Auction-based Network Slicing Architecture and Experimentation on SD-RANs

Qiaofeng Qin, Leandros Tassiulas
Department of Electrical Engineering and Institute for Network Science, Yale University, USA

ABSTRACT

The 5G communication era brings innovations to the Radio Access Networks (RANs) to support increasing traffic demands and a wide variety of network services, where the network slicing is a key technology. However, heterogeneous access technologies (e.g., LTE, WLAN and WiMAX) have their unique mechanism representing and allocating radio resources. In addition, with the coexistence of multiple RANs, network providers may compete with each other in offering accesses and resources. In order to converge to an optimal network slicing scheme under this circumstance, we propose an architecture based on Software Defined RANs (SD-RANs) and auction theories. Our proposed system supports modular deployment to heterogeneous RANs and is capable in achieving slicing through the negotiation with all network providers. The efficiency and scalability of the system are verified by a prototype that we implement with open-source SD-RAN platforms FlexRAN and 5G-EmPOWER.

CCS CONCEPTS

• **Networks** → **Network architectures**; **Network management**.

KEYWORDS

SD-RAN; Network Slicing

ACM Reference Format:

Qiaofeng Qin, Leandros Tassiulas. 2020. Auction-based Network Slicing Architecture and Experimentation on SD-RANs. In *Proceedings of ACM MobiSys (OpenWireless '20)*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3396865.3398690

1 INTRODUCTION

In the 5G era, wireless networks are facing with a tremendous increase of traffic demands. Meanwhile, they are expected to support service categories with heterogeneous requirement such as enhanced Mobile Broadband (eMBB), Ultra-Reliable Low Latency Communications (UR-LLC) and massive Machine Type Communications (mMTC), where the radio resources have to be managed and allocated in a more sophisticated way.

This publication was supported partly by the National Science Foundation in U.S. under Grant CNS 1815676, and the the Army Research Office under Agreement Number W911NF-18-10-378.

OpenWireless '20, June 19, 2020, Toronto, ON, Canada

ACM ISBN 978-1-4503-8011-9/20/06…$15.00
https://doi.org/10.1145/3396865.3398690

By separating the network control and data planes, Software Defined Networking (SDN) provides a centralized management and programmability to networks. It becomes a solution when applied to radio access networks (RANs). There have several Software Defined RAN (SD-RAN) projects such as FlexRAN [1] and 5G-EmPOWER [2] bringing flexibility to the radio resource allocation. Network slicing policies can be applied on these platforms dividing radio resources in separate groups, i.e., slices, to meet the requirement of different services.

However, the emergence of SD-RAN also brings challenges for the optimal allocation of radio resources. On the one hand, SD-RAN projects simplify the deployment of mobile networks. It is expected to be a common case that multiple SD-RANs owned by different providers may coexist to offer network accesses in a competitive manner. Moreover, most mobile devices nowadays such as laptops and smartphones can connect to multiple RANs at the same time. On the other hand, the state-of-the-art SD-RAN projects usually design their own protocols for the communication between control and data planes, making it difficult to apply a policy across different SD-RANs, especially when it involves heterogeneous radio access technologies (RATs), e.g., LTE and WLAN. Such phenomenon significantly increases the difficulty of developing a network slicing scheme over heterogeneous SD-RANs.

We solve the heterogeneous SD-RAN slicing problem by designing a novel architecture consisting of a centralized Slicing Orchestrator and an agent for each SD-RAN as a module running on it, as shown in Figure 1. Within this architecture, we propose a new network abstract facilitating the adoption of slicing policies to heterogeneous RANs in a unified way. The architecture is also capable to optimize the resource allocation with multi-party network providers by utilizing game theory and auction mechanisms.

We develop a prototype of proposed architecture on two open-source SD-RAN platforms FlexRAN and 5G-EmPOWER to verify the feasibility of our design. Evaluation results measured with real network devices and application scenarios demonstrate the low cost, high efficiency and scalability of the proposed architecture.

The rest of the paper is organized as follows. Section II summarizes the related work of SD-RAN projects and network slicing approaches. In Section III, we describe our system design realizing heterogeneous RAN slicing. We specify in Section IV how we develop a prototype of proposed architecture on several state-of-the-art SD-RAN platforms. The evaluation results of our system are presented in Section V. Finally, Section VI concludes the paper and lists some directions for future researches.

