Toward Operator-to-Waveform 5G Radio Access Network Slicing

Salvatore D'Oro, Francesco Restuccia, and Tommaso Melodia

The authors discuss a novel framework for operator-to-waveform 5G RAN slicing. In the proposed framework, slicing operations are treated holistically, including MNOs' selection of base stations and maximum number of users, down to the waveform-level scheduling of resource blocks.

ABSTRACT

RAN slicing refers to a vision where multiple MNOs are assigned virtual networks (i.e., slices) instantiated on top of the same physical network resources. Existing work in this area has addressed RAN slicing at different levels of network abstraction, but has often neglected the multitude of tightly intertwined inter-level operations involved in practical slicing systems. For this reason, this article discusses a novel framework for operator-to-waveform 5G RAN slicing. In the proposed framework, slicing operations are treated holistically, including MNOs' selection of base stations and maximum number of users, down to the waveform-level scheduling of resource blocks. Simulation results show that the proposed framework generates RAN slices where 95 percent of allocated resources can be used to perform coordination-based 5G transmission technologies, and facilitates the coexistence of multiple RAN slices providing up to 120 percent improvement in terms of SINR experienced by mobile users.

INTRODUCTION

Radio access network (RAN) slicing is expected to be a pivotal component of next-generation 5G networks [1] and the Internet of Things [2, 3]. By leveraging virtualization of physical network resources, infrastructure providers (IPs) assign mobile network operators (MNOs) one or more *slices* of the RAN, each representing a virtual network built on top of the underlying physical RAN. For each slice, the IP specifies the amount of network resources, including base stations (BSs), spectrum, and transmission power, among others, that can be used by MNOs to provide services to mobile users (MUs).

A key aspect of RAN slicing is that MNOs are assigned slices that are strictly independent of one another. In other words, the physical-layer allocation of radio resources (e.g., resource blocks) is completely up to the MNO, who can allocate them based on the proffered demand, pricing schedules, and quality of experience (QoE) levels. This provides MNOs with a great deal of flexibility, since they can leverage the IP's RAN infrastructure without the need to share business-specific information or low-level scheduling policies with the IP.

Thanks to these core features, RAN slicing has recently attracted considerable interest from academia and industry alike [4–7]. This is not at all surprising as RAN slicing provides a cost-effective and flexible solution to core challenges such as:

- The scarcity of networking resources (e.g., spectrum, antennas, and BSs)
- The need for cost-effective resource allocation strategies
- The ever-increasing demand for service differentiation through slices tailored to provide services with diverse QoE requirements; for example, video content delivery, web browsing, and real-time surveillance monitoring, among others

Indeed, MNOs can adapt their RAN slice requests to subscribers' requirements and traffic patterns in real time (e.g., by increasing the resource demand when MUs request high-data-rate services, or reducing it in small crowded areas), thus avoiding extra costs due to overbuying of network resources.

By allowing the coexistence of multiple virtual networks on top of the same physical resources, IPs can finally overcome the dreaded resource underutilization issue, which necessarily comes by using static and exclusive allocation policies, and has plagued previous network generations [8]. This latter aspect makes RAN slicing a beneficial technology for MNOs — who are now obliged to compete with each other to provide the best possible service — and a profitable business model for IPs [7]. On the other hand, IPs expand their business to the continuously growing market of flexible, high-performance, and on-demand network deployment for differentiated service typical of 5G networks.

Despite being one of the most promising 5G technologies, RAN slicing and its application to next-generation networks do not come without key challenges [5]. To be considered an integral part of future 5G networks, RAN slicing will have to be simultaneously *cost-effective*, *easy to use*, as well as providing *high-level network performance*. Poor performance, high prices, and a cumbersome slicing interface may indeed discourage potential MNOs from joining the RAN slicing market. Similarly, high maintenance and management costs may make RAN slicing an unprofitable business for IPs. Furthermore, to obtain the tens of gigabits per second [9] level data rates envisioned for 5G, deployed RAN slices will have to support advanced transmission technologies such as coordinated multipoint (CoMP), and multi-user and massive multiple-input multiple-output (MIMO) [6]. These technologies, however,

Digital Object Identifier:
10.1109/MCOM.001.1900316

The authors are with Northeastern University, Boston.

