Side Channel Attack-Aware Resource Allocation for URLLC and eMBB Slices in 5G RAN

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Side Channel Attack-Aware Resource Allocation for URLLC and eMBB Slices in 5G RAN

YAJIE LI[1], **YONGLI ZHAO**[1], **JUN LI**[2], **JIAWEI ZHANG**[1],
XIAOSONG YU[1], **AND JIE ZHANG**[1]
[1]State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunication, Beijing 100876, China
[2]Chalmers University of Technology, 41296 Gothenburg, Sweden

Corresponding author: Yongli Zhao (yonglizhao@bupt.edu.cn)

ABSTRACT Network slicing is a key enabling technology to realize the provisioning of customized services in 5G paradigm. Due to logical isolation instead of physical isolation, network slicing is facing a series of security issues. Side Channel Attack (SCA) is a typical attack for slices that share resources in the same hardware. Considering the risk of SCA among slices, this paper investigates how to effectively allocate heterogeneous resources for the slices under their different security requirements. Then, a SCA-aware Resource Allocation (SCA-RA) algorithm is proposed for Ultra-reliable and Low-latency Communications (URLLC) and Enhanced Mobile Broadband (eMBB) slices in 5G RAN. The objective is to maximize the number of slices accommodated in 5G RAN. With dynamic slice requests, simulation is conducted to evaluate the performance of the proposed algorithm in two different network scenarios. Simulation results indicate that compared with benchmark, SCA-RA algorithm can effectively reduce blocking probability of slice requests. In addition, the usage of IT and transport resources is also optimized.

INDEX TERMS Network slicing, resource allocation, side channel attack, 5G RAN.

I. INTRODUCTION

As the foundation of Networked Society, the objective of the fifth generation of mobile networks (5G) is to provide ultra-high data rate, ultra-low latency and nearly ubiquitous connectivity to users and machines. According to the definition in [1], the application scenarios of 5G are classified into three types, including Enhanced Mobile Broadband (eMBB), Ultra-reliable and Low-latency Communications (URLLC), and Massive Machine Type Communications (mMTC). Thus, network operators are expected to meet the diversified service requirements with the same substrate network for different application scenarios.

Network slicing is a promising solution to realize the provisioning of customized services in 5G paradigm [2], [3]. Different network slices can share the same substrate infrastructure in the manner of logical isolation. Each network slice consists of heterogenous resources, including radio, transport and IT resources (e.g., computing and storage). According to the definition in [4], an end to end network slice is from antennas side to 5G core and spans radio access network (RAN),

The associate editor coordinating the review of this manuscript and approving it for publication was Ting Yang.

transport network and core network. This paper focuses on RAN slicing, which is a primary part in 5G network slicing.

In 5G RAN, based on the flexible split of baseband processing functions [5]–[7], base band unit is divided into two parts, Distributed Unit (DU) and Central Unit (CU). DU is responsible for processing latency sensitive tasks while CU consists of latency tolerant functions. Network function virtualization (NFV) is considered as a key enable technology to realize the slicing of 5G RAN. With the leverage of NFV in 5G RAN, both CUs and DUs can be realized in the form of virtual network functions (VNFs) running on commercial servers [8], [9]. Thus, each server can accommodate multiple instances of virtual DU (vDU) or virtual CU (vCU). For each slice, the traffic from antennas side must go through vDU and vCU in sequence to perform baseband processing.

Due to logical isolation instead of physical isolation, network slicing is facing up with a lot of security issues, such as exhaustion of security resources, denial of service (DoS) [10]. This paper focuses on the issue of side channel attack (SCA) among slices [11], [12], which is a class of attack on implementations of cryptography. Supposing two slices respectively use two VNF instances (i.e., virtual machines (VMs)) in the same hardware, a malicious VM in one slice may extract

fine-grained information from a victim VM in the other slice. Specifically, if the two slices have very different security levels, SCA could create a significant benefit for the attacker. Therefore, a specific policy of resource allocation is desired to cope with the potential SCA among different slices in 5G RAN.

In this paper, a SCA-aware Resource Allocation (SCA-RA) algorithm is proposed for URLLC and eMBB slices in 5G RAN. With the constraint of avoiding SCA, the objective in this paper is to maximize the number of slices accommodated in 5G RAN by optimizing resource allocation for slices. The main contributions of this paper can be summarized as follows.

- To the best of our knowledge, this paper investigates the issue of SCA among network slices in 5G RAN for the first time. We define the scenario of network slicing in 5G RAN and describe the resource allocation of eMBB and URLLC slices. Meanwhile, this paper clarifies how resource allocation for slices is affected by the risk of SCA.
- Considering dynamic slice requests, this paper designs a heuristic algorithm to effectively allocate heterogeneous resources for URLLC and eMBB slices under different security requirements.
- The performance of the proposed algorithm is evaluated through simulation with two different cases. Simulation results indicate that compared with the benchmark, the proposed SCA-RA algorithm can effectively reduce blocking probability of slice requests, as well as the usage of servers and transponders.

The rest of the paper is organized as follows. Section II introduces the related work of network slicing. Section III defines network scenario and presents the problem addressed in this paper. Then, Section IV describes the proposed SCA-RA algorithm and analyzes its time complexity. Section V evaluates the performance of SCA-RA algorithm in two network scenarios. Finally, Section VI concludes this paper.

II. RELATED WORK
In this section, we introduce the related works about network slicing, including network slice isolation and security in network slicing.

A. NETWORK SLICE ISOLATION
Since slice isolation is an effective solution to guarantee slice security, many researches focus on how to realize or optimize the slice isolation. The authors in [13] outlined the recent trends and proposed a set of challenges to realize end-to-end user's security based on slices isolation. For the combination of private and public network infrastructure, several deployment models were introduced in [14] to provide strong network slice isolation for the third parties. However, they just showed the qualitative analysis of different deployment approaches. The authors in [15] investigated dynamic network slicing strategies with mixed traffics in virtualized RAN to improve resource utilization and QoS satisfaction. A connection admission control mechanism was used to achieve effective isolation while reaping the capacity benefits of dynamic network slicing [16]. A two-layer scheduler was proposed for an efficient and low complexity RAN slicing approach to achieve trade-offs between isolation and efficiency [17]. A demonstration was given in [18] to show how Flex Ethernet (FlexE) technology could be leveraged to guarantee hard isolation between slices.