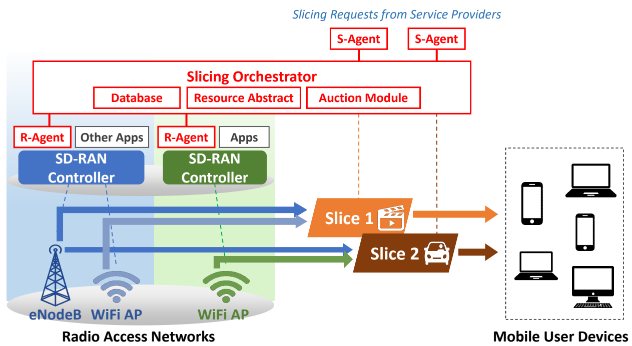


Figure 1: Network slicing across heterogeneous software-defined radio access networks.

2 BACKGROUND

2.1 SD-RAN Projects

Efforts have been made in introducing the concepts of SDN into different radio access technologies. FlexRAN [1] is a project separating the control and data planes of the open-source LTE eNodeB implementation OpenAirInterface [3]. It achieves network slicing by allocating resource blocks (RBs) among different slices. Orion [4] further improves this architecture by enabling the functional isolation. CAPWAP [5] is a protocol enabling managing multiple WLAN access points through a central controller. Odin [6] proposes light virtual access point (LVAP) abstraction facilitating the programmability of WLANs. Based on LVAP, 5G-EmPOWER [2] is another SD-RAN approach achieving WLAN slicing. It also provides SDN control to LTE eNodeBs and plans to develop 5G New Radio (5G NR) solutions.

The centralized management and programmability brought by these SD-RAN approaches make it possible to realize various applications including network slicing at RANs. However, each SD-RAN usually defines its own protocol, making it difficult to develop a scheme allocating network resources across heterogeneous SD-RANs, which is a challenge we aim to solve in this work.

2.2 RAN Slicing Approaches

There exist many works developing theories [7] and architectures [8, 9] for network slicing over heterogeneous RANs, e.g., cellular networks and WLANs. However, they usually assume all RANs are owned by the same provider and neglect the possible competitions between them. Though a few works like [10] address the competitions during slicing by applying the game theory, they do not contain evaluation results with a real networking testbed. Overall, only a small portion of slicing works such as [11] contain implementations and experiments with real devices. Moreover, these slicing architectures require deep modifications to existing network infrastructures. On the contrary, we develop and deploy a real system with cutting-edge SD-RAN technologies while taking these realistic issues into consideration.

3 SYSTEM DESIGN

3.1 Overview

In this section, we propose our system design for the heterogeneous RAN slicing. We expect to have an architecture satisfying the following requirements:

- **Support of various RATs and multi-connectivity**. Different RATs have diverse mechanisms handling connectivity and radio resource allocations. To realize slicing across multiple RANs, we must find a way to unify different standards. Besides, most of the modern mobile end devices (e.g., laptops and smartphones) have the capacity to connect to multiple RANs at the same time. This case must be considered when slicing the network.
- **Modularity**. We prefer to deploy our system as a module over SD-RANs rather than replace the existing control architecture. Such approach has several advantages. First, our system can be quickly adopted in an SD-RAN without modifications to the network infrastructure. Second, each SD-RAN usually has implemented some useful applications such as the access control and handover. Working as another module will make our system compatible with an SD-RAN's existing features.
- **Slicing as multi-party games**. Though the resources can be allocated efficiently by solving a global optimization problem, it is usually not realistic because both slices and RANs may owned by different providers. They keep their private information and pursue their own profits. Therefore, we do not assume that our system has a full control over every SD-RAN. Instead, we consider slicing in a game theory framework.