0163-6804/20/$25.00 © 2020 IEEE

have strict timing requirements (on the order of a few milliseconds), require in-depth understanding of low-level details of the physical network, and demand coordination among BSs, resulting in increased networking and operational costs.

From the above discussion, it is clear that the ability of IPs to efficiently trade off between business and networking aspects is a cornerstone to the success of RAN slicing. On the business side, this crucial trade-off can only be reached when IPs implement *holistic*, *operator-to-waveform* 5G slicing solutions, where MNOs express high-level networking need, and obtain and retain full control of the right amount of networking resources to achieve it through waveform-level operations, such as user scheduling and wireless transmissions.

Simply put, we envision a framework where an MNO can ask the IP something along the lines of "There is an event between 7:30 and 10:30 at TD Garden in Boston, where approximately 1000 spectators will be present. I expect each user to generate traffic of 10 MB/s." The framework will then *automatically* generate RAN slices and assign them with physical resources to achieve this goal, thus leaving to the MNO the task of handling the waveform-level scheduling operations.

Although existing work has already tackled slicing-related problems, to the best of our knowledge a unified operator-to-waveform 5G RAN slicing framework is still missing. Thus, in this work we discuss the road ahead to achieve this ambitious but crucial goal. Specifically, we propose a suitable architecture and discuss it in detail, and we identify crucial aspects to enforce isolation among different slices while simultaneously enabling advanced communication technologies (e.g., CoMP and MIMO) for 5G networks.

The remainder of this article is organized as follows. The following section gives an overview of the RAN slicing problem and its application to 5G networks. Then we discuss the architecture design of a unified RAN slicing ecosystem spanning across MNOs' and IPs' domains. The effectiveness of the proposed architecture is then assessed, and concluding remarks are given in the final section.

RAN Slicing: Requirements and Challenges

Figure 1 provides an overview of the main procedures involved in RAN slicing. In a nutshell, we can divide the core operations into three broad but logically distinct phases. First, each MNO declares the desired slice configuration (*slice request generation* phase), for example, how many resources are required, which BSs should be included in the slice, and the desired QoE level, among others. Then the IP decides which MNO requests can be accommodated and how many networking resources should be allocated to each slice (*network slice generation phase*). Virtual RAN slices are generated by allocating physical networking resources (e.g., resource blocks) to each MNO. Finally, each MNO takes control of the slice to schedule MUs' transmissions (*MUs scheduling and transmission phase*).

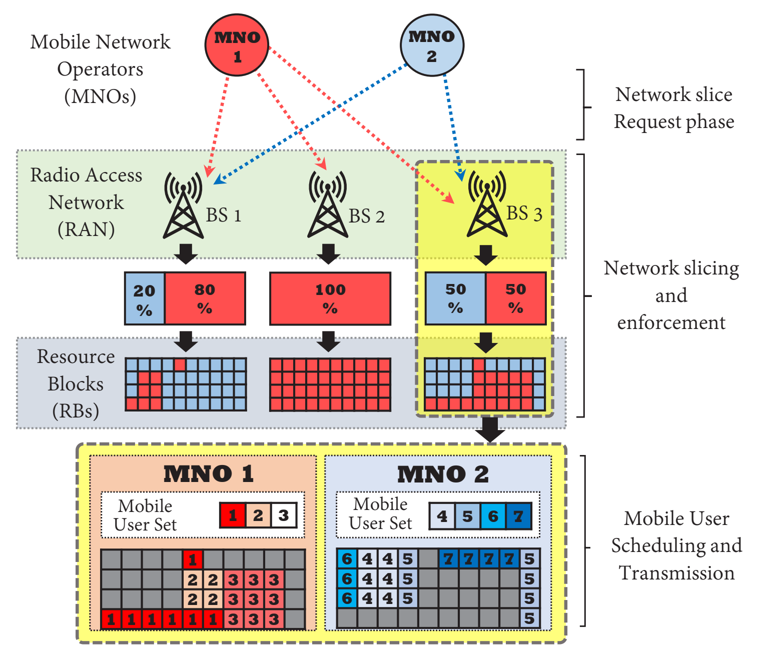

Figure 1. An overview of RAN slicing procedures with three BSs and two MNOs. We focus on BS 3 whose resources are sliced among MNO 1 and 2 serving three and four MUs, respectively.