B. SECURITY IN NETWORK SLICING
Due to sharing the same physical network, network slices face a lot of security issues. The authors in [19] proposed the network slice trust degree concept and established the trust degree calculation model for 5G network slices. The concept of Security Trust Zone (STZ) was introduced into network slicing architecture, and the performance and isolation capabilities of STZ approach were assessed in terms of detecting and mitigating simulated threats [20]. An application-aware framework was proposed in [21] to enable Security as a Service (SECaaS) within network slices using Software Defined Networking (SDN) and NFV technologies. An efficient and secure service-oriented authentication framework was proposed in [22] to support network slicing and fog computing for 5G-enabled IoT services. Two heterogeneous signcryption schemes were proposed for 5G network slicing to achieve mutual communications between different public key environments [23]. The authors in [24] abstracted security issues of network slicing into the security requirement and security level instead of delving into specific security issues. They formulated an Integer Linear Programming (ILP) model and designed a heuristic algorithm to realize efficient and secure 5G core network slice provisioning. The authors in [25] used slice isolation to mitigate the impact of DDoS attacks on slice authentication and the solution was evaluated by a combination of simulation and an experimental testbed. However, the solutions proposed in [24], [25] are only suitable for core network slicing.

To the best of our knowledge, resource allocation policy of RAN slicing in 5G is not available in the context of avoiding SCA. Since SCA brings constraint on resource allocation of slices, the flexibility of resource allocation will be degraded as well as resource utilization. Thus, resource allocation of slices needs to be addressed carefully under the constraint of security requirements to improve the resource utilization in 5G RAN.

III. NETWORK SCENARIO AND PROBLEM STATEMENT
A. DEFINITION OF NETWORK SCENARIO
As shown in Fig. 1(a), this paper assumes a Dense Wavelength Division Multiplexing (DWDM) centric transport network with a ring-type topology [26]. The Active Antenna Units (AAUs) are connected to Metro Nodes (MNs) via Access Edge (AE) nodes. In turn, MNs can reach the 5G
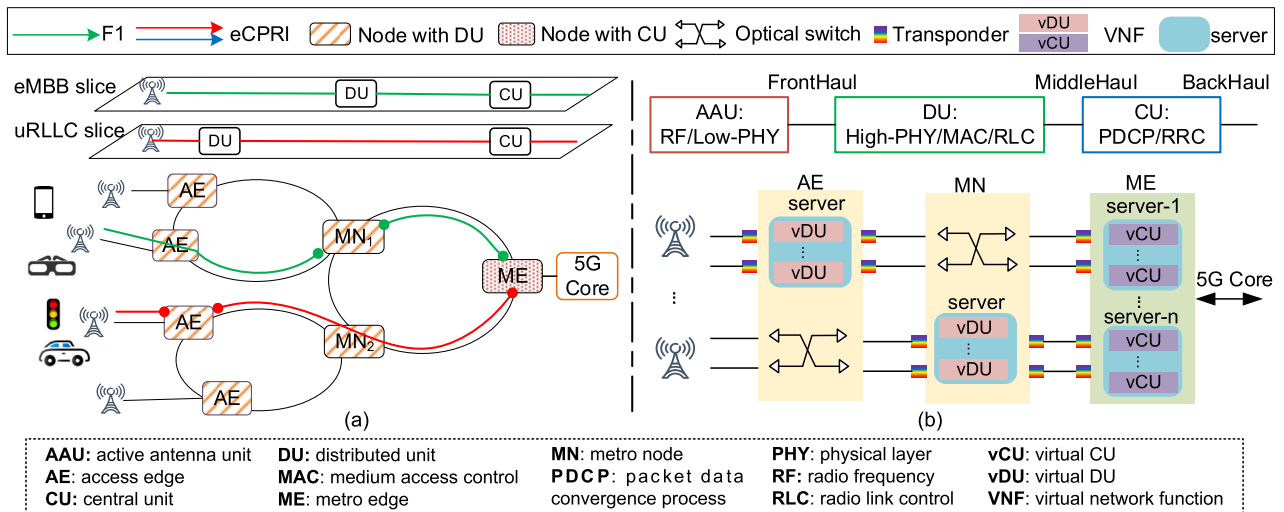
FIGURE 1. Illustration of network scenario in this paper.

Core via a Metro Edge (ME) node [27]. For the placement of radio functionalities, CUs are placed at the ME while DUs are deployed at both the MNs and AEs. In Fig. 1(b), the interface between CU and DU is called F1 [28]. The transport capacity requirement of F1 is similar to that for the conventional backhaul interface. The interface between AAU and DU is defined as evolved Common Public Radio Interface (eCPRI) [29], [30]. This paper adopts packet aggregation and wavelength switching for F1 and eCPRI traffic flows, respectively. Note that both DUs and CUs can be virtualized as vDUs and vCUs running on commercial servers. Each server can accommodate multiple instances of vCU or vDU.

With the logical isolation, each slice contains a set of heterogeneous resources from AAU to 5G core. Specifically, AAU/sub-AAU is allocated for each slice to collect the user signals and transport resources are assigned to carry eCPRI and F1 traffic. Meanwhile, IT resources (e.g., computing and storage) are required for each slice since the traffic must go through vDU and vCU in sequence to perform the baseband processing. As shown in Fig. 1(a), this paper considers two kinds of typical slices, i.e., eMBB and URLLC slice. Note that the Quality of Service (QoS) requirements of slices affect the location of the vDU to be used. For instance, the traffic from URLLC slice should be processed by the vDU hosted in AE to meet the low latency requirement. By contrast, since the number of vDUs in MN is much higher than that in AE, the flow in eMBB slice should be terminated in MN to enable the large amount of baseband processing from broadband services.

Due to the differences of service characteristics and application scenarios, each type of slices may have different security levels. This paper assumes that both URLLC and eMBB slices have two kinds of security levels, i.e., high and low security level. We consider dynamic traffic scenario where slice requests dynamically come and depart. Besides, the security level of each slice request is randomly assigned as

high or low security level. According to the security recommendations in [10], slices with different security levels cannot use the vDUs or vCUs in the same server to avoid the risk of SCA. In other words, each sever is supposed to provide vDUs or vCUs for the slices with the same security level.

For ease of description, the notations and corresponding definitions are listed as follows.

- $G(V, E)$, ring topology with DWDM transport for 5G RAN, where V refers to the set of nodes including AAUs, AEs, MNs and ME. E is the set of fiber links.
- $(r_{id}, r_t, \eta_{id}, b_{FH}, b_{F1}, d_{IT})$, a tuple denoting slice request, where r_{id} is the ID of slice request and r_t is the type of slice request. η_{id} is a binary variable, $\eta_{id} = 1$ for high security level. Otherwise, the value is 0. b_{FH} and b_{F1} are the bandwidth demand of eCPRI and F1 traffic, respectively. d_{IT} is the demand of IT resources.
- a_u^v, the vth slice of the uth AAU, $u \in [1, n_{aau}]$, $v \in [1, n_{slice}]$. n_{aau} is total number of AAUs and n_{slice} is slice number supported by each AAU.
- ψ_{AE_i}, set of AEs, including AE_i and its neighbor AEs with no more than one hop distance.
- ψ_{MN_i}, set of MNs, including MN_i and its neighbor MNs with no more than one hop distance.
- $\{s^i\}$, server list in MN, $i \in [1, m]$.
- $\{s^j\}$, server list in AE, $j \in [1, n]$.
- $s = \{vDU_\eta^k\}$, $k \in [1, \Gamma]$, a server consisting of several vDUs, Γ is total number of vDU.
- vDU_η^k, the kth vDU providing IT units for slices with security level η.
- χ, Capacity of IT units in each vDU.

B. PROBLEM STATEMENT

In this subsection, two examples are given to further explain how the resource allocation strategies are affected by
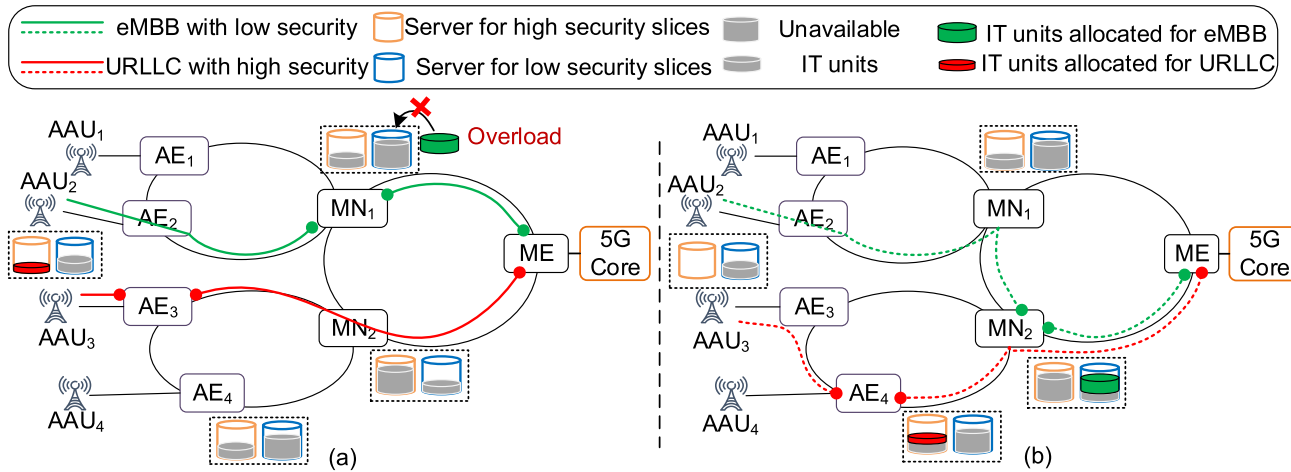
FIGURE 2. Two options of resource allocation for slices with different security levels.

special constraints (e.g., the selection requirement of vDU), which are shown in Fig. 2. We can see that both AEs and MNs have two servers, which provide vDUs for slices with high and low security level, respectively. The slice requests include one URLLC slice with high security level from AAU_3 to 5G core and one eMBB slice with low security level from AAU_2 to 5G core.

The policy adopted in Fig. 2(a) is to always use the nearest vDUs for each slice request. To be specific, the URLLC slice request will occupy the vDU in the server with high security level in AE_3. We can see that a free server provides the desired vDU for this slice. Similarly, the eMBB slice request attempts to use the vDU in the server with low security level in MN_1. However, the remaining resource in the selected server is not enough to meet the vDU requirement of this slice request. Thus, this eMBB slice request will be rejected due to lack of vDU. In Fig. 2(b), the selection of vDU for slice request is not limited to the nearest AE or MN. The URLLC slice uses the vDU in the server with high security level in AE_4. In this case, the server with high security in AE_3 remains free and the number of used servers reduces by one. For the eMBB slice, the vDU in the server with low security in MN_2 is allocated to avoid the overload in MN_1.

Comparing Fig. 2(a) with Fig. 2(b), we can find that the security requirements of slices introduce the constraint of avoiding SCA in the resource allocation for slices. In this regard, different policies of resource allocation for slices may potentially affect the resource usage and network performance (e.g., blocking probability of slice requests). To solve this issue, this paper investigates how to effectively allocate heterogeneous resources for the slices under their different security requirements. The objective is to maximize the number of slices accommodated in 5G RAN. In addition, it is important to note that since all vCUs are placed in ME, the allocation of IT resources focuses on vDUs assignment for slices.

IV. SCA-AWARE RESOURCE ALLOCATION ALGORITHM FOR URLLC AND eMBB SLICES

A. ALGORITHM DESCRIPTION

This paper designs a heuristic algorithm of resource allocation for URLLC and eMBB slices under the constraints of different security requirements. Considering each slice consists of radio, IT and transport resources, the aim of this work is to minimize the blocking probability of slice requests by optimizing the allocation of heterogenous resources. The pseudo-code of the algorithm is shown in Table 1.

Without loss of generality, we take an eMBB slice request as an example and its security level is high. For URLLC slice requests, the resource allocation can be realized similarly. The difference is that URLLC slices need to occupy IT resources (i.e., vDU) in AE instead of in MN. For each slice request, the proposed algorithm consists of three phases. The first phase is to allocate radio resource (i.e., AAU) for each slice request. Note that each AAU can be shared by several network slices. If a AAU is available for the new slice request, the algorithm steps into the second phase. Otherwise, the slice request will be blocked for lack of AAUs.

The second phase is to search a vDU to supply the required IT resource for the slice request. First, we will get the local MN MN_i, which the selected AAU in the first phase belongs to. Then, the set ψ_{MN_i} is composed of MN_i and its neighbor MNs with no more than one hop distance. It is important to note that MN_i is ranked as the first element in ψ_{MN_i} and its neighbor MNs are sorted in the ascending order of the connection length from AAU to 5G core. All the elements in ψ_{MN_i} are traversed to search a vDU for the high security slice request. Meanwhile, to avoid the risk of SCA, each server can only provide vDUs for either high security slices or low security slices. To be more specific, a server is assumed to provide vDUs for a set of slices, which is denoted as $R = \{r_{id}\}, id \in [1, N]$. Then, the following constraint must be

TABLE 1. SCA-aware resource allocation algorithm.

Input: network resource, eMBB slice request $(r_{id}, r_t, \eta, b_{FH}, b_{F1}, d_{IT})$

Output: resource allocation for slice request

1:   search available a_u^v for slice request
2:   **If** a_u^v cannot be found
3:     slice request is blocked due to the lack of AAUs
4:   **else**
5:     get the local MN MN_i to which a_u^v belongs
6:     calculate the set ψ_{MN_i} for MN_i
7:     **for** each MN in ψ_{MN_i}
8:       search vDU in the occupied servers to supply IT resource d_{IT} ;
9:       **if** vDU for high security slice can be found
10:         mark the ID of vDU;
11:         **break**;
12:       **else**
13:         **continue**;
14:     **end**
15:     **if** the demand d_{IT} cannot be satisfied,
16:       try to use vDU in new server in local MN
17:       **if** new server can be found
18:         mark the ID of vDU;
19:       **else**
20:         slice request is blocked for lack of IT resources
21:     **end**
22:     routing and wavelength assignment for eCPRI and F1 traffic
23:     **if** resource can be found for both eCPRI and F1 traffic
24:       update resource usage, including radio, IT and transport
25:     **else**
26:       slice request is blocked for lack of transport resources
27:   **end**

TABLE 2. Parameters configuration in simulation.

Network resource	Ring topology	ME number	1
		MN number	5
		AE number	75
		AAU number	450
	Max slice number supported by each AAU		3
	TP capacity for each wavelength (Gb/s)		25
	Number of servers	in each AE	2
		in each MN	6
	Number of vDUs in each server		10
	Capacity of each vDU (IT unit)		10
Service profile	Number of slice requests		50000
	URLLC	eCPRI (wavelength channel)	1
		F1(Gbit/s)	1
		vDU (IT unit)	5
	eMBB	eCPRI (wavelength channel)	2
		F1(Gbit/s)	5
		vDU (IT unit)	10

phase is expressed as $O(|A|)$, where $|A|$ is total number of AAUs. The worst case is to traverse all the AAUs to search an available AAU for a slice request. The second phase is the allocation of IT resources by assigning a vDU for the slice request. Considering the worst case where all the MNs in ψ_{MN} are traversed, the time complexity is represented as $O(|\psi_{MN}| * m * \Gamma)$, where $|\psi_{MN}|$ is the number of neighbor MNs, m is the number of servers in each MN and Γ is the number of vDUs in each server. The third phase is to perform routing and transport resource allocation and its time complexity can be denoted as $O((R_{AE} + R_{MN}) * \omega)$, where R_{MN} and R_{AE} are the number of MNs and AEs in each MR and AR, respectively. Besides, ω is the number of wavelengths provided by each fiber link. Thus, the time complexity of SCA-RA algorithm is $\max\{O(|A|), O(|\psi_{MN}| * m * \Gamma), O((R_{AE} + R_{MN}) * \omega)\}$.

V. SIMULATION AND RESULTS ANALYSIS
A. SIMULATION SETUP
The parameter configurations in the simulation are shown in the Table. 2. A DWDM centric network with a ring topology is considered in the simulation. There are 5 MNs and 1 ME. Each MN is connected to 15 AEs, each one with 6 AAUs. Thus, the ring topology consists of 450 AAUs in total. Note that each AAU can be shared by three slices at most. 25Gb/s transponder is adopted for each wavelength. For server deployment, each AE accommodates two servers while each MN consists of six servers. Note that due to the potential risk of SCA, each server cannot provide vDUs for high security slices and low security slices at the same time. Besides, each server consists of ten vDU instances, each of which can provide ten IT units for slices.

In terms of service profile, 50000 service requests are generated following Poisson distribution and their holding

satisfied to avoid the SCA among slices in this server.

$$\sum_{id=1}^{N} \eta_{id} = \begin{cases} N & only \text{ high security} \\ 0 & only \text{ low security} \end{cases}$$

Note that to reduce the number of working servers, only vDUs in the being occupied servers are checked and the vDUs in new servers are not considered in this step. If the vDU can be found in an occupied server, the ID of the vDU is marked. Otherwise, MN_i is checked whether a new server is available to provide the required vDU. If so, the ID of the vDU is marked. Otherwise, slice request is blocked for lack of IT resources.

The third phase is to perform the routing and transport resource assignment. For each slice request, a connection from AAU to 5G core is established, which must go through the selected vDU in the phase two. If transport resources are adequate for both eCPRI and F1 traffic, the slice request can be served successfully and the resource usage will be updated then. Otherwise, the slice request is blocked due to the scarcity of transport resources.

B. TIME COMPLEXITY ANALYSIS
The time complexity of SCA-RA algorithm is analyzed as follows. As mentioned in above section, SCA-RA algorithm consists of three phases. The time complexity of the first

time follows exponential distribution. The AAU in each slice request is randomly selected from all the AAUs. For URLLC and eMBB slices, the detailed capacity requirements of eCPRI and F1 interfaces are listed in Table 2. For each slice, eCPRI traffic occupies one or multiple wavelength by using wavelength switching while F1 traffic can share one wavelength channel through packet aggregation. The vDU of each eMBB slice and URLLC slice consumes ten and five IT units, respectively. In addition, routing and wavelength assignment are computed according to the strategy presented in [20]. Simulation results are obtained by averaging 20 experiments. All the presented results have a confidence interval not exceeding 6%, with 95% confidence level.

B. RESULTS ANALYSIS

The performance of the proposed algorithm is evaluated through simulation experiments. We focus on two types of performance metrics in this paper, i.e., blocking ratio of slice requests and network resource usage. The blocking ratio is the ratio of the blocked slice number to the total request number. This paper focuses on dynamic traffic, where slice requests dynamically arrive and depart. If the slice request cannot be served immediately at the arrival time, it will be blocked. Thus, blocking probability is calculated after all the slice requests arrive. Two kinds of network resources are considered in the simulation, i.e., IT and transport resources. The former is quantified as the total number of servers occupied by slices while the latter refers to the wavelengths and transponders (TPs) allocated for slices. Specifically, for wavelength usage, we calculate the number of wavelengths used on the most loaded link (MLL) in metro ring (MR) and access rings (AR), respectively. The TP usage refers to the total number of TPs used to aggregate F1 traffic flows. Since the usage of network resource varies with time during the simulation, we record the resource usage every 5000 slices. Then, with 50000 slice requests in total, each metric of resource usage is the average value of 10 samples in each simulation. In addition, the benchmark algorithm is to select the nearest AE or MN to provide the required vDUs for slice requests.

Two network scenarios are considered in the simulation. In the first scenario, the performance of SCA-RA algorithm is evaluated in different traffic loads. Besides, both AAUs and transport resources are enough for the slice requests. Thus, IT resource is the only bottleneck that may lead to service rejection in high traffic load. By contrast, the second scenario focuses on different ratios of eMBB to URLLC slice number among all slice requests with a fixed traffic load. Meanwhile, only AAUs are sufficient for slice requests. Both wavelength and IT resources affect how many slice requests can be accommodated in 5G RAN.

1) PERFORMANCE EVALUATION IN DIFFERENT TRAFFIC LOADS

In this case, traffic load increases from 700 to 1150 Erlangs with the increment of 50 Erlangs. Note that traffic load is calculated as the ratio of the mean holding time to the mean
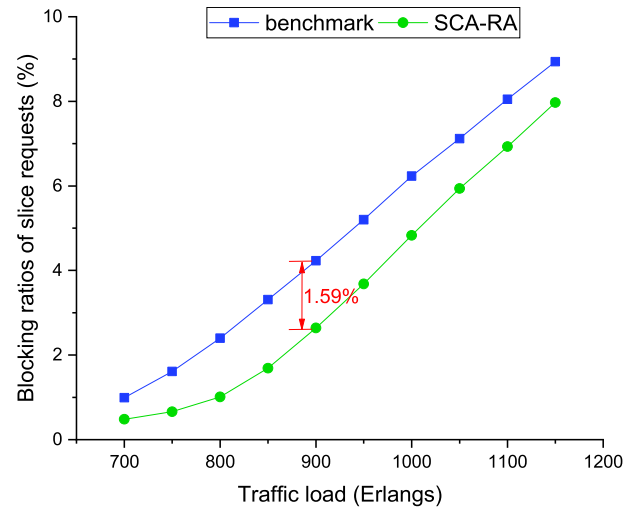

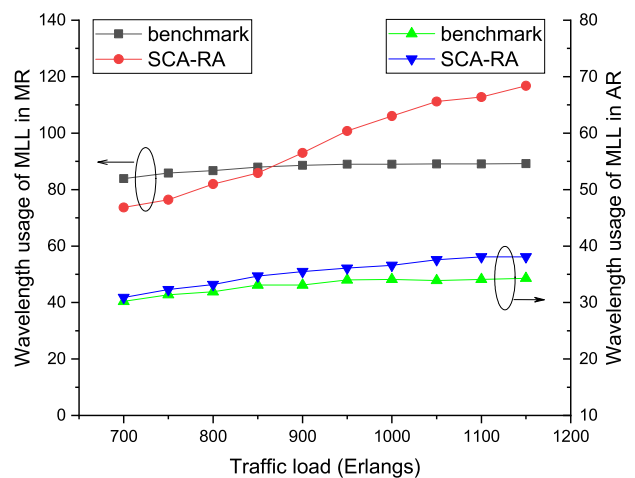**FIGURE 3.** Blocking probability in different traffic loads.

FIGURE 4. Wavelength usage in different traffic loads.

inter-arrival time of network slices. For the composition of slice requests, the ratio of eMBB to URLLC number is 1:2 [31]. Besides, each fiber link has enough wavelengths to carry eCPRI and F1 traffic.

Figure 3 shows the blocking probability of slice requests in different traffic loads. We can see that with traffic load increasing from 700 to 1150 Erlangs, a growing number of slice requests are blocked for lack of IT resources for both the benchmark and the proposed SCA-aware algorithm. Compared with benchmark, the proposed algorithm can make more slice requests served by the network in different traffic loads. For instance, if the traffic load is set to 900 Erlangs, nearly 4.23% of service requests are blocked with the benchmark while the blocking ratio can be reduced to 2.64% with the proposed SCA-RA algorithm. This is because multiple MNs or AEs are available to provide the required vDU for each slice request in SCA-RA algorithm. However, the benchmark always selects the nearest MN or AE, which makes slice requests blocked with a higher probability.

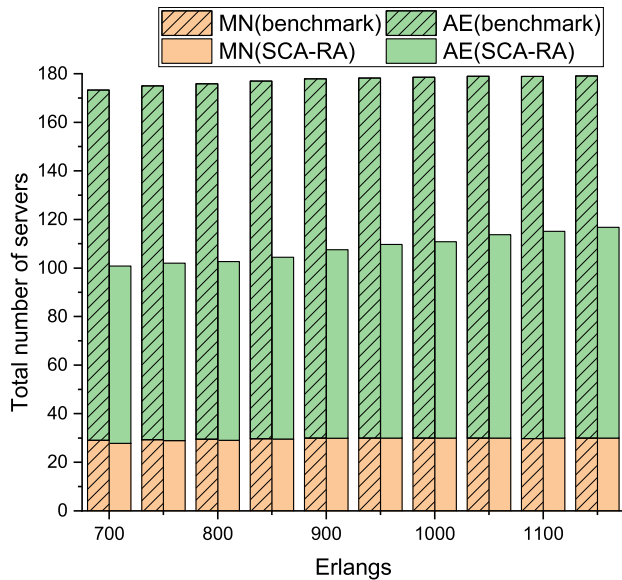Figure 4 shows the wavelength usage on MLL in MR and ARs. When traffic load increases from 700 to 1150 Erlangs,

the number of wavelengths used on MLL in MR increases from 84 to 90 with the benchmark. By contrast, with SCA-RA algorithm, the number of wavelengths used on MLL in MR increases from 74 to 117. Besides, it is interesting to note that two curves cross around 850 Erlangs. This can be explained as follows. The wavelengths in MR are occupied by both eMBB and URLLC slices. For eMBB slices, since the traffic may be processed by vDUs in neighbor MNs, SCA-RA algorithm requires more wavelengths in MR than the benchmark. For URLLC slices, since more F1 flows can be aggregated in SCA-RA algorithm, the wavelength usage in MR is less than the benchmark. When traffic load is less than 850 Erlangs, wavelength reduction in MR brought by URLLC slices surpasses the wavelength increment caused by eMBB slices. Thus, SCA-RA algorithm can achieve a smaller value for this metric than the benchmark. However, when traffic load is higher than 850 Erlangs, wavelength reduction in MR brought by URLLC slices is less than wavelength increment caused by eMBB slices. Therefore, SCA-RA algorithm requires more wavelengths in MR than the benchmark in high traffic load. Meanwhile, the wavelength usage on MLL in AR is shown in the lower half of Fig. 4. We can see that SCA-RA algorithm consumes more wavelengths than the benchmark on MLL in AR with different traffic loads, which can be analyzed similarly. Since the traffic in URLLC slices may be processed by vDUs in neighbor AEs, SCA-RA algorithm needs more wavelengths in AR than the benchmark. Besides, for eMBB slices, SCA-RA algorithm makes little difference on the wavelength usage in ARs. Therefore, SCA-RA algorithm requires more wavelengths in ARs than the benchmark.

The total number of servers and TPs are shown in Fig. 5 and Fig. 6, respectively. Since servers are deployed in both MNs and AEs, we calculate the number of servers used in MNs and AEs, respectively. In Fig. 5, we can see that the total number of servers used in MNs and AEs increases with the growth of
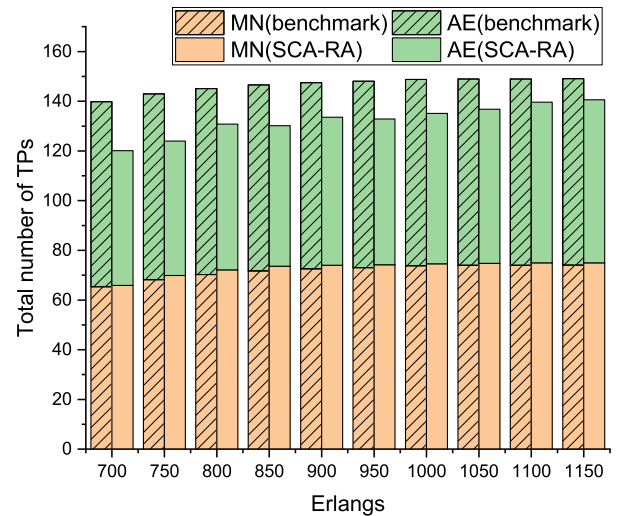
traffic load. The server usage in MNs with the benchmark is very close to the case with SCA-RA algorithm. With only five MNs in total, SCA-RA algorithm just brings slight reduction of server number in MNs. By contrast, the SCA-RA algorithm can bring significant reduction in terms of server usage in AEs. This is because 75 AEs in total leave a huge margin for the optimization of server usage. Compared with the benchmark, the total number of servers needed by SCA-RA algorithm is reduced by ~39%. For instance, when traffic load is 900 Erlangs, the SCA-RA algorithm and the benchmark consume 108 and 178 servers, respectively. Similarly, total number of TPs consists of two parts, TPs in MNs and AEs. For TP usage in MNs, SCA-RA algorithm requires slightly more TPs than the benchmark. Besides, we can see that SCA-RA algorithm makes a difference on the TP usage in AEs. Compared with the benchmark, SCA-RA algorithm brings 27% reduction of TP usage in AEs when traffic load is 700 Erlangs.

2) PERFORMANCE EVALUATION IN DIFFERENT PROFILES OF SLICE REQUESTS

In this case, the traffic load is fixed at 900 Erlangs and only 80 wavelengths are available for each fiber link. We consider five different ratios of eMBB to URLLC slice number among 50000 slice requests, i.e., eMBB: URLLC = 1:1, 1:1.25, 1:1.5, 1:2 and 1:3. The lower ratio corresponds to a higher number of URLLC slices generated among all slice requests.

Figure 7 shows the blocking ratio of slice requests in different ratios of eMBB to URLLC slices. Two curves are going down when the number of URLLC slices increases. This is because each URLLC slice requires less transport and IT resources than an eMBB slice. Besides, the proposed SCA-RA algorithm can make more slices accommodated than the benchmark. For instance, when the ratio is 1:1.5, the blocking ratio can be reduced by 1.08%. The number of wavelengths used on MLL in MR and ARs are listed in Fig. 8, respectively. For SCA-RA algorithm, the wavelength usage
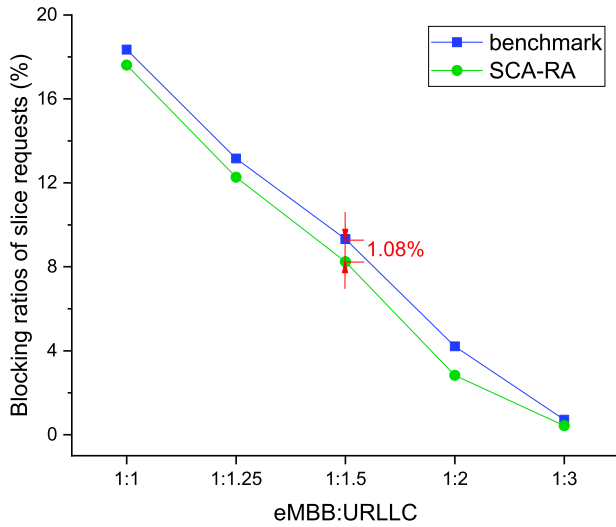
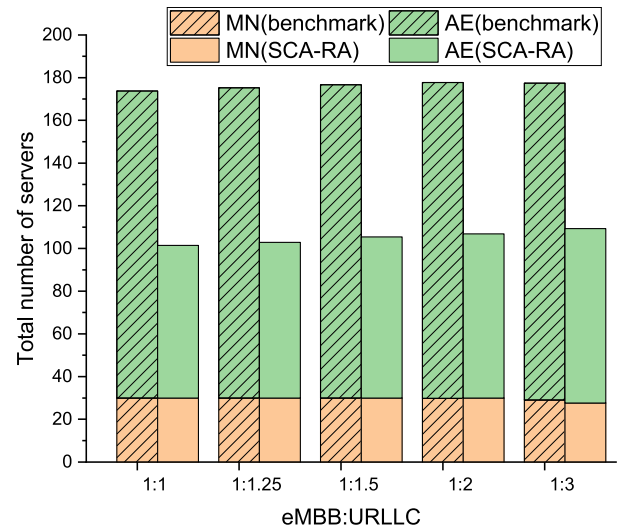**FIGURE 7.** Blocking probability in different ratios of eMBB to URLLC slices.
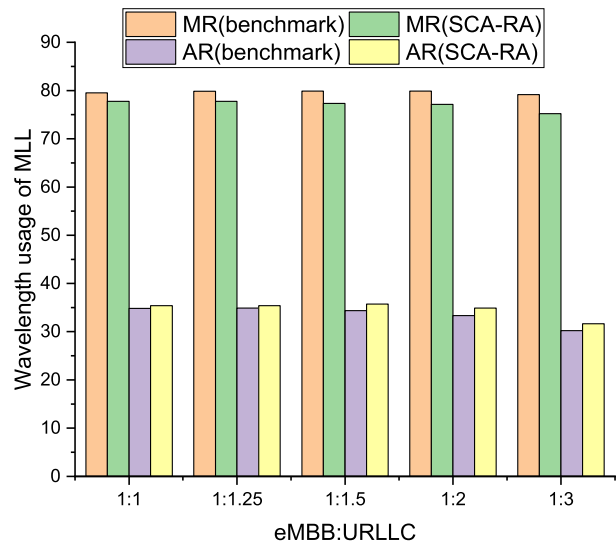

FIGURE 8. Wavelength usage in different ratios of eMBB to URLLC slices.

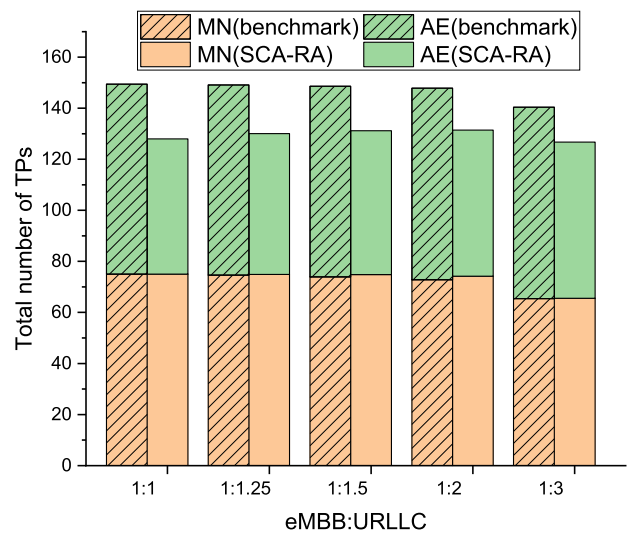**FIGURE 9.** Total number of servers in different ratios of eMBB to URLLC slices.

FIGURE 10. Total number of TPs in different ratios of eMBB to URLLC slices.

in ARs is slightly higher than the benchmark while the wavelength usage in MR is slightly less than the benchmark. This can be explained as follows. The wavelengths in ARs and MR are occupied by both eMBB and URLLC slices. Since the traffic in URLLC slices may be processed by vDUs in neighbor AEs, SCA-RA algorithm needs more wavelengths in AR than the benchmark. Meanwhile, for eMBB slices, SCA-RA algorithm makes little difference on the wavelength usage in ARs. Therefore, SCA-RA algorithm requires more wavelengths in ARs than the benchmark. The wavelength usage in MR can be analyzed similarly. For eMBB slices, SCA-RA algorithm requires more wavelengths in MR than the benchmark. For URLLC slices, since more F1 flows can be aggregated in SCA-RA algorithm, the wavelength usage in MR is less than the benchmark. With the increase of URLLC slice number, the wavelength reduction in MR brought by URLLC slices will surpass the wavelength increment caused

by eMBB slices. That is why SCA-RA algorithm can reduce the number of wavelengths used on MLL in MR.

Figure 9 and 10 show the usage of servers and TPs in different ratios of eMBB to URLLC slices. In Fig. 9, for a given ratio, we can see that SCA-RA algorithm hardly affects the server usage in MNs. In contrast, the server usage in AEs can be greatly reduced due to a large number of AEs. For instance, when the ratio is 1:1.5, the server usage in AEs can be reduced by ~49%. Therefore, SCA-RA algorithm can effectively reduce the server usage by optimizing resource allocation for slices. Besides, when the number of URLLC slices increases, more servers are required to provide IT resources for slices. The usage of TPs in MNs and AEs are shown in Fig. 10. Similar with the server usage in MNs, SCA-RA algorithm makes little difference on the TP usage in MNs. However, the TP usage in AE can be reduced by ~24% with SCA-RA algorithm. This is because SCA-RA algorithm

can aggregate F1 flows from local AE as well as neighbor AEs.

According to the above two cases, compared with the benchmark, the proposed SCA-RA algorithm can make more slices accommodated in 5G RAN by optimizing the allocation of heterogenous resources. Meanwhile, the proposed SCA-RA algorithm can bring savings in terms of server and TP usage, with the cost of slightly higher wavelength usage. It is worth noting that the server usage in AEs is significantly reduced by SCA-RA algorithm.

VI. CONCLUSION

With the constraint of avoiding SCA, this paper investigates how to optimize the resource allocation for eMBB and URLLC slices in 5G RAN. A SCA-RA heuristic algorithm is proposed to address this problem, with the objective of maximizing the number of slices accommodated in 5G RAN. Simulation is carried out to evaluate the performance of the proposed SCA-RA algorithm and two different network scenarios are considered. Simulation results show that the proposed SCA-RA algorithm can effectively reduce the blocking ratios of slice requests, compared with the benchmark. In addition, the usage of IT and transport resources is also reduced by optimizing the allocation of heterogeneous resources for slices. In future work, we plan to utilize artificial intelligence algorithms to improve the security of network slicing in 5G RAN [32].

REFERENCES

[1] *View on 5G Architecture*, 5G-PPP Architecture WG, 2016.
[2] *System Architecture for the 5G System Release-15, V.2.0.1*, document TS 23.501, 3GPP, 2017.
[3] *Description of Network Slicing Concept Version 1.0*, NGMN Alliance, 2016.
[4] Q. Li, G. Wu, A. Papathanassiou, and L. Wei, "End-to-end network slicing in 5G wireless communication systems," in *Proc. ESTI Workshop Future Radio Technol.*, 2016, pp. 27–28.
[5] *Study on New Radio Access Technology; Radio Access Architecture and Interfaces V2.0.0*, document 3GPP TR 38.801, 2017.
[6] D. Harutyunyan and R. Riggio, "Flex5G: Flexible functional split in 5G networks," *IEEE Trans. Netw. Service Manag.*, vol. 15, no. 3, pp. 961–975, Sep. 2018.
[7] J. Zhang, Y. Xiao, D. Song, L. Bai, and Y. Ji, "Joint wavelength, antenna, and radio resource block allocation for massive MIMO enabled beamforming in a TWDM-PON based fronthaul," *J. Lightw. Technol.*, vol. 37, no. 4, pp. 1396–1407, Feb. 15, 2019.
[8] K. Ramantas, A. Antonopoulos, E. Kartsakli, P. Mekikis, J. Vardakas, and C. Verikoukis, "A C-RAN based 5G platform with a fully virtualized, SDN controlled optical/wireless fronthaul," in *Proc. Int. Conf. Transp. Opt. Netw. (ICTON)*, 2018, pp. 1–4.
[9] *C-RAN: The Road Towards Green RAN-V2.5*, China Mobile, Hong Kong, 2011.
[10] *5G Security Recommendations, Package #2: Network Slicing v1.0*, NGMN Alliance, 2016.
[11] A. Oprea, A. Juels, Y. Zhang, and M. K. Reiter, "HomeAlone: Co-residency detection in the cloud via side-channel analysis," in *Proc. IEEE Symp. Secur. Privacy (SSP)*, May 2011, pp. 313–328.
[12] F. Liu, Y. Yarom, Q. Ge, G. Heiser, and R. B. Lee, "Last-level cache side-channel attacks are practical," in *Proc. IEEE Symp. Secur. Privacy (SSP)*, May 2015, pp. 605–622.
[13] Z. Kotulski, T. Nowak, M. Sepczuk, M. Tunia, R. Artych, K. Bocianiak, T. Osko, and J. Wary, "On end-to-end approach for slice isolation in 5G networks. fundamental challenges," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, 2017, pp. 783–792.

[14] P. Schneider, C. Mannweiler, and S. Kerboeuf, "Providing strong 5G mobile network slice isolation for highly sensitive third-party services," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.
[15] G. Sun, K. Xiong, G. O. Boateng, D. Ayepah-Mensah, G. Liu, and W. Jiang, "Autonomous resource provisioning and resource customization for mixed traffics in virtualized radio access network," *IEEE Syst. J.*, vol. 13, no. 3, pp. 2454–2465, Sep. 2019.
[16] X. Yang, Y. Liu, I. C. Wong, Y. Wang, and L. Cuthbert, "Effective isolation in dynamic network slicing," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Marrakesh, Morocco, Apr. 2019, pp. 1–6.
[17] D. Marabissi and R. Fantacci, "Highly flexible RAN slicing approach to manage isolation, priority, efficiency," *IEEE Access*, vol. 7, pp. 97130–97142, 2019.
[18] N. Huin, P. Medagliani, S. Martin, J. Leguay, L. Shi, and S. Cai, "Hard-isolation for network slicing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr./May 2019, pp. 955–956.
[19] B. Niu, W. You, H. Tang, and X. Wang, "5G network slice security trust degree calculation model," in *Proc. IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 1150–1157.
[20] D. Schinianakis, R. Trapero, D. S. Michalopoulos, and B. G.-N. Crespo, "Security considerations in 5G networks: A slice-aware trust zone approach," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–8.
[21] Y. Khettab, M. Bagaa, D. L. C. Dutra, T. Taleb, and N. Toumi, "Virtual security as a service for 5G verticals," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.
[22] J. Ni, X. Lin, and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 3, pp. 644–657, Mar. 2018.
[23] J. Liu, L. Zhang, R. Sun, X. Du, and M. Guizani, "Mutual heterogeneous signcryption schemes for 5G network slicings," *IEEE Access*, vol. 6, pp. 7854–7863, 2018.
[24] X. Li, C. Guo, L. Gupta, and R. Jain, "Efficient and secure 5G core network slice provisioning based on VIKOR approach," *IEEE Access*, vol. 7, pp. 150517–150529, 2019.
[25] D. Sattar and A. Matrawy, "Towards secure slicing: Using slice isolation to mitigate DDoS attacks on 5G core network slices," in *Proc. IEEE Conf. Commun. Netw. Secur. (CNS)*, Jun. 2019, pp. 82–90.
[26] M. R. Raza, M. Fiorani, A. Rostami, P. Öhlen, L. Wosinska, and P. Monti, "Demonstration of dynamic resource sharing benefits in an optical C-RAN," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 8, no. 8, pp. 621–632, Aug. 2016.
[27] Y. Li, J. Mårtensson, M. Fiorani, B. Skubic, Z. Ghebretensaé, Y. Zhao, J. Zhang, L. Wosinska, and P. Monti, "Flexible RAN: A radio access network concept with flexible functional splits and a programmable optical transport," in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, 2017, pp. 1–3.
[28] *F1 Application Protocol (F1AP)*, document NG-RAN TS 38.473, 3GPP, 2018.
[29] *Common Public Radio Interface: eCPRI Interface Specification, V2.0*, CPRI, Bengaluru, India, 2019.
[30] J. Zhang, Y. Ji, S. Jia, H. Li, X. Yu, and X. Wang, "Reconfigurable optical mobile fronthaul networks for coordinated multipoint transmission and reception in 5G," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 9, no. 6, pp. 489–497, Jun. 2017.
[31] T. Truong-Huu, P. M. Mohan, and M. Gurusamy, "Service chain embedding for diversified 5G slices with virtual network function sharing," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 826–829, Mar. 2019.
[32] Y. Zhao, B. Yan, Z. Li, W. Wang, Y. Wang, and J. Zhang, "Coordination between control layer AI and on-board AI in optical transport networks [invited]," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 12, no. 1, pp. A49–A57, Jan. 2020.

YAJIE LI received the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications (BUPT), in 2018. He was a Visiting Doctoral Student with the KTH Royal Institute of Technology, from October 2016 to December 2017. He is currently a Postdoctoral Researcher with BUPT. His research interests include software-defined optical networks, edge computing, and 5G optical transport networks.

YONGLI ZHAO received the B.S. degree in communication engineering and the Ph.D. degree in electromagnetic field and microwave technology from the Beijing University of Posts and Telecommunications (BUPT), in 2005 and 2010, respectively. He is currently a Professor with the Institute of Information Photonics and Optical Communications, BUPT. He has published over 150 journals and conference papers. His research interests include software-defined optical networks, flexi-grid optical networks, and network virtualization.

JUN LI received Ph.D. degree from the KTH Royal Institute of Technology, Sweden, in 2019. He was a Visiting Ph.D. Student with Princeton University, from September 2018 to March 2019. He is currently a Postdoctoral Researcher with the Chalmers University of Technology, Sweden. His research interests focus on fog/edge computing, optical network, and distributed machine learning.

JIAWEI ZHANG received the Ph.D. degree from the State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, China. He was a Joint-Supervised Ph.D. Student with the University of California at Davis, Davis, CA, USA. He is currently an Associate Professor with BUPT. His current research interests include 5G RAN transport networks, network function virtualization, software-defined radio, and optical access networks, with an emphasis on the various fronthaul-/midhaul-/backhaul-based optical communication technologies. He served on the Technical Program Committees for the IEEE DRCN 2019 and for the Workshop on Cloud Computing Systems, Networks, and Applications at the IEEE GLOBECOM 2014, ICC 2015, and INFOCOM 2018 conferences. He also severed as a Guest Editor of the special issue Resilience in future 5G Photonic Networks of *Photonic Network Communications Journal* (Springer).

XIAOSONG YU received the Ph.D. degree from the Beijing University of Posts and Telecommunications (BUPT), China, in 2015. He is currently an Assistant Professor with the Institute of Information Photonics and Optical Communications (IPOC), BUPT. His research interests include spatial division multiplexing-enabled elastic optical networks (SDM-EONs), software-defined optical networking (SDON), data centre networking, and optical network security.

JIE ZHANG received the bachelor's degree in communication engineering and the Ph.D. degree in electromagnetic field and microwave technology from the Beijing University of Posts and Telecommunications (BUPT), China. He is currently a Professor and the Dean of the Information Photonics and Optical Communications Institute, BUPT. He has published over 300 technical articles. He has authored eight books, submitted 17 ITU-T recommendation contributions, and six IETF drafts. He holds 17 patents. His research interests include architecture, protocols, and standards of optical transport networks. He has served as a TPC Member of a number of conferences, such as ACP, OECC, PS, ONDM, COIN, and ChinaCom.

● ● ●