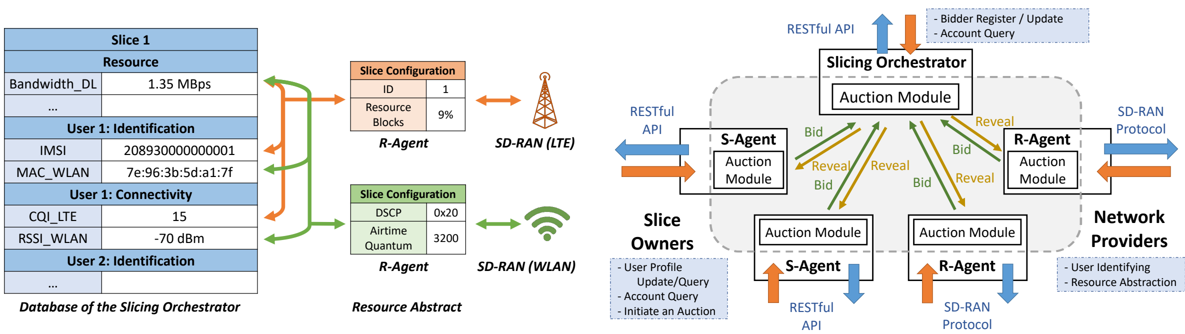
Addressing the issues above, we design a system containing following components, which are marked with red frames in Figure 1:

- **Slicing Orchestrator**. Following the manner of SDN, we deploy a centralized entity communicating with each RAN and managing slices.
- **Agent of each slice (S-Agent)**. We setup an agent for every network service provider in need of a slice. RESTful APIs are provided for them to request from RANs through the Slicing Orchestrator.
- **Agent of each RAN (R-Agent)**. Similar, an agent will be deployed at each SD-RAN as a module, following the protocols defined by that SD-RAN. It works as a bridge for the Slicing Orchestrator to query radio resources in the RAN. This agent has to be developed separately for each different SD-RAN control architecture. However, we will show in the implementation section that this module is lightweight and suitable for fast system deployment.

With these components, we are able to perform the key functions required by heterogeneous RAN slicing. In the next subsections, we will describe the way this system works.

3.2 Database and Abstraction

To apply slicing policies, the Slicing Orchestrator should first have a database storing necessary information, which is impacted by the diversity of different RAN protocols and standards:



(a) Database and resource abstract design. (b) Auction-based slicing mechanism.

Figure 2: The design of crucial components and protocols in the system.

- **User Identification**. A RAN identifies its user in its own way. For example, cellular networks will check a user's international mobile subscriber identity (IMSI), while a WLAN access point may manage the user access depending on the MAC address.
- **User Connectivity**. Rather than an on-off binary state, it is better to track the signal strength between each user and each RAN, making it possible to compensate users with low-quality links when making the slicing decisions. However, different indicators may be in use simultaneously, such as Channel Quality Indicator (CQI) and Received Signal Strength Indicator (RSSI).
- **Resources**. The same type of radio resource (e.g., bandwidth) can also be represented in different ways. For example, an LTE eNodeB adjusts a user's available bandwidth by allocating RBs, while a WLAN access point may divide the transmission airtime instead.

We develop a new resource abstract for the compatibility of heterogeneous RANs, as depicted in Figure 2(a). All types of user identification and connectivity are stored in key-value pairs in the Slicing Orchestrator, and an R-Agent will only get access to the necessary subset of them, depending on specific RAT standards. For radio resources, the database only maintains high-level representations (e.g., bandwidth and latency), and each R-Agent is equipped with an algorithms to convert them to the units adopted by RANs (e.g., RBs and airtime portions).

3.3 Auction-based Slicing Mechanism

In the multi-party setting, a slicing policy can only be determined if each slice owner's demand is satisfied by each RAN's offer. For example, considering a scenario with K RANs and M slices, we can set up a model mathematically assuming that each slice m requests resource $\mathbf{x_m} = (x_{m1}, ..., x_{mK})$ from each RAN which leads to a utility $U_m(\mathbf{x_m})$. On the other hand, if each RAN k offers resource $\mathbf{y_k} = (y_{k1}, ..., y_{kM})$ to slices, it will incur a cost $V_k(\mathbf{y_k})$. A slicing system may concern some forms of the social welfare, e.g., the gross utility of all participants $\sum_{m=1}^{M} U_m(\mathbf{x_m}) - \sum_{k=1}^{K} V_k(\mathbf{y_k})$. At the same time, a slicing scheme makes sense only if $x_{mk} = y_{km}$ is true for every pair of RAN/slice. However, if each slicing participant is selfish and aims at maximizing/minimizing its own utility/cost, it is unlikely to acquire either the optimal social welfare or the balance between resource offers and requests.

To solve this problem, the Slicing Orchestrator provides a platform for all participants to negotiate with each other. We achieve this by implementing an auction-based mechanism as depiected in Figure 2(b). An auction contains one or multiple rounds. In each round, participants submit bids on $\mathbf{x_m}$ and $\mathbf{y_k}$ (i.e., how much resource it is willing to request/offer) to the Slicing Orchestrator through S-Agents and R-Agents. Then, according to a specific set of payment rules, the orchestrator will announce a scheme that pays utility $h_k(\mathbf{y_k})$ to each RAN k based on how much resource it bids to offer, as well as charges each slice m utility $g_m(\mathbf{x_m})$ based on how much resource it bids to request. The procedure will repeat until no participant changes its bid. And the payment and allocation of the last round will be applied as the final slicing scheme.

During the auction process, the payment rules are crucial in guiding each participants to bid towards the maximization of the social welfare. There are multiple auction algorithms providing the payment rules in different scenarios. For example, if slices compete for a single type of resource with a fixed total amount (e.g., downlink bandwidth), we can define x_{mk} and y_{km} as non-negative real values and apply a double auction algorithm presented in [12]. It has been proved that the optimal social welfare can be achieved within a limited amount of bidding rounds under specific circumstances.

Under the same framework, it is also possible to adopt other auction algorithms in distinct scenarios, e.g., combinatorial auctions [13] for discrete resources of multiple types, where we can define each x_{mk} and y_{km} as a set containing finite elements.

We propose two protocols to implement the auction mechanism. A set of RESTful APIs is provided to human operators owning a slice/RAN to support functions such as registering a slice, initiating an auction on a specific resource or querying the account. Once the auction is initiated, the procedure will be finished by a second protocol between the Slicing Orchestrator and agents. For efficiency, it no longer requires manual operations except setting a high-level goal at the beginning. Instead, the auction is processed by a built-in optimizer of each agent (calculating the optimal bid) and the orchestrator (calculating payments and allocations).

Role	Hardware	Software
Slicing Orchestrator	Desktop Computer	-
WLAN Control Plane	HP ProLiant DL360 Server	EmPOWER Controller
WLAN Data Plane	TP-Link AC1750 Router	OpenWrt, EmPOWER Agent
LTE Control Plane/Core Network	HP ProLiant DL360 Server	OpenAirInterface EPC, FlexRAN Controller
LTE Data Plane	Desktop Computer, USRP B210	OpenAirInterface eNodeB
Service Provider	Desktop Computer	MPTCP, Apache HTTP Server, VLC Player
User Equipment	HP Omen Laptop, Huawei E3372 LTE USB Modem	MPTCP
User Equipment	Nexus 6P Smartphone	Ubuntu Chroot Environment

Table 1: Details of equipment and software deployed for experimentation.

4 IMPLEMENTATION

We develop a prototype of the proposed Slicing Orchestrator and S-Agent, as well as the R-Agent modules in two different SD-RAN controllers, FlexRAN managing LTE eNodeBs and 5G-EmPOWER managing WLAN access points.

We choose the down-link bandwidth as a representative radio resource to examine our slicing scheme. The SD-RANs have different inner logic to control the bandwidth of each user. FlexRAN assigns a user by mapping its IMSI to a slice ID. To control the bandwidth, it allocates a specific number of RBs, which is the resource unit of an LTE frame to that slice ID. On the contrary, 5G-EmPOWER adopts the OpenFlow [14] rules matching the MAC address of users and marks them in the Differentiated Service Code Point (DSCP) header. Then, the controller schedules the transmission airtime of access points in a weighted Round Robin manner, where weights are determined by the DSCP values. We develop R-Agent modules bridging the high-level bandwidth demands to these inner mechanisms. As a result, these parts only account for around 10% lines of codes compared with other components (i.e., database and auction modules). Therefore, it is possible to deploy our system in an existing SD-RAN even with a new RAT (e.g., 5G NR) with minor efforts and low costs.

We prepare a network testbed to deploy the prototype, which sets up LTE eNodeBs with USRP Software Defined Radio (SDR) devices, and WLAN access points with OpenWrt routers. The testbed also contains servers as the SDN controllers and the core networks. MPTCP is adopted enabling user devices' multi-connectivity to RANs. The details of the testbed are shown in Table 1. In the next section, we will demonstrate a realistic slicing scenario using this testbed.

5 EVALUATION

5.1 Scenario Setup

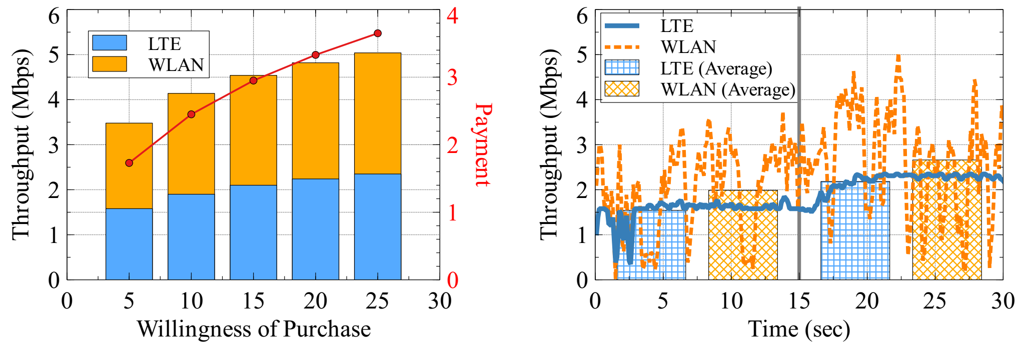
We set up one LTE and one WiFi access network in the testbed described in the last section. First, we start with a relatively simple scenario, having only one slice with HTTP video streaming service, where the user (a laptop) connects to the video server with MPTCP v0.93 protocol [15]. We apply slicing policies on the download bandwidth, which is determined with the double auction algorithm proposed in [12]. We assume the RAN's cost is a square function of the bandwidth consumed, and the slice's utility is an exponential function $U_1(\mathbf{x_1}) = w_1 \cdot \sum_{k=1}^{2}(1 - e^{-\alpha x_{1k}})$, where w_1 is a weight representing the willingness of purchasing bandwidth. This function is commonly used to capture the performance of network applications [16]. Then, we create another scenario with one additional slice, which is a web server for two mobile phones to download web pages. We define its utility $U_2(\mathbf{x_2})$ as the average delay of loading web pages, multiplied by another weight w_2.

5.2 Results

In the first scenario with only one slice, we focus on how the two SD-RANs cooperate to satisfy the bandwidth request. We apply slicing following the double auction algorithm with several different w_1 values, i.e., willingness of purchasing bandwidth. Figure 3(a) indicates that both SD-RANs offer bandwidth for a portion of requested bandwidth. The slice can determine its request by adjusting its willingness of purchase, resulting in corresponding different slicing policies achieved by the system. In all cases, the Slicing Orchstrator charges the slice an amount of utility which is the same with the payment to SD-RANs. This amount increases with the bandwidth requested.

The system should also be capable to react to network dynamics in real time. Therefore, we then measure the throughput of the user's both wireless interfaces for a 30-second video transmission. We assume that the user changes its willingness w_1 from 5 to 15 to require a larger bandwidth, therefore initiates a new auction at the 15-th second. Both the transient and average throughput during the two 15-second periods are recorded in Figure 3(b). The plot demonstrates that the Slicing Orchestrator succeeds in processing the request and allocates more bandwidth in the second period through both WLAN and LTE access networks. Moreover, the traces indicate that the new slicing policy is applied quickly to SD-RANs. For example, it takes less than 1 second to finish the auction and less than 5 seconds to achieve the new expected throughput with FlexRAN.

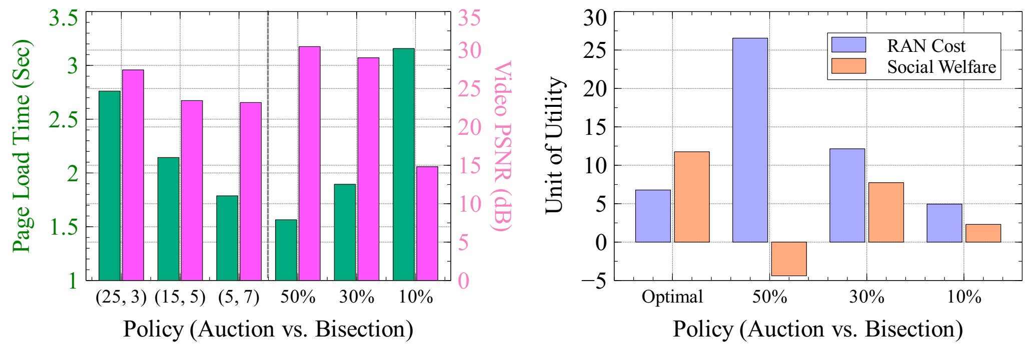
In the second scenario with two slices, we examine how to reach a balanced resource allocation between services that leads to the optimal social welfare. We run auctions with different combinations of (w_1, w_2) values in Figure 4(a). We measure the actual performance of two services, i.e., load time of web pages and peak signal-to-noise ratio (PSNR) of received videos. The tendency of service performance is consistent with the chosen utility functions. As a comparison, we implement a strawman slicing approach that simply allocates the same amount of bandwidth to both services (denoted as Bisection in the figures). The results demonstrate that it is harder to reach an efficient slicing scheme. For example, when increasing the allocated bandwidth from 30% to 50%, the performance has only limited improvement. However, as depicted in Figure 4(b),



(a) Different slicing polices.

(b) Dynamic slicing performance.

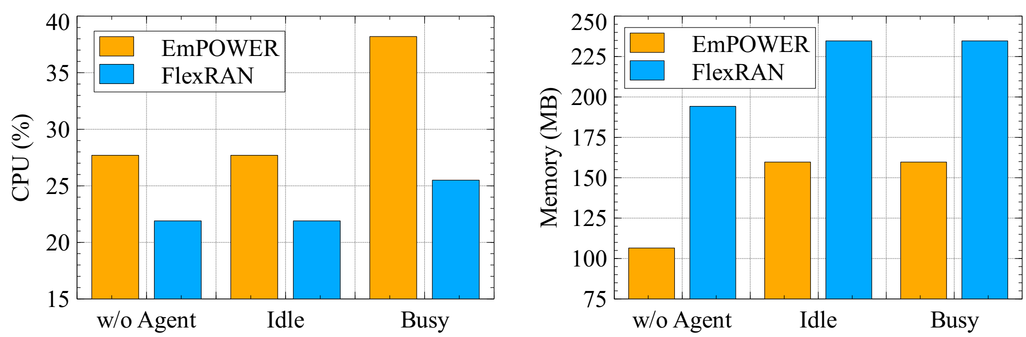
Figure 3: A slicing scenario with the video streaming service and two RANs.



(a) Different slicing polices.

(b) Social welfare improvement.

Figure 4: A slicing scenario with two services, video streaming and a web server.



(a) CPU utilization.

(b) Memory cost.

Figure 5: Evaluation results on the operating costs of proposed system.

it results in a significant higher cost to RANs and therefore makes the social welfare much worse than the optimal value acquired through the auction method.

We are also concerned about the operating costs incurred by our R-Agents to SD-RAN controllers. In Figure 5(a) and 5(b), we plot the CPU and memory consumption of both FlexRAN and 5G-EmPOWER controllers. The deployment of R-Agents (marked as Idle state) adds little burden to the CPU utilization and only costs a small extra portion of the memory. Then, we send new slicing requests to the orchestrator every second, emulating a high load to the system (marked as Busy state). This requires extra CPU resource, while the CPU is still far from being fully utilized. Therefore, our system is lightweight and scalable enough to handle a large number of slicing requests.

6 CONCLUSION

In this paper, we proposed an architecture realizing network slicing in heterogeneous SD-RANs in a modular manner. We deployed auction-based algorithms and protocols to properly handle the competition among different network and service providers. We implemented a prototype of our system on several state-of-the-art open-source SD-RAN platforms and evaluated its performance with real network devices and application scenarios. Our next step is to adopt similar solutions to more RATs, e.g., the 5G New Radio (5G NR) and low profile Internet of Things (IoT) protocols.

REFERENCES

[1] Xenofon Foukas, Navid Nikaein, Mohamed M Kassem, Mahesh K Marina, and Kimon Kontovasilis. Flexran: A flexible and programmable platform for software-defined radio access networks. In *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*, pages 427–441. ACM, 2016.

[2] Estefanía Coronado, Shah Nawaz Khan, and Roberto Riggio. 5g-empower: A software-defined networking platform for 5g radio access networks. *IEEE Transactions on Network and Service Management*, 16(2):715–728, 2019.

[3] Navid Nikaein, Mahesh K. Marina, Saravana Manickam, Alex Dawson, Raymond Knopp, and Christian Bonnet. Openairinterface: A flexible platform for 5g research. *SIGCOMM Comput. Commun. Rev.*, 44(5):33–38, October 2014.

[4] Xenofon Foukas, Mahesh K Marina, and Kimon Kontovasilis. Orion: Ran slicing for a flexible and cost-effective multi-service mobile network architecture. In *Proceedings of the 23rd annual international conference on mobile computing and networking*, pages 127–140. ACM, 2017.

[5] Massimo Bernaschi, Filippo Cacace, Giulio Iannello, Massimo Vellucci, and Luca Vollero. Opencapwap: An open source capwap implementation for the management and configuration of wifi hot-spots. *Computer Networks*, 53(2):217–230, 2009.

[6] Lalith Suresh, Julius Schulz-Zander, Ruben Merz, Anja Feldmann, and Teresa Vazao. Towards programmable enterprise wlans with odin. In *Proceedings of the first workshop on Hot topics in software defined networks*, pages 115–120. ACM, 2012.

[7] Saehoon Kang and Wonyong Yoon. Sdn-based resource allocation for heterogeneous lte and wlan multi-radio networks. *The Journal of Supercomputing*, 72(4):1342–1362, 2016.

[8] Tao Chen, Honggang Zhang, Xianfu Chen, and Olav Tirkkonen. Softmobile: Control evolution for future heterogeneous mobile networks. *IEEE Wireless Communications*, 21(6):70–78, 2014.

[9] Ali Saeed Dayem Alfoudi, Mohammed Dighriri, Gyu Myoung Lee, Rubem Pereira, and Fung Po Tso. Traffic management in lte-wifi slicing networks. In *2017 20th Conference on Innovations in Clouds, Internet and Networks (ICIN)*, pages 268–273. IEEE, 2017.

[10] Menglan Jiang, Massimo Condoluci, and Toktam Mahmoodi. Network slicing in 5g: An auction-based model. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.

[11] Amina Boubendir, Fabrice Guillemin, Christophe Le Toquin, Marie-Line Alberi-Morel, Frédéric Faucheux, Sylvaine Kerboeuf, Jean-Luc Lafragette, and Barbara Orlandi. Federation of cross-domain edge resources: a brokering architecture for network slicing. In *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, pages 415–423. IEEE, 2018.

[12] George Iosifidis, Lin Gao, Jianwei Huang, and Leandros Tassiulas. A double-auction mechanism for mobile data-offloading markets. *IEEE/ACM Transactions on Networking (TON)*, 23(5):1634–1647, 2015.

[13] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, New York, NY, USA, 2007.

[14] Nick McKeown, Tom Anderson, Hari Balakrishnan, Guru Parulkar, Larry Peterson, Jennifer Rexford, Scott Shenker, and Jonathan Turner. Openflow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 38(2):69–74, 2008.

[15] Damon Wischik, Costin Raiciu, Adam Greenhalgh, and Mark Handley. Design, implementation and evaluation of congestion control for multipath tcp. In *NSDI*, volume 11, pages 8–8, 2011.

[16] Veselin Rakocevic, John Griffiths, and Graham Cope. Performance analysis of bandwidth allocation schemes in multiservice ip networks using utility functions. In *Teletraffic Science and Engineering*, volume 4, pages 233–243. Elsevier, 2001.