The Need for a Unifying Slicing System: At first blush, RAN slicing might look like a minor variation of already well-established cloud-related slicing technologies [10, 11]. However, RAN slicing is an intrinsically different problem, since:

- Spectrum is a scarce resource for which overprovisioning is not possible.
- The network capacity is dynamic and heavily depends on rapidly fluctuating RAN-specific factors such as location of both MUs and BSs.
- Electromagnetic interference rapidly varies over time and prevents isolation across different slices.
- The agreements with MNOs usually impose stringent requirements on the subscriber QoE, which, however, strongly depends on channel conditions and MU mobility patterns.

Even though they refer to different components of the system, these aspects are tightly intertwined with each other, and to tackle them separately results in sub-optimal RAN slicing solutions. Thus, we need a radically novel approach that jointly tackles the above resource virtualization (high-level), interference management, and data transmission (low-level) aspects through a unified slicing framework.

Different Timescales: A crucial aspect is that the three phases in Fig. 1 work at different temporal scales. Indeed, as shown in Fig. 2, while the request and generation of virtual RAN slices take place within a *slicing window* whose overall duration fluctuates from a few tens of milliseconds to several months [12], MU scheduling and data transmission follow more minute dynamics. For example, scheduling and data transmission in 4G/5G networks is performed on time slots lasting 1 ms only — the duration of a transmission time interval (TTI). This is because scheduling requires network state information, which is available only at network runtime,, including MU position, channel state information (CSI), size of

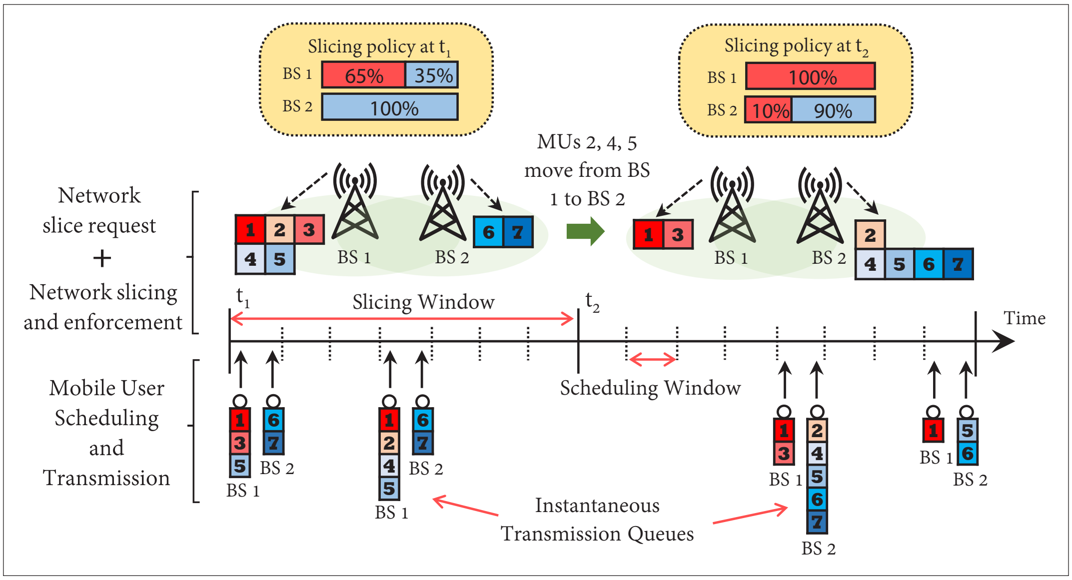

Figure 2. Timing aspects in RAN slicing and scheduling with two BSs and seven MUs. It is shown that MNOs adapt their slicing strategies to MUs' mobility patterns (long timescale), while transmission queues rapidly vary in time (short timescale).

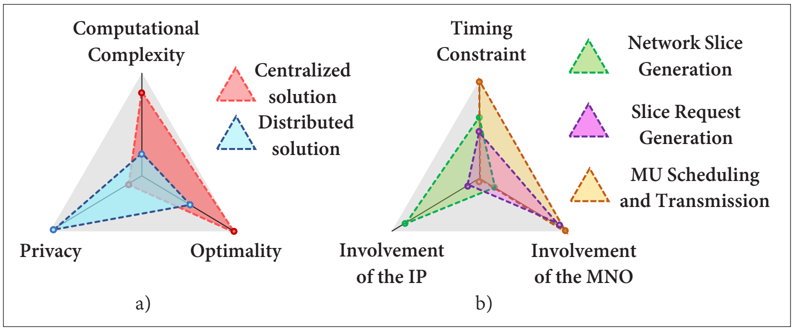

Figure 3. Triple constraint diagrams summarizing the RAN slicing optimization constraints: a) the complexity vs. optimality vs. privacy trade-offs; b) features and requirements for each phase of the RAN slicing problem.

scheduling queues, and so forth.

Enabling Emerging Multi-BS Wireless Technologies: The promise of high-speed 5G communications heavily relies on key multi-BS technologies including CoMP, massive MIMO, and beamforming [6], which ground their effectiveness on the tight cooperation and coordination among the different serving BSs. However, traditional slicing algorithms neglect to consider this core aspect, and thus do not fully embrace the potential of these technologies. Only recently [13] has it become evident that designing fine-grained RAN slicing algorithms that partition network resources to enable cooperation- and coordination-based communications will become more and more crucial in the years to come.

Complexity vs. Optimality vs. Privacy: A main challenge in designing an operator-to-waveform RAN slicing framework is the need to balance among *complexity*, *optimality*, and *privacy*, determine the role of IPs and MNOs, and regulate their interactions. These aspects can be represented with the two triple constraints in Fig. 3.

Figure 3a shows that centralized approaches produce optimal slicing policies to the detriment of privacy. Indeed, these strategies assume that the IP is aware of MNOs' relevant information, for example, subscribers' number and position, and bidding and business requirements. MNOs, however, are extremely reluctant to disclose such sensitive information. Also, centralized formulations of the RAN slicing problem are generally NP-hard [12, 13]. For this reason, approaches that divide the problem into multiple sub-problems with lower complexity, such as distributed solutions [14], are often considered. Although these approaches are usually sub-optimal, they compute solutions locally and rapidly, and therefore are more desirable in highly dynamic scenarios.

Addressing these triple constraints requires deep understanding of the dynamics governing each phase of the RAN slicing problem. As shown in Fig. 3b, the *slicing request generation* phase has mild timing constraints and mostly involves MNOs only. The *network slicing generation phase*, instead, requires direct control of the IP to efficiently apportion the available network resources while reducing intra-MNO interference and enabling coordination-based communications. Lastly, the *MU scheduling and transmission* phase is individually controlled by each MNO, does not involve the IP at all, but has very strict timing constraints.

The above discussion suggests that *the RAN slicing problem has diverse and generally opposing requirements and temporal constraints*. Recent research work has suggested letting MNOs handle the *slicing request generation* and *MU scheduling and transmission* phases, while leaving the network slicing generation phase to the IP [14]. Conversely, the *network slicing generation* phase is left to the IP, which possesses the global network view required to implement fine-grained RB allocation [13]. This approach offers a balanced tradeoff between centralization, complexity, optimality, and privacy, but also captures requirements and distinctive interactions among IPs and MNOs.

SYSTEM ARCHITECTURE

The proposed operator-to-waveform RAN slicing framework is illustrated in Fig. 4. It consists of a

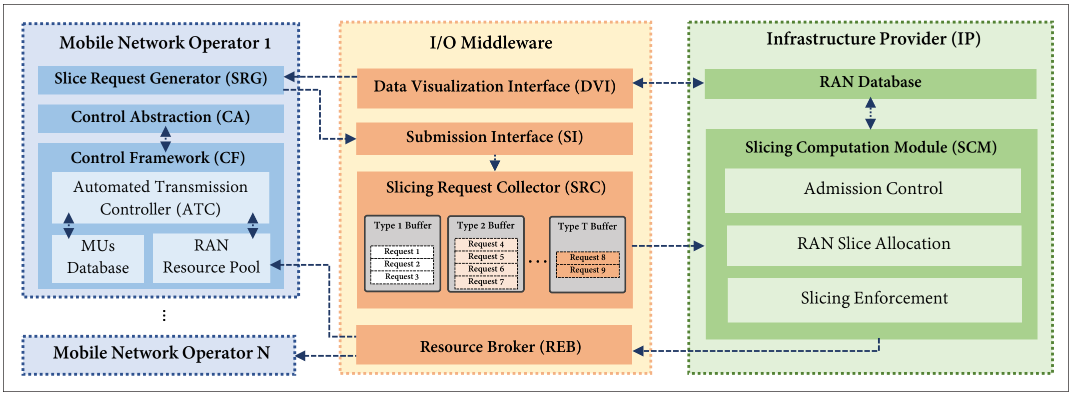

Figure 4. The architecture of the proposed three-tier RAN slicing framework.

The promise of high-speed 5G communications heavily relies on key multi-BS technologies including CoMP, massive MIMO, and beamforming [6], which ground their effectiveness on the tight cooperation and coordination among the different serving BSs. However, traditional slicing algorithms neglect to consider this core aspect, and thus do not fully embrace the potential of these technologies.

three-tier architecture. The *IP space* enables IPs to process MNOs' requests and generate RAN slices according to such requests. The MNO space concerns functionalities required by MNOs to visualize the available network resources, generate slice requests, and control the obtained RAN slices and network resources. Finally, the *I/O middleware* enables and regulates interactions among MNOs and between every MNO and the IP.

The I/O Middleware Tier

Database Visualization Interface (DVI): The DVI provides MNOs access to the list of available BSs, their location, leasing price, and coverage details, as well as useful metrics such as provided quality of service (QoS) (e.g., average QoS class identifier [QCI] and throughput) and the current congestion and utilization levels of each BS, among others. In some cases, IPs might in turn be MNOs, willing to increase profits by renting portions of their infrastructure to other MNOs. In such cases, they may be reluctant to share sensitive information such as the exact location of each BS. In this case, the DVI hides any sensitive information and only exposes more general coverage data. In line with 3rd Generation Partnership Project (3GPP) specifications, the DVI also provides a list of slicing templates to request predefined and feasible slice configurations.

Submission Interface (SI): The slicing requests generated by the *Slice Request Generator* block are submitted to the IP through the SI. The design of the SI can rely on web-based services [14] or client-based software with dedicated graphic user interfaces (GUIs) to facilitate the submission process [13].

Slicing Request Collector (SRC): This block collects MNOs' requests submitted through the SI. As discussed earlier, MNOs are expected to submit RAN slice requests that substantially differ from one another. For example, some MNOs might be willing to include a specific BS to their slice [14, 13]. Instead, other MNOs could be interested in covering a specific point of interest (PoI) (e.g., landmarks, schools, theaters, and other locations that might be of interest for MNOs) disregarding which BS is used to serve MUs [5]. Requests are organized into multiple classes, where each class contains homogeneous requests in terms of requested RAN configuration, QoS, or slice template.

Resource Broker (REB): This block is used by the IP to send slice admission/rejection notifications to each requesting MNO and, if accepted, to specify which BSs have been included in the current slice, the amount of networking resources that have been assigned to the MNO, and the price to be paid to rent the physical infrastructure. When a slice request is accepted, this information is stored inside the resource pool (RP).

The IP Tier

RAN Database: It contains detailed information on the network topology (e.g., position of each BS and its interfering BSs), leasing price, and the amount of available resources at each BS (e.g., number of antennas, operational frequencies, and RB availability). Furthermore, the database has information on coverage and performance properties of each BS (e.g., whether or not a BS covers a given PoI and what are the congestion and QCI levels at specific points of the network). The RAN database contains information that the IP uses for maintenance and monitoring purposes (e.g., how many resources are allocated to each MNO, MNO preferences toward a specific subset of BSs, and historical data on the RAN slice requests received in the past). Since the IP might be reluctant to share such abundant information with MNOs, they can visualize information contained in the RAN database through the DVI only. This way, the DVI obfuscates any IP-related sensitive information.

Slicing Computation Module (SCM): This block is responsible for computing optimal RAN slicing strategies based on MNOs' slice requests received by the Slicing Request Collector (SRC) module. This block also takes care of deciding which requests can be admitted, computing a slicing policy to allocate the available resources to the admitted MNOs, and enforcing the computed slicing policy on the underlying physical RAN. The IP specifies a desired objective function (e.g., interference minimization [13], slicing efficiency [12]) subject to one or more constraints (e.g., guaranteed service level agreements [SLAs] or minimum QoS level). The SCM computes a slicing strategy that meets an IP's directives via the following three procedures.

Admission Control (AC): This procedure determines which requests can be admitted. If an MNO submits an infeasible request (e.g., it demands an excessive minimum data rate guarantee for subscribers in a very congested area, or a large amount of resources on a BS already assigned to other MNOs), the AC procedure

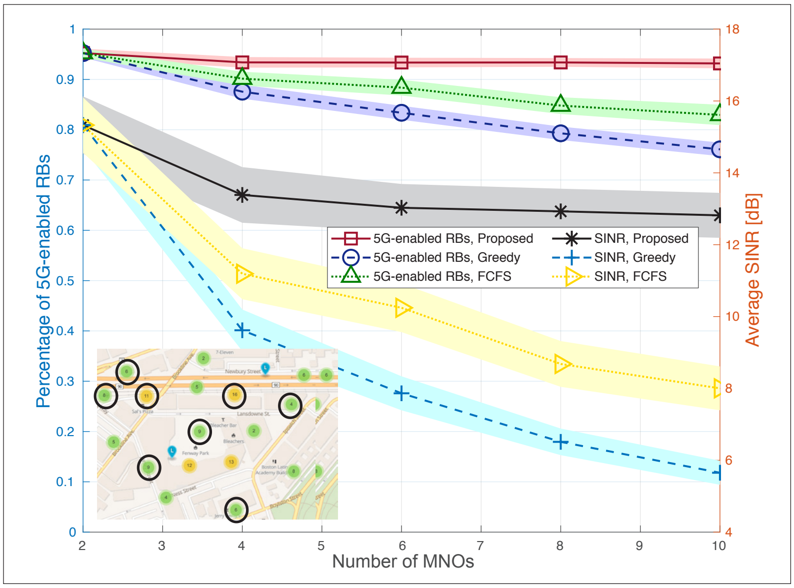

Figure 5. Simulated RAN deployment (BSs are highlighted with black circles)}, percentage of 5G-enabled RBs, and average SINR level under different RAN slicing strategies. Curves show average values, shaded areas display 95 percent confidence intervals over 200 simulation runs.

refuses the request and notifies the corresponding MNO through the REB;

RAN Slice Allocation (RSA): Admitted requests are assigned a portion of the available networking resources commensurate with the amount of resources or QoS level they requested. An example is illustrated in Fig. 1, where MNO 1 obtains 20, 100, and 50 percent of the available resources on BS 1, 2, and 3, respectively;

Slicing Enforcement (SE): This procedure completes the RAN slicing process. An example is depicted in Fig. 1, where we show how the slicing policy (50 percent, 50 percent) generated by the RSA procedure for BS 3 is then enforced over the RB grid by assigning specific RBs to the two requesting MNOs. This procedure is also aware of the network topology, specifically of adjacency (or interference) matrices [13]. These matrices provide information with respect to adjacency among close BSs, and are used to allocate the same RBs (in the time/frequency domain) to the same MNO when BSs are close enough to interfere with each other [13], thus enabling coordination-based communications such as MIMO, beamforming, and CoMP, and reducing inter-slice interference, which eventually results in isolation among slices.

We remark that the above three procedures are tightly intertwined, and their impact on network performance and IP profit is extremely significant. As such, