for their users. This paper focuses on tenant profit optimization in 5G Radio Access Network (RAN) slicing and analyzes slice resource reservation and reconfiguration operation. A slice cost model is proposed, in which a base cost is included to compensate reconfiguration management overhead for the network providers. Increasing the accuracy of tenant resource requirement prediction can decrease tenant cost while guaranteeing QoS of its users. Further, we analyze the price strategy based on user traffic, and propose several alternative price scheme. In slicing across multiple Base Stations (BSs) scenario, we study two slice management operations, called slice mergence/split, similar but not identical to the breathing behavior of conventional cellular BS management. The tenant profit optimization problem across adjacent BSs is formulated, which can be used in slice mergence/split decision making. Taking VR (Virtual Reality) application as an example, simulation in allusion to the typical user arrival process is performed. The simulation indicates how these two operations can increase tenant profits.

**INDEX TERMS** RAN slicing, tenant profit, cost model, resource reservation, slice mergence/split.

## I. INTRODUCTION

Traditional “one size fits all” network architecture is no more efficient meeting variety of services and applications requirement. The 5G has been expected to support diverse applications and mobile devices simultaneously. Toward this end, network slicing virtualizes the 5G mobile network in a flexible and programmable way to meet customization requirements [1]. Network slicing can provide customized virtual subnetworks for different customized services in 5G system. These subnetworks, called slices, share the common physical infrastructure and maintains logical isolation with each other.

While the virtualization of 5G Core Networks (CN) has been widely studied, the research of RAN slicing is at its infancy [2]. Meanwhile, the 5G RAN, faced with unprecedentedly heterogenetic requirements, (e.g., high reliability, low latency, high data rates, and low energy consumption) becomes a promising issue [3]. In RAN slicing, user

traffic variations and network state dynamics (such as dynamic channel conditions, user mobility) make it very challenging to ensure diverse quality of service (QoS) of individual users. On the other hand, reducing the energy consumption is also an important issue in real RAN implementation. Thus, a good trade-off should be made between user QoS guarantee and system power efficiency.

In 5G wireless networks, the application of massive MIMO in the BS provides a way to utilize the degree of spatial freedom, which can vastly increase spectral and energy efficiencies. Meanwhile, the employment of orthogonal non-overlapping beams provides technical support to reuse the same time-frequency resource blocks and maintain the isolation between slices in spacial domain. Moreover, it offers a physical layer technology to slice a physical BS to several virtual BSs with different coverage range areas and enhances the feasibility of 5G service-oriented networks. However, channel state information (CSI) estimation in a massive MIMO environment poses a serious challenge and significantly complicates the resource allocation problem.

The associate editor coordinating the review of this manuscript and approving it for publication was Cunhua Pan <sup>1</sup>.

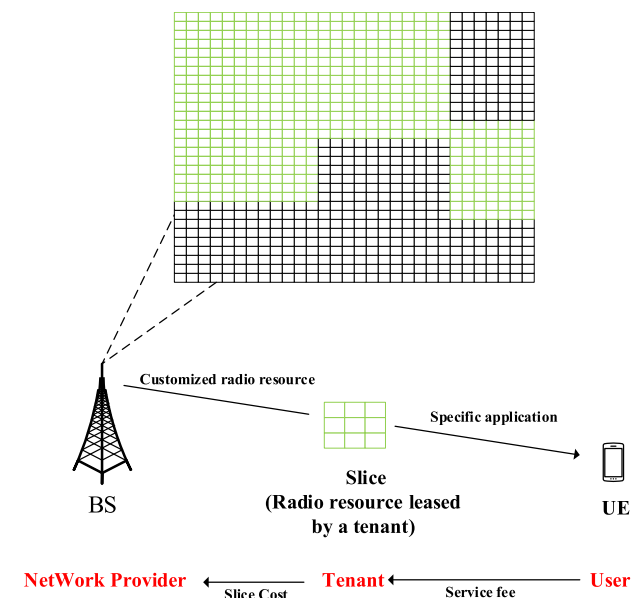


FIGURE 1. An illustration of relationships among network provider, a tenant and users.

To fully unleash the power of network slicing in dynamic environments, reconfiguration of slices must be enabled [4]. However, frequent reconfiguration will cause additional system costs (such as signaling overhead and computing resource consumption) and diminish slice isolation, a prerequisite for network slicing. Thus, frequent reconfiguration will be forbidden to protect slice isolation. In 5G, the change of BS configuration will be notified to users via system information (MIB, SIB). Moreover, system information are periodically broadcast about per 80/100ms by BS, which may cause a delay of tens of milliseconds for mobile devices to receive BS configuration updates. Therefore, it is impossibly difficult for mobile devices to acquire the newly updated BS configuration information instantaneously. Due to the traffic dynamics in individual slices, service QoS might be degraded if a sudden increase in user traffic happens. Furthermore, as it is impossible to predict individual user requests without error due to the stochastic nature of users, the traffic variation cannot be perfectly handled in advance [4]. In traditional RAN, the transmission employs a best-effort strategy without resource reservation, which cannot guarantee service QoS. For several low-latency applications (VR, automatic drive), best-effort would rise catastrophic service losses.

In the perspective of network providers, network slicing is born as an emerging business to operators by allowing them to sell the customized slices to various tenants at different prices [5]. In the perspective of tenants, tenants are aimed to maximize their profits from the customized slices leased from network providers. On the one hand, tenants provide their specific services to their users and acquire service fees. On the other hand, tenants need to pay the cost of network slices to network providers. Typically, service fees are dependent on the real user data traffic and expenses of network slices are depend on the number of physical resources [6].

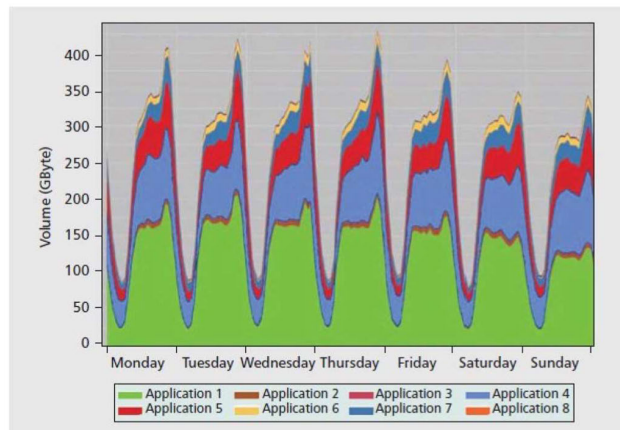


FIGURE 2. Measured weekly traffic volumes of a mobile cellular network [5].

The costly slices with more resources are better to guarantee service QoS. Nevertheless, such a choice may cause resource redundancy due to user traffic variation and decrease the tenant profit. Even in some cases, such as midnight when user data traffic is relatively low, the superfluous resource cost may exceed profits from users. For tenants, the accurate traffic prediction of data traffic can support to reduce the cost of redundant resources.

In previous researches, tenants will compete to acquire more wireless resources for their slices from network providers and gain more profits with the assumption that each tenant’s data traffic is large enough and can saturate its own slice capacity. However, peak-to-average-ratio of real network traffic is generally high, and a tenant should request wireless resources from network providers on demand to reduce cost. In 5G, enhanced Mobile BroadBand (e-MBB) is one of the typical application scenarios and requires high data rate and reliable broadband access over large areas. For an e-MBB tenant serving just a few active users, a small fluctuation of the number of active users may obviously influence the total data traffic of slice if each user’s data traffic is extremely high and the number of active users is small. Therefore, we focus on tenant profit improvement with the e-MBB user traffic model and slice management operation in this paper. We study slice resource reservation based on the traffic model of a specific application (VR), which is a typical e-MBB application, and propose a comprehensive tenant cost model from the slice tenant profit perspective. We find that edge users can greatly influence the total profit in several scenarios, where the user number is small but data traffic of each user is extremely large. Therefore, a fixed resource price strategy is not suitable, and we study some dynamic resource price strategy. In traditional cellular network, BS breathing behavior is an effective way to save system power consumption. Inspired by cell breathing, we propose two slice management operations, called slice merge/split. By slice merge/split, a tenant can dynamically adjust the slice BS coverage to match the real user distribution. Furthermore, we formulate

a tenant profit optimization problem and perform a simulation to verify slice merge/split operation effect on the tenant profits. We summarize our major contributions in three aspects.

- 1) We analyze slice resource reservation, propose dynamic resource price strategies in typical scenarios, and formulate the tenant profit optimization problem based on a comprehensive tenant cost model consisting of resource and basic operation expenses.
- 2) Two slice management operations, called slice merge/split across BSs, are suggested, whose differences with existing similar technologies (such as BS breathing) are pointed out in terms of control signal coverage and physical layer operation mechanism. The economic benefits and technology implementation complexity of these operations, which are performed independent of BS breathing, are discussed.
- 3) For the slice merge/split scenarios, the tenant profit optimization problem across adjacent BSs is formulated, in which inter-cell interference and system information broadcasting overhead are considered. A distributed control algorithm for profit optimization to make dynamic slice merge/split decision is proposed.

The rest of the paper is organized as follows. In Section II, we briefly review relevant researches that study network slice profit optimization. In Section III, slice framework and service attribute of our system model are presented. In Section IV, we analyze the slice resource reservation and propose the profit model in a cell. In Section V, we propose and study two slice management operations, called slice merge/split, formulate tenant profits optimization across adjacent cells, and perform typical simulations. Finally, the paper is concluded in Section VI.

## II. RELATED WORKS

Recently, researchers have investigated network slicing and sharing for multitenant cellular networks. In [7], the impact of network slicing on RAN in terms of resource utilization was discussed. In the perspective of the network welfare, Wang *et al.* [8], [9] proposed a resource allocation for network slices in 5G with network resource pricing, aimed at maximizing the network welfare of slice provider (SP) and slice customers (SCs) profits. Nonetheless, these two papers, in allusion to entire 5G network, did not perform detailed analysis on the BS slicing tenant profit problem. Online auction was applied to solve the profit maximization of entire network welfare [10]. [11] studied optimization of a network providers profits in RAN slicing with prediction of the arrival of two type slices for elastic and inelastic traffic. Sciancalepore *et al.* [12] proposed an optimization design based on network slice traffic prediction and admission control decision to maximize the system resource utilization. Reference [13] studied dynamic network slicing scheme for multitenant in heterogeneous RAN. Frequent slice reconfiguration incurred certain cost and might cause service

interruption and the cost model of network slice reconfiguration in CN was proposed in [4]. These papers mainly focused on the maximization of network utilization from the perspective of the network providers with enough tenant requirements, and less attention was paid to the single-tenant profit aspect.

The author in [14] suggested to reserve radio resources for each tenant and to admit users based on resource availability of the associated network slice, and similar schemes are used in this paper. In [15], a system with RAN slicing was considered, and each SPs competed with other SPs to auction non-overlapping orthogonal channels based on a machine learning approach. Reference [16] studied maximization of the C-RAN operator's revenue, including both long-term and short-term revenues, by properly admitting the slice requests. In detail, long-term revenues were defined by the slice parameters, and short-term revenue was obtained by saving system power consumption in each frame. These studies generally applied the sum achievable rate or power consumption as the equivalent of revenues, which was only theoretically possible income and different from slice tenants' real income by charging serviced users for real traffic. Han *et al.* [17] proposed the measurement of network slice revenue based on the specific service KPI. Reference [18] classed users into two types, which cared the bandwidth and bandwidth/latency, and these two different service types of users were charged different prices.

In general, all these aforementioned researches performed from the perspective of the network providers and mainly aimed at maximization of network welfare. The optimization of a single-tenant welfare has received litter attention, which was the main issue of our research.

## III. SLICE FRAMEWORK AND SERVICE ATTRIBUTE

In the following, we describe the various aspects related to our system model.

*Players:* In our system model, we focus on the maximization of tenant profit. Assuming that the network providers can provide slices with arbitrary resources and the total potential resources is sufficient, there are the following players:(i) the tenants, which request to the network provider to acquire network resources, and utilize these resources to offer specific service to their users, (ii) the users, which are subscribers of the service provided by a tenant and pay service fees.

*Network Model:* Our wireless cellular network is composed of a set of BSs  $\mathbf{B}$ , lies in the two-dimensional area  $\mathbf{A}$ . For each BS  $b \in \mathbf{B}$ , we let  $V_b$  denote the maximum amount of the resources that BS can afford. Let  $\mathbf{U}$  denote the set of users served by the tenant in the network. Each user is associated with and served by the BS with the best signal strength included in the slice.

*Traffic Model:* The users are sorted into groups according to their various QoS requirements. Assuming the users are provided a typical e-MBB service, VR video flow, and the data traffic follows a gamma-based model [19]. The users randomly arrive according to a Poisson distribution with

arrival rate  $\lambda$ . Traffic service time follows an exponential distribution with a mean  $1/\mu$ . VR video stream is a typical 5G application, and its I frames have relatively larger frame size than P frames. In general, P frame size is approximately one order of magnitude less than I frame size.

**Channel Model:** The channel response between the BS and the user is modeled by:  $h = \sqrt{l}g$ , where  $\sqrt{l}$  denotes a large scale fading coefficient between UE and BS.  $g$  stands for small scale fading coefficient.

**Network Slice Model:** The network can be logically divided in different network slices for each tenant, and different slices can camp in one BS. We assume that isolation between slices in the same BS can be ensured by assigning orthogonal resources to different slices. In our slice model, the tenant needs to choose the proper set of BSs and request the radio resources from them to fulfill the user rate requirements. Once the user data traffic decreases greatly, maintaining of redundant resources in slice cause unnecessary costs and tenant can request to reconfigure slice to release redundant resources or reselect BSs. However, the reconfiguration of the slice results in some QoS losses of connected users, and signaling overheads (e.g. new configuration information (MIB/SIB) need to broadcast). To guarantee service stability and isolation between slices, frequent reconfiguration is forbidden. Let  $T_{th}$  represents the threshold of the interval of reconfiguration. Such tenant network slices requests are characterized by:

- Slice service duration  $T$ : this is the length of the time interval of slice request from tenants.
- Reserved resource: the amount of RBs in each chosen BS.
- Cost: the cost a tenant should pay to the network providers for leasing resources. The price is per time unit, and the total cost is the sum of all chosen BS' cost. For each chosen BS, cost consists of base cost and resource cost.

We summarize the important parameters in Table 1.

**TABLE 1. Description of parameter symbol.**

Symbol	Description
$B$	a set of base station
$b$	one base station
$U$	the set of users in the network
$p_b^s$	cost of slice $s$ on base station $b$
$q_b^s$	revenue of slice $s$ on base station $b$
$\psi_b^s$	the profit of slice $s$ on base station $b$
$T$	slice service duration
$T_{th}$	the threshold of the interval of reconfiguration

## IV. ANALYSIS OF SLICE PROFIT MODEL IN A CELL

### A. PROFIT MODEL

In this section, we discuss profit problem of a slice operated only in single cell. In our system, a tenant should continuously pay for the base cost of the slice once the slice request is accepted and slice is created. The base cost is considered as compensation to the network provider for maintenance of

network equipment and network management, and can be assumed linear to the time length of slice time duration [6]. Meanwhile, the tenant should pay for the occupied radio resources of its slice. During the whole slice time duration, the tenant can dynamically request to reconfigure the number of radio resources to adjust to dynamic data traffic.

In this paper, a Physical Resource Block (PRB), consisting of 12 subcarriers and 1 slot, is assumed as the minimum resource unit that network providers can allocate to a slice and, every RB in the same BS is assumed to have the same price.  $\rho_b$  denote the price per RB, which is decided by the network providers, e.g., BSs with high power consumption or high load now can set higher  $\rho_b$ .

Let  $T$  denote the slice service duration. We assume that  $T$  is an integer multiple of transmission time interval (TTI), a minimum time unit of schedule. In BS  $b$  of slice  $s$ , the cost  $p_b^s$ , which a tenant pay for the network provider during  $T$  can be expressed as:

$$p_b^s = p_{b,base}^s + p_{b,rb}^s \quad (1)$$

$$p_{b,base}^s = \rho_{b,base} \cdot T \quad (2)$$

$$p_{b,rb}^s = \rho_b \cdot N_b^s = \rho_b \cdot \sum_{t \in T} n_b^s(t) \quad (3)$$

where  $p_{b,base}^s$  represents the base cost linear to slice time duration and  $p_{b,rb}^s$  represents the cost of required radio resources;  $N_b^s$  and  $n_b^s(t)$  represent respectively the number of required RBs of slice during whole slice time duration or per TTI.

In mobile communication network, charging by data traffic volume is widely used in network billing model (e.g., \$5 per GB). Meanwhile, some application service is charged by service duration, such as certain VPNs. As for diverse streams in e-MBB services, high data rate and massive data traffic volume are its typical characteristics. Thus, it is more rational to charge the services by their data traffic volume. In this paper, the traffic charge mode is used and, the revenue is a linear function of user traffic data volume.

The total revenue of slice  $s$  on BS  $b$  during its lifecycle is:

$$q_b^s = \rho_s \cdot R_b^s = \rho_s \cdot \sum_{t \in T} \sum_{u \in U} r_{b,u}^s(t) \quad (4)$$

where  $\rho_s$  represents price per bit of slice  $s$ ;  $R_b^s$  represents the amount of total data traffic volume during  $T$  and  $r_{b,u}^s(t)$  represents the total required data traffic of user  $u$  in  $t_{th}$  TTI. From Shannon equation, the minimum numbers of RBs allocated to user  $u$  in  $t_{th}$  TTI is:

$$k_{b,u}^s(t) = c \cdot \frac{r_{b,u}^s}{\log(1 + \frac{p_{b,u} |h_{b,u}|^2}{N_0})} \quad (5)$$

where  $p_{b,u}$  represents transmission power from BS  $b$  to UE  $u$ ,  $h_{b,u}$  represents channel gain from BS  $b$  to UE  $u$ ,  $N_0$  is the variance of white Gaussian noise, and  $c$  is a constant.

To guarantee the quality of service (QoS) that we need to meet,  $n_b^s(t) > \sum_{u \in U} k_{b,u}^s$  in each TTI.

Obviously, the profit of slice  $s$  on BS  $b$  is:

$$\begin{aligned} \psi_b^s &= q_b^s - p_b^s \\ &= \rho_s \cdot \sum_{t \in T} \sum_{u \in U} r_{b,u}^s(t) - \rho_{b,base} \cdot T - \rho_b \cdot \sum_{t \in T} n_b^s(t) \end{aligned} \quad (6)$$

To guarantee the slice isolation, the resource of one slice can not be unduly frequently reconfigured. Thus  $n_b^s(t)$  should be remain unchanged in  $aTTI * N_{th}^t$ , where  $aTTI$  denotes the duration of a TTI, and  $N_{th}^t$  is an integer equal to  $\lceil \frac{T_{th}}{aTTI} \rceil$ .

Furthermore, slice reconfiguration will cause extra overhead. When the slice reconfiguration occurs, the new configuration information will be delivered via broadcasting the new SIB1. Normally, UEs will initiate random access procedure to establish RRC connection and acquire specific physical downlink control channel (PDCCH), named control resource set (CORESET), once SIB1 is successfully received [20]. UEs will monitor their assigned CORESET to acquire their own downlink control information (DCI), indicating uplink resources of individual UE. Hereafter, it is not necessary for UEs to periodically receive SIB1 every 160ms, and BS can transmit specific RRC assignment to specific via PDCCH. However, the whole connected UEs in a slice should receive new SIB1 to acquire new CORESET if slice reconfiguration occurs, which will cause certain overheads. Moreover, the assigned CORESET positions in the time-frequency dimension can change significantly when the slices in a BS reconfigure like what happens in a handoff process, which may degrade the QoS of ongoing service. In our proposal model, to alleviate the loss of QoS, BS uses PDCCH to inform UE before the reconfiguration and UE can send reference signal beforehand to estimate the channel of new assigned frequency band. Nevertheless, the reconfiguration will result in extra overhead costs. In the perspective of the network providers, slice reconfiguration will cause extra overhead and it can increase the base cost or RB per price, which can drive a tenant to tend to request fewer reconfiguration.

**B. TRAFFIC BURSTINESS AND DYNAMIC RESOURCE PROVISION FOR eMMB: VR SLICE INSTANCE**

Due to the huge traffic of e-MMB services, the cost of wireless resources is very high. On the other hand, a big difference exists between peak and trough data rate. Peak data rate only appears sporadically, and e-MBB service is bursty. Therefore, it is potential for tenants to reserve wireless resources dynamically to save resource cost. In this section, taking the typical VR service as an example, and based on a well-known mathematical model of video traffic and transmission delay requirement, we give the calculation method of VR slice resource usage, which provides the basis for various simulations in subsequent sections. Thus, we assume that the slice provides VR service to users. The payload data in a VR streaming is composed of I-frame (intra-picture) and P-frame (predicted-picture) video packets as video compression output at VR server. The average size of I-frame is multiple of P-frame, and I-frame presents sporadically in the video frame sequence. Unsuccessful decoding of I-frame

will cause the decoding of the entire group of picture (GOP) to fail. Consequently, I-frame with great influence to QoS causes a bursty traffic increase. Field-of-view (FOV) is conveyed by a sequence of video frames, which are supposed to be compressed by H.264 standard with 1080p resolution and 75 frame-per-second (FPS) frame rate. The average size of one I-frame is 500 kilo-Bytes, while the average size of one P-frame is 50 kilo-Bytes [21]. The time interval between two frames (regardless of I-frame or P-frame) is  $\frac{1000}{75} \approx 13.3$  milliseconds. In a GOP, the number of I frames compared with P frames is 1 to 29. A gamma-based model [19] is used for video frame size in this paper. Figure.3 illustrates an example.

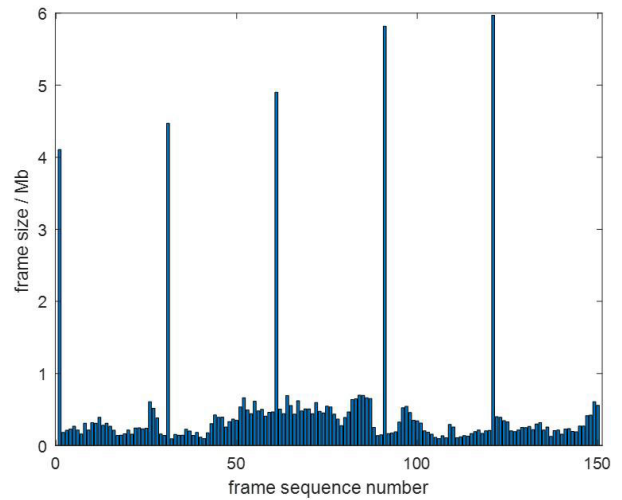


FIGURE 3. Frame sequence follows a gamma-based model.

In the perspective of a tenant, user data traffic will increase suddenly during transmission of I-frame, and keep relatively much lower in the rest time of GoP. Thus, a tenant can lease relatively fewer resources from network providers to save cost when transmitting P-frame. However, it is necessary to ensure that I-frame is successfully transmitted, since failure in transmitting I-frame will cause failure in transmitting subsequent P-frames. To this end, redundant reserved resources for I-frame is reliable for QoS requirement though it will increase tenant resource cost. Therefore, the accurate prediction of I-frame size and arrival time can effectively reduce the cost of resources and increase the profit.

Let  $n_{fps}$  denote the frames per second of VR video stream. To avoid the visual perceptual loss, the successful transmission and decoding of the next frame should be completed within  $\frac{1}{n_{fps}}$  from the start of play of the current frame play. Assuming that maximum of the sum of other delays (e.g., CN transmission delay and decoding delay) are constant, the transmission delay constraint in BS is  $t_{constraint} \leq \frac{1}{n_{fps}} - t_{ods}$ , where  $t_{ods}$  denotes the sum of other delays. Let  $D_t = aTTI \cdot INT(t_{constraint}/aTTI)$  with  $aTTI$  denoting the duration of a TTI, and  $INT(t_{constraint}/aTTI)$  being the upper bound of number of TTI that  $t_{constraint}$  can have.  $D_t$  is the

upper bound of the transmission time that a video frame can be allocated.

When next video frame of length  $m_u$  bits of user  $u$  arrivals at BS in the  $t_u$ th TTI, the required number of RB is:

$$k_{b,u}^s = c \cdot \frac{r_{b,u}^s(t)}{\log(1 + \frac{p_0 \cdot h_u^2}{N_0})}$$

$$= c \cdot \frac{m_u/D_t}{\log(1 + \frac{p_0 \cdot h_u^2}{N_0})}, t \in [t_u, t_u + D_t) \quad (7)$$

The total resource demand of this slice with set of users  $U$  in  $t_u$ th TTI is defined as:

$$k_{b,U}^s(t) = \sum_{u \in U} k_{b,u}^s(t) \quad (8)$$

### C. PRICING STRATEGY

The simulation parameters are summarized in Table.2. Fig.4 illustrates the accumulative revenues during 2 hours. In fig.4, the revenue is negative when the user arrival rate is lower and service charge can not compensate the base cost of slice.  $\alpha$  is the large scale fading factor.  $\frac{\rho_s}{\rho_b}$  is the ratio of price of user data traffic and RB cost decided by the average channel conditions and we set  $\frac{\rho_s}{\rho_b} = \rho_0 \cdot c_{ration}$  for this simulation. Abnormally, the obvious fluctuation appears with high user arrival rate. In other words, the revenue decreases at some occasions when the user arrival rate rises. The reasons are as follows: in VR video scenario, the traffic flow of each user is orders of magnitude larger than traditional scenario. The user with bad channel condition consumes orders of magnitude RBs larger than the one with good channel condition even

TABLE 2. Used simulation parameters.

Symbol	value
$T$	7200second
$\mu$	20minute
$\rho_0$	$3.4528e - 08$
$\alpha$	2

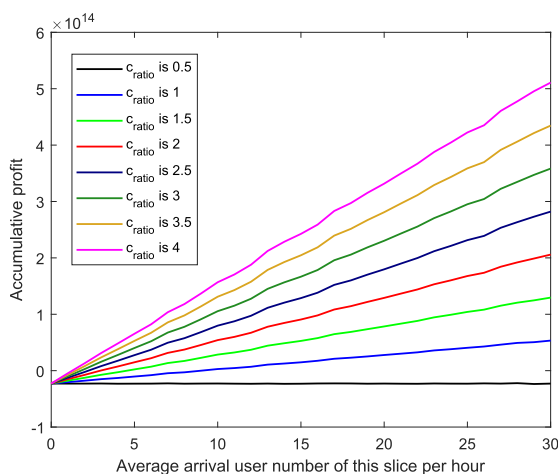


FIGURE 4. Revenue vs different user arrival rate with fixed price per RB.

their traffics is the same. The profit of a slice with more users, but mostly edge users, is lower than the profit of a slice with fewer users but located in the center of the cell. In this case, a tenant provides more traffic services, contrarily earns less profit. Therefore, the fixed price per RB strategy may not be fair and suitable to a certain tenant, especially in scenarios with relatively fewer user numbers, like VR, since it overemphasizes the cost of edge users with relatively low profitability ratio. Users with highly-guaranteed QoS service, are willing to pay relatively higher service fees to the tenant for more reliable service, which can bear high price of potential edge resource consumption. This factor should be considered in the way that a tenant charges its users. From the aforementioned analysis, in order to alleviate the impact of edge users on the slice profit, tiered RB price strategy are proposed.

For tenants, the designed price strategy must insure that the service revenue of edge users exceeds the cost of RBs; otherwise, the slice should reject the user service request, which is an explicit strategy for user connection admission control in a slice. Such a strategy may also be adopted in the traditional mobile networks, e.g., in LTE, the assignment algorithm of RB will limit the amount of RBs for edge users to increase cell traffic capacity. However, it will extremely degrade the QoS of edge users. In 5G, massive MIMO fully exploits the potential capacity in the spacial domain, and beam-based spatial multiplexing improves spectrum efficiency and reduces inter-cell interference. Compared with LTE, the total number of RBs in one cell is relatively sufficient. However, the total transmission power in a BS is limited, and the power resource is relatively scarce. Thus, the total amount of available RBs is still limited. Instead of limiting the number of RBs occupied by edge users as in LTE, the customized slice should be encouraged by a reasonable price strategy to provide better QoS for edge users in 5G. To ensure fairness to tenants, the tiered pricing strategy for RB is better and the average price of RBs will be cheaper when the required amount of RBs during the unit interval (e.g., 10ms) is larger.

Considering that two UE have the same traffic data volume but different channel conditions, the small scale fading model in this paper has time ergodicity. Thus, the long-term RB consume ratio relies mainly on the large scale fading, which is dependent on the distance between UE and BS. The average channel gain is determined by the distance between UE and BS [15]. In this paper, large scale fading coefficient  $l$  equals  $d^{-\alpha}$  according to an exponential fading model, where  $d$  denotes the distance between UE and BS. Stair price strategy is a type of cumulative discounts, which is a discount scheme that can be cumulative in effect [22]. Moreover, stair price strategy provides a fixed discount to the edge users, of which distances to BS are larger than thresholds. Furthermore, a dynamic discount factor can make a price strategy with more flexibility [23]. The logarithmic price strategy and negative exponential pricing strategy are proposed in this paper. In these two strategies, a discount factor is decided by the network providers during a discount

duration (e.g., 10ms or 20mins) and its value can be adjusted in different discount duration. The network providers can set several discount RB number thresholds, as  $[dis_0, dis_1, \dots]$  in these strategies and price of RBs in one threshold is different from RBs in other thresholds.  $dis_j$  represents a specific number of RBs. As for logarithmic pricing strategy, the price  $\rho_b$  of RBs in  $[dis_j, dis_{j+1})$  is  $\rho_b^{(j)} = \frac{\epsilon_{log}}{\log N_j}$ , where  $\epsilon_{log}$  represents the discount factor decided by the network providers and  $N_j$  depends on the RB number, like  $N_j = \frac{dis_j + dis_{j+1}}{2}$ . In negative exponential pricing strategy,  $\rho_b^{(j)} = (\epsilon_{exp})^{-N_j}$ . Particularly, the size of each discount interval can be set to 1. The tenants can acquire more revenue with the user traffic volume (required RB number) increases by these three price strategy. In the perspective of a tenant, stair price strategies offer the least revenue. Meanwhile, the logarithmic price strategy, which can be regarded as a two-order stair price strategy, offers the most revenue. The revenue of a negative exponential pricing strategy is between the aforementioned price strategies. However, mild discounts such as stair price strategy may be more acceptable to network providers in reality. Since the base cost for slices is adopted in our price model, network providers can ensure a future profit serving tenants with low RB requirements. A negative exponential pricing strategy is relatively flexible by adjusting the negative exponential factor. By providing discounts to tenants, network providers promote tenants to consume more resources, and their own profits can also increase.

The simulation is performed to analyze the effect of different price strategies on tenant profits. The simulation parameters are set as follow: the user arrival rate is 0.9, discount factor of negative exponent price strategy is 1.05 and the percentage of edge users is 0.15. Admitting all users with no user rejection is simulated as a comparison. The simulation result of accumulative tenant profit, cost and rejected user number in 2 hours are shown as fig.5-7. From fig.5-7, it is shown that when the user arrival rate is relatively high (0.9) the suitable discount can both increase the tenant profits and

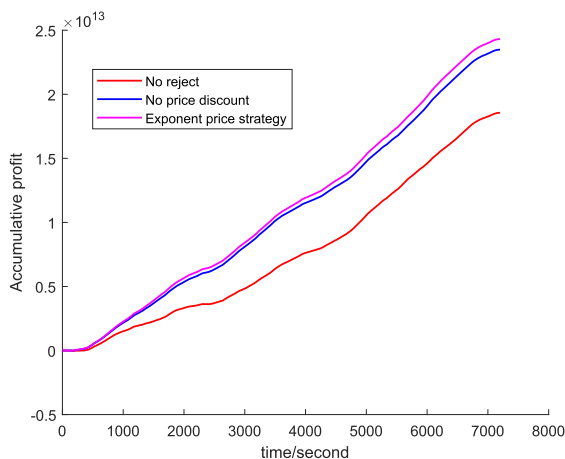


FIGURE 5. Accumulative tenant profits in different price strategy.

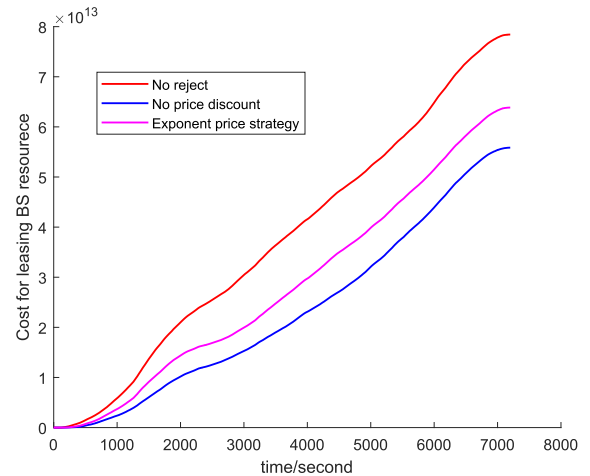


FIGURE 6. Accumulative cost paid to the network provider.

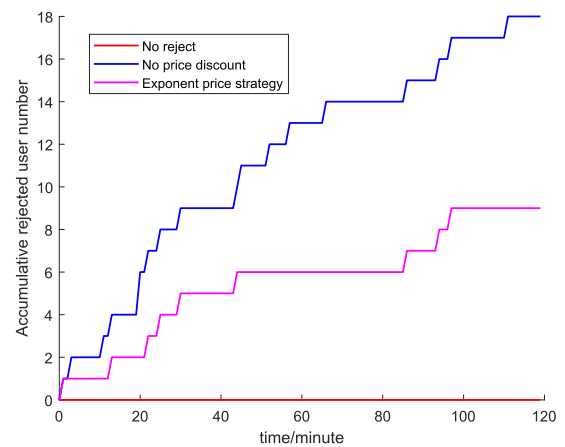


FIGURE 7. Accumulative rejected user number.

the network provider incomes. Meanwhile, it can decrease the rejection rate of user access, which improves the user quality of experience.

Although tiered price strategies are more reasonable, it is not a easy task to determine tiered price in the actual network. For the sake of research convenience, the resource price is only set depending on the radius of cell coverage in the rest of this paper. The resource price is higher in the cell with larger radius since expanding coverage will cause more power consumption, and these BSs can set higher  $\rho$  as aforementioned.

### V. SLICE MERGENCE/SPLIT OPERATIONS AND TENANT PROFITS OPTIMIZATION ACROSS ADJACENT CELLS

As aforementioned, single slice profit will be negative on several occasions (e.g., at night or in suburbs) since the user arrival rate is low and, the revenue can not exceed the base cost for the slice. From a tenant perspective, maintaining the slice will cause more profit losses, and a tenant can decide to terminate this slice lifecycle. To avoid service outage, the area originally covered by a newly closed slice is taken over by adjacent slices of the same type, which can prevent destructive loss of QoS to the active users. Such operation is

called Slice Mergence in this paper. Slicing mergence offers an effective scheme to reduce slice operation costs for tenants in a multi-cell scenario. Conversely, when the user data traffic served by a slice with a relatively large coverage area dramatically increases and becomes greater than a certain threshold, Slice Split can be exercised. By Slice Split, the large coverage slice can be split into multiple small coverage slices, and each small coverage slice obtained by the split serves only a part of the original users. Generally, the slice merging and splitting operations are similar to the breathing behavior of conventional cellular BS management, and it enables the slice deployment to adjust to the state of the network dynamically. The technical characteristics and economic utility of slice merging and splitting operations are analyzed below.

#### A. ANALYSIS OF SLICE MERGENCE PROCESS AND IMPACT

In conventional cellular BS management, breathing behavior is a way of dynamically adjusting BS state to save energy. The BSs with low traffic data volume will switch to low power mode, such as completely switch-off [24] or opportunistic sleep [25], and adjacent BSs will take over its original coverage. However, a tenant only concerns traffic data of its own users, and user data traffic of other types of slices are not relevant to its profit. Considering that BS sleep mode can only be deployed when the total payload of the BS is low, several frequency reuse technologies, such as FFR [26], SFR [27], are devised to dynamically expand the specific BS coverage for the takeover of users nearer to other BS. In detail, slice mergence is not as same as these BSs and FFR/SFR scheme in the physical layer, and it can be achieved by different physical layer facilities of network providers.

In the proposed scheme, a tenant, leasing several slices in an area, sends a request of slice mergence to the network providers when user data traffic of one or more slices is low and obtains negative profits. Then, the slice mergence process will start once the network providers accept the request. The initial slice will firstly broadcast specific SIB to inform the connected users to prepare for slice mergence. As fig.8 illustrates, the MIBs and SIBs of different BSs deployed with slices (red and blue) are both broadcast in the blue BS coverage. Such overlapping of broadcast channel coverages also exists in ultra-dense network and they are assigned orthogonal resources in time, frequency or spatial dimension [28]. When a slice is merged by adjacent slices of the same type, the coverage range of the adjacent slices (called donor slices) will expend to cover the merged one, and new system information will be broadcast throughout the new coverage area via SIB. However, users there access to the adjacent slices, maybe consuming more RB(or power) than they consume in the previous slice due to the farther distances to serving BSs. Thus, a tenant will lease more resources to serve these users, and pay more for network providers. It is a rational way to compensate for power consumption or signal overhead for network providers. Nevertheless, the total cost of the tenant is reduced since the base cost saved by closing a slice is greater than the increase of the operating cost caused

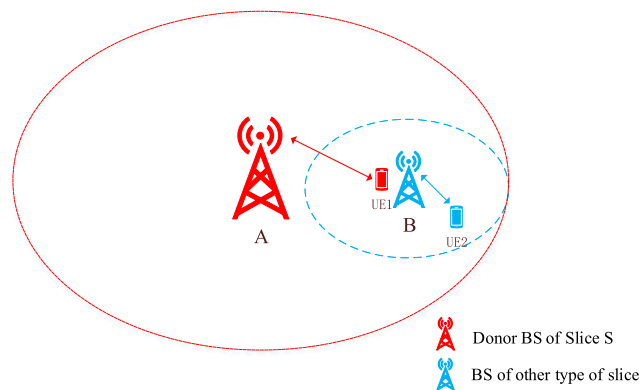


FIGURE 8. An illustrate example of slice mergence of two cell.

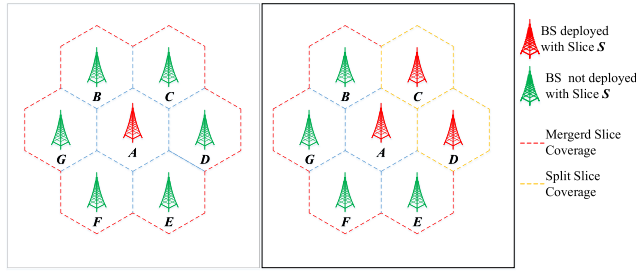
by the expansion of the donor BS coverage range. On the other hand, a tenant triggers slicing mergence to increase its own profit without considering the overall status of BS with various slices. Slicing mergence will effect object BSs status and other various slices deployed in these BSs. Meanwhile, slicing mergence prompts a BS with a specific type of slice to expand coverage to areas still covered by adjacent BSs with other types of slices, increasing inter-cell interference and energy consumption. However, the negative impact is small and acceptable since slice mergence only occurs when the network traffic load is relatively low, and various technologies can be applied to alleviate inter-cell interference.

In this paper, our research focuses on the slice mergence/split operations of one single tenant. Assuming that the data traffic of different application slices is independent and the quantity of users or data traffic of other types of slices is invariable or irrelevant, adjacent BSs themselves will not be entirely emerged/split, whereas adjacent BSs remains switch-on serving other types of slices. To serve the users in merged slice, more resources (RBs and power) will be consumed and reduce the resources available to other types of slices in the adjacent BS. Furthermore, the network providers can decide whether the slice mergence request is adopted based on the BS payload status. Thus, slice mergence can not be accepted in BS of high payload status of other different slices.

#### B. ANALYSIS OF SLICE SPLIT PROCESS AND IMPACT

Slice split is the inverse operation of slice mergence, which split a slice with broad coverage to several small slices. As similar to switch-on operation of sleeping BSs to increase system capacity when the load rises approaching certain threshold, the tenant requests to open a new but small slice in adjacent BS when a large slice's data traffic load trends to exceed slice's capacity. Moreover, areas originally covered by a larger slice will be divided to several small areas individually served by different small slices. As a tenant, it can earn more profits by serving more users. From another perspective, costs can be reduced since the cost of RBs is comparable to the base cost of opening a new slice when the





**FIGURE 9.** An illustrate example of effect of slice merge and slice split in cellular network. The aggregated data user traffic of type  $s$  slice in this area is very low and the slices are merged to a large slice deployed in BS A. The large slice is split to three small slice, separately deployed in BS A, BS C, BS D, since the data user traffic in cell C and cell D grows.

proportion of edge users rises highly in a large slice. With decreasing the average distance of users to BSs that they associated with, the cost of resources (RBs or power) reduces and the cell interference is alleviated, which is consistent with the network providers' intention to increase system capacity. Consequently, slice split can both increase revenue and reduce cost when the user data traffic is high. However, slice split may not be achievable when the payload of objective BSs is high and the resources is not enough to create new small slices.

**C. PROBLEM FORMULATION**

**1) PROBLEM ANALYSIS**

Different from the problem analyzed in single slice presented in Section III, the transmission power and adjacent cell interference should be concerned in slice merge. As for a tenant, the slice merge can reduce the base cost of BS, nevertheless this operation may aggravate inter-cell interference and reduce network system capacity, which network provider may be unwilling to abide. (5) in multi-cell scenario is modified to:

$$k_{b,u}^s(t) = c \cdot \frac{r_{b,u}^s}{\log(1 + \frac{p_{b,u} \cdot |h_{b,u}|^2}{\sum_{b' \in B \setminus b} p_{b',u} \cdot |h_{b',u}|^2 + N_0})} \quad (9)$$

As compensation, the price per RB for users in the merged slice can be moderately raised by the network providers, and an increase of base cost for merged slice can be enforced by the network providers with an alternative price option. Furthermore, the change of the slice user traffic served by a specific tenant is not completely consistent with the change of the BS payload. The implementation of the slice merging request by one tenant may degrade the interests of other types of slice tenants since the total network traffic capacity decreases. To ensure sufficient network traffic capacity, the slice merging request can be refused by the network providers when the BS payload in the merged area is large enough.

Slice merge/split operation is generally accepted when the network traffic load is relatively low. We assume that

enough orthogonal frequency/spatial resources are provided when slice merge/split operation happens and users can be assigned in orthogonal beams for a tenant.  $\sum_{b' \in B \setminus b} p_{b',u} \cdot |h_{b',u}|^2$  can be negligible and price per RB is raised for inter-cell interference in slice merge/split scenarios.

As for a tenant, when it detects user data traffic is low and the profit is negative for a given duration it will decide to merge slice. The profit of single slice is presented as equation.(6). Considering the alternative price strategy for network providers, it is modified as:

$$\begin{aligned} \psi_b^s &= q_b^s - p_b^s \\ &= \rho_s \cdot q_b^s - \alpha_{b,base} \cdot p_{b,base}^s - \alpha_{b,rb} \cdot p_{b,rb}^s \end{aligned} \quad (10)$$

where  $\alpha_{b,base}$  and  $\alpha_{b,rb}$  represents respectively the price factor of base cost and RB cost for the merged slice decided by the network providers. Both  $\alpha_{b,base}$  and  $\alpha_{b,rb}$  is no less than 1 since a tenant should pay more for the merged slice with larger coverage range and occupation of adjacent cell resources. We consider that  $\alpha_{b,base}$  and  $\alpha_{b,rb}$  are linear to the cover area of the slice in this paper.

The total profit of multiple slices during the slice lifecycle is denoted as:

$$\psi_{B_s}^s = \sum_{b \in B_s} \psi_b^s \cdot a_b \quad (11)$$

where  $B_s$  donates the set of the whole BSs within coverage of slice  $s$ , and  $\psi_b^s$  is the profit of slices in one BS, which is zero if this BS is not requested during this slice lifecycle or merged by adjacent BSs.  $a_b \in \{0, 1\}$  is the indicator of BS  $b$ , that is decided by the slice merge/split strategy. The indicator  $a_b$  is 1 when the slice of type  $s$  is deployed in the BS  $b$ ; otherwise the indicator is 0, such as type  $s$  in the BS  $b$  is closed or merged and profit of slice  $s_1$  will change according to (12) when slice  $s$  has been merged into slice  $s_1$ .

To maximize the profit of the tenant, the slice merge/split operations will be performed according to the real user data traffic distribution, which can be represented by determining the value of indicator  $a_b$  in mathematical expression. Thus, vector  $\mathbf{a} = [a_1, a_2, \dots, a_{|B_s|}]$  is the vector of the BS indicator.

$$\begin{aligned} \max_{\mathbf{a}} \quad & \psi_B^s \\ \text{s.t.} \quad & \sum_{s \in S} N_b^s \leq N_{b,rb}^{max} \\ & \sum_{u \in U_b} P_u \leq P_{b,rb}^{max} \end{aligned} \quad (12)$$

where  $N_{b,rb}^{max}$  reflects the maximum of resource that can be occupied by the total slices in the BS  $b$ . It is dynamically decided by the network providers based on the overall network load in the area. Both slice merge request and slice split request are refused when the desired resources in objective BSs exceed  $N_{b,rb}^{max}$ .  $P_{b,rb}^{max}$  is the max transmission power of BS  $b$ , which the sum of assigned power for all users can not exceed.

The problem in (12) can be classified as an NP-complete 0-1 knapsack problem. We assume that the slice can only be merged by the adjacent slices to avoid serious inter-cell interference. Thus, the dimension of state space is dramatically reduced. In reality, one BS is usually occupied by many different tenants and one tenant can not acquire global network load state information since user traffic records are important digital assets and business secrets of tenants. Therefore, executing the algorithm distributed in each BS is more suitable for solving the problem.

The execution cycle of the merge/split decision equals reconfiguration time threshold  $T_{th}$ , to decide whether data traffic in a certain area slice is lower/higher than a certain threshold. If the aforementioned condition is satisfied, the decision will be made according to formula (12), and the decision result will determine whether to perform merge/split operation. This procedure is shown in Fig.10, where each slice will independently manage its operation. We consider that a distributed strategy can be widely used in universal scenarios. In reality, a slice may belong to a specific tenant, like a company. In other way, a slice may be discriminated by its types of applications, like VR slice, which may simultaneously serve many different commercial organizations. Other tenants will not share their decision information to compete for business and a tenant will not be able to know state of slices owned by other tenants in the adjacent cell. Assuming that the network providers can completely decide the acceptance or rejection of reconfiguration requests, it is convenient to solve the profit problem without considering dependency management. In this paper, the total resource in each BS is fixed, which is donated by  $N_{b,rb}^{max}$ . We define load factor  $\gamma_b$  to reflect the load state of BS  $b$ .

$$\gamma = \mathbf{E} \left\{ \frac{\sum_{s \in S} N_s^b}{N_{b,rb}^{max}} \right\} \quad (13)$$

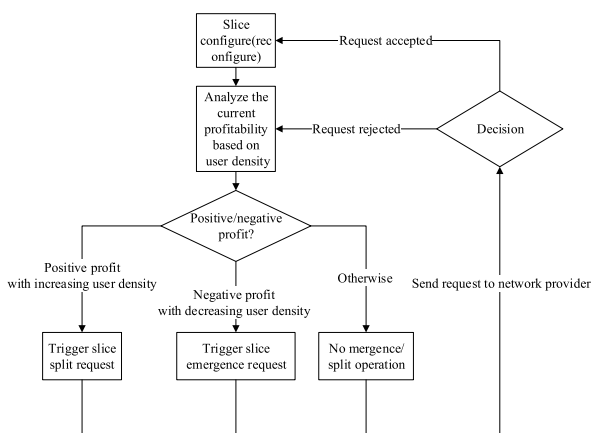


FIGURE 10. The procedure of slice reconfiguration.

The network providers will reject the slice split or merge request if the load factor of the target BS  $b$  maintains higher than the load threshold  $\theta_b$ . Nevertheless, this factor will not

influence the trigger of the slice split or merge request by a tenant.

## 2) ALGORITHM DESCRIPTION

In this section, we propose a distributed BS switching algorithm and suggest its protocol-level implementation.

*Slice Merge/Split Algorithm:* The decision criterion defined in (12) only depends on information for a slice and its neighboring the same type of slices. Thus, it is possible that the emerge-split decision can be localized as a problem at each BS deployed with this type of slice. The proposed distributed emerge-split algorithm is simple: a slice detects its user arrival rate keeping increasing or decreasing, and predicts the trend will continue. Thus, it can decide to trigger slice split or slice merge. The system information such as slice assignment and system load are periodically shared among BSs and UEs, and each BS decides whether it should be emerged or split by the slice. The slice merge/split request will be controlled by the network providers. If the request is refused, the slice will maintain its previous configuration for a certain period and restart emerge-split decision.

The slice split algorithm involves two parts as follows:

- a. **Initial request:** The user arrival rate of a big merged slice keeps increasing and active user data traffic is approaching to the service load limitation. To increase service capacity and acquire more profits, the tenant requests to split its big merged slice into some small slices.
- b. **Request control:** Slice split operation will not aggravate inter-cell interference since overlap of BS coverage is reduced and small subslices are individually served by its own BSs. Network providers are generally willing to accept split request except when the load state of objective BSs is high.

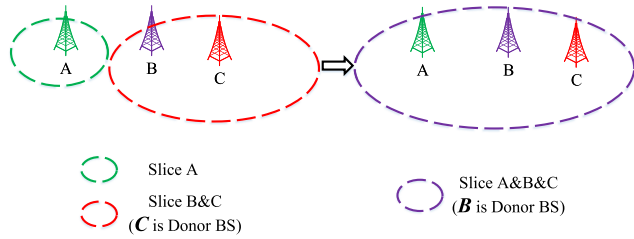
Compared with slice split, the slice merge algorithm is relatively complex and involves three parts as follows:

- a. **Initial request:** A slice detects few active users and gain negative profits for a certain time. The slice request to merge into one of its neighboring the same type of slices.
- b. **Target slice decision:** The slice will detect all its neighbors with the same type of slices and make it a priority to merge into a neighboring slice with the highest potential profit after merge.
- c. **Request control:** Network providers decide whether to accept or reject slice merge requests depending on overall network state and load of target BSs.

We find that when a slice request to merge into a merged slice, the donor BS may change to alleviate inter-cell interference as an example shown in fig.11. Meanwhile, the load of target BS will greatly increase if it becomes donor BS in a merged slice. Thus, we come to Principle 1.

*Principle 1:* The donor BS of a merged slice must be the BS that is adjacent to all other BSs.

Different from BS breathing completely switching off target BSs, BS still works, and services for other types of slices



**FIGURE 11.** An example of donor BS change in slice merger. BS C is donor BS of slice BC and BS B becomes donor of slice ABC.

after a specific slice is merged. As example shown in fig.11, BS C is the donor BS of slice B&C. Slice A request to merge into neighboring slice B&C, and BS A is only adjacent to BS B. To alleviate inter-cell interference, BS B becomes the donor BS in slice A&B&C. Nevertheless, the switch of donor BS causes have-over overheads in slice management. Donor BS switching may even happen with no slice mergence in a big merged slice to adjust to real user location distribution. To compensate for these overhead costs, we come to Principle 2.

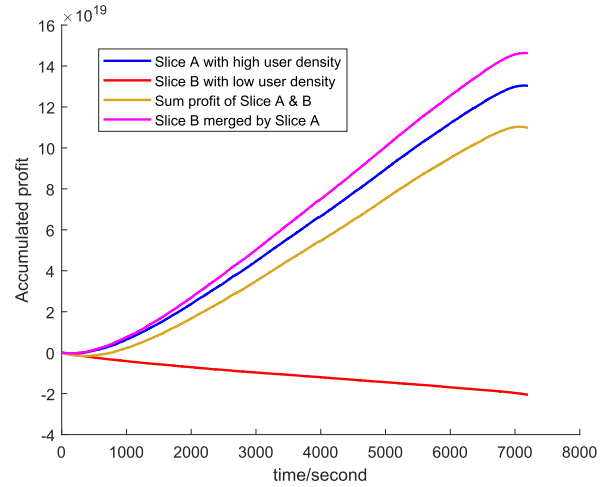
*Principle 2:* The base cost of a merged slice should be raised according to the number of merged BSs and the network topology.

Slice mergence is proposed as a slice management operation for tenants to save cost and increase profits. The average distance between donor BS and users will be increased once a huge merged slice is formed, causing potential QoS degradation of edge users. A reasonable price strategy should be determined to promote tenants to achieve a trade-off between benefits and service quality. In the aspect of network providers, the management of a big merged slice can be more complex and hand-over overheads increases with the number of BSs in merged slice. The bast cost of donor BS in merged slice should be raised to compensate extra management overhead of network providers. The base cost will be raised according to the expanded coverage size of donor BS in merged slices. In a network consisting of heterogeneous BSs, BS attribution will be a factor in pricing, e.g., a tenant should pay more for a macro BS than a micro BS.

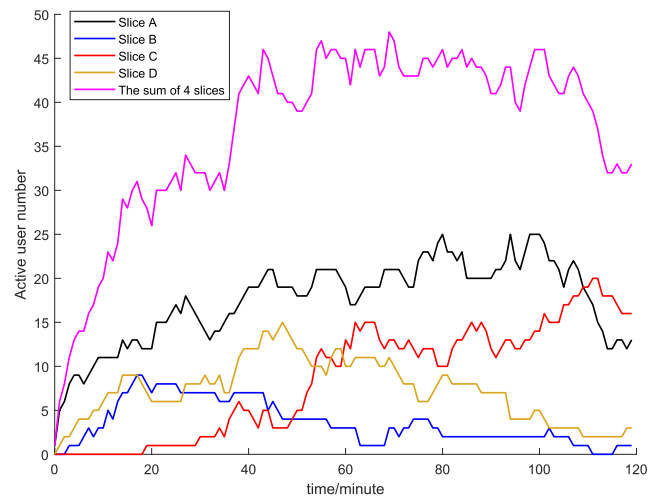
**D. PERFORMANCE**

In our simulation, we firstly consider the mergence of two adjacent slices, called slice A and slice B. The cell radii of these two slice are both 200 meters. Slice A, with a high user arrival rate, 0.75, persistently makes profits throughout the slice duration. Meanwhile, slice B, with a low user arrival rate 0.1, only causes losses of profits. Fig.12 illustrates that the total profit increases by merging slice B into slice A, forming slice A&B.

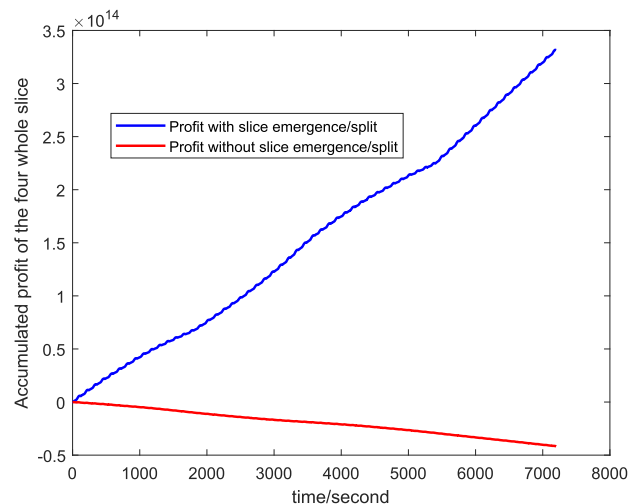
Then we perform simulation to analyze the performance of slice mergence/split in a complicated scenario. We consider 4 cells and each BS serves many different tenants. We focus on only one tenant, and other tenants in the simulation are



**FIGURE 12.** Comparison of profit between the merged two slices.



**FIGURE 13.** The active user number of 4 slices.



**FIGURE 14.** The sum of accumulative profit of 4 slices.

considered to be aggregated to a big tenant, and they will influence the load state of BSs.

The active user number of each slice is shown in fig.13. In fig.13, it shows that slice *B* data traffic kept decreasing causing negative profit. Thus, slice *B* was merged in to slice *A&C*, combining into slice *A&B&C*. Then, active user number of slice *C* increased, triggering slice split, and slice *A&B&C* was split into slice *A&B* and slice *C*. Next, slice *D* was merged in slice *D*, forming slice *C&D*, when it had few users. In fig.14, it illustrates the variation of total accumulative profit with or without slice merge/split operation, and indicates that slice merge/split operation can increase tenant profits.

## VI. CONCLUSION

In this paper, we focused on the problem of slice management for maximizing profits of slice tenants. Compared with the traditional slice scheme, in which slices were deployed consistently with the BS physical topology, we considered that a slice can overlap several cells according to the actual traffic load of the total network. In particular, we suggested a slice adjusting scheme based on two suggested slice operations, named slice split and slice merge. Taking into account the implementation difficulty, we proposed a model to maximize the profit of a tenant. We considered the loss of network system performance due to inter-cell interference, as an extra cost to a tenant. Furthermore, we proposed a distributed algorithms performed by each slice, and the network provider can control the access of slice management request.

The problem will be more complex when slice deployment are inconsistent with the physical BS deployment. The model we proposed is initial. We assumed that the user numbers/data traffic volume of other types of slices were invariable or irrelevant, and adjacent BSs themselves with other types of slices would not perform merge/split. Furthermore, the price strategy of RB should be different in varieties of scenarios, and accuracy of measurement of cost caused by slice merge/split should be improved in the model. For the merge/split of adjacent BSs, the problem of slice merge/split with various types of slices needs to be further studied.

## REFERENCES

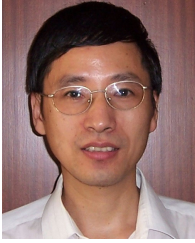
- [1] M. Yan, G. Feng, J. Zhou, Y. Sun, and Y.-C. Liang, "Intelligent resource scheduling for 5G radio access network slicing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7691–7703, Aug. 2019.
- [2] X. Foukas, M. K. Marina, and K. Kontovasilis, "Orion: RAN slicing for a flexible and cost-effective multi-service mobile network architecture," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2017, pp. 127–140.
- [3] C. Sexton, N. Marchetti, and L. A. DaSilva, "Customization and trade-offs in 5G RAN slicing," *IEEE Commun. Mag.*, vol. 57, no. 4, pp. 116–122, Apr. 2019.
- [4] G. Wang, G. Feng, T. Q. S. Quek, S. Qin, R. Wen, and W. Tan, "Reconfiguration in network slicing—Optimizing the profit and performance," *IEEE Trans. Netw. Service Manage.*, vol. 16, no. 2, pp. 591–605, Jun. 2019.
- [5] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.
- [6] L. A. DaSilva, "Pricing for QoS-enabled networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 3, no. 2, pp. 2–8, 2nd Quart., 2000.
- [7] I. da Silva, G. Mildh, A. Kalokylos, P. Spapis, E. Buracchini, A. Trogolo, G. Zimmermann, and N. Bayer, "Impact of network slicing on 5G radio access networks," in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun. 2016, pp. 153–157.
- [8] G. Wang, G. Feng, W. Tan, S. Qin, R. Wen, and S. Sun, "Resource allocation for network slices in 5G with network resource pricing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.
- [9] G. Wang, G. Feng, S. Qin, R. Wen, and S. Sun, "Optimizing network slice dimensioning via resource pricing," *IEEE Access*, vol. 7, pp. 30331–30343, Mar. 2019.
- [10] L. Liang, Y. Wu, G. Feng, X. Jian, and Y. Jia, "Online auction-based resource allocation for service-oriented network slicing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8063–8074, Aug. 2019.
- [11] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, and X. Costa-Perez, "A machine learning approach to 5G infrastructure market optimization," *IEEE Trans. Mobile Comput.*, vol. 19, no. 3, pp. 498–512, Mar. 2020.
- [12] V. Sciancalepore, K. Samdanis, X. Costa-Perez, D. Bega, M. Gramaglia, and A. Banchs, "Mobile traffic forecasting for maximizing 5G network slicing resource utilization," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, May 2017, pp. 1–9.
- [13] Y. L. Lee, J. Loo, T. C. Chuah, and L.-C. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2146–2161, Apr. 2018.
- [14] T. Guo and R. Arnott, "Active LTE RAN sharing with partial resource reservation," in *Proc. IEEE 78th Veh. Technol. Conf. (VTC Fall)*, Sep. 2013, pp. 1–5.
- [15] X. Chen, Z. Zhao, C. Wu, M. Bennis, H. Liu, Y. Ji, and H. Zhang, "Multi-tenant cross-slice resource orchestration: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2377–2392, Oct. 2019.
- [16] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast eMBB," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.
- [17] B. Han, S. Tayade, and H. D. Schotten, "Modeling profit of sliced 5G networks for advanced network resource management and slice implementation," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2017, pp. 576–581.
- [18] S. Chen and B. Krishnamachari, "Differential pricing of 5G network slices for heterogeneous customers," in *Proc. 10th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2020, pp. 0705–0711.
- [19] M. Frey and S. Nguyen-Quang, "A gamma-based framework for modeling variable-rate MPEG video sources: The GOP GBAR model," *IEEE/ACM Trans. Netw.*, vol. 8, no. 6, pp. 710–719, Dec. 2000.
- [20] *NR: Physical Layer Procedures for Control (Rel-15)*. Accessed: Sep. 2019. [Online]. Available: [http://www.3gpp.org/ftp/Specs/archive/38\\_series/38.213/](http://www.3gpp.org/ftp/Specs/archive/38_series/38.213/)
- [21] M. Huang and X. Zhang, "MAC scheduling for multiuser wireless virtual reality in 5G MIMO-OFDM systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.
- [22] P. Xavier, "Price discount schemes and inter-country comparisons of telecommunication tariffs," *Telecommun. Policy*, vol. 22, nos. 4–5, pp. 289–314, May 1998. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0308596198000111>
- [23] A. Li, X. Liao, Z. Gao, and Y. Yang, "Price discount strategy for WSP to promote hybrid access in femtocell networks," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2014, pp. 1–6.
- [24] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.
- [25] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3801–3813, Nov. 2014.
- [26] S.-E. Elayoubi, O. Ben Haddada, and B. Fourestie, "Performance evaluation of frequency planning schemes in OFDMA-based networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1623–1633, May 2008.
- [27] M. Qian, W. Hardjawana, Y. Li, B. Vucetic, X. Yang, and J. Shi, "Adaptive soft frequency reuse scheme for wireless cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 1, pp. 118–131, Jan. 2015.
- [28] P. Kela, H. Lundqvist, M. Costa, K. Leppanen, and R. Jantti, "Connectionless access for massive machine type communications in ultra-dense networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.



**XIAO FU** (Graduate Student Member, IEEE) received the B.E. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2017. He is currently pursuing the Ph.D. degree with the National Mobile Communications Research Laboratory, Southeast University, China. His research interests include 5th generation networks and wireless networks.



**HONGWEI HOU** received the B.E. degree in information engineering from Southeast University, Nanjing, China, in 2020, where he is currently pursuing the M.E. degree with the National Mobile Communications Research Laboratory. His research interests include wireless communications and software implementation.



**QINGGUO SHEN** (Member, IEEE) received the Ph.D. degree from the PLA University of Science and Technology, Nanjing, China, in 1994. From 2001 to 2017, he was a Professor with the College of Communications Engineering, PLA University of Science and Technology. Since 2018, he has been a Special-term Professor with the National Mobile Communications Research Laboratory, Southeast University, Nanjing. His research interests include next generation mobile networks and M2M.



**XIQI GAO** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997.

He joined the Department of Radio Engineering, Southeast University, in 1992. From 1999 to 2000, he was a Visiting Scholar with the Massachusetts Institute of Technology, Cambridge, MA, USA, and Boston University, Boston, MA, USA. Since 2001, he has been a Professor of information systems and communications. From 2007 to 2008, he visited the Darmstadt University of Technology, Darmstadt, Germany, as a Humboldt Scholar. His current research interests include broadband multicarrier communications, MIMO wireless communications, channel estimation and turbo equalization, and multirate signal processing for wireless communications. He received the Science and Technology Awards of the State Education Ministry of China in 1998, 2006, and 2009, the National Technological Invention Award of China in 2011, and the 2011 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communications theory. He has served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2007 to 2012 and the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2009 to 2013. He also serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS.



**WENJIN WANG** (Member, IEEE) received the Ph.D. degree in communication and information systems from Southeast University, Nanjing, China, in 2011. From 2010 to 2014, he was with the School of System Engineering, University of Reading, Reading, U.K. He is currently an Associate Professor with the National Mobile Communications Research Laboratory, Southeast University. His research interests include advanced signal processing for future wireless communications and satellite communications. He was a recipient of the Best Paper Award at IEEE WCSP 2009 and the first grade Technological Invention Award of the State Education Ministry of China in 2009.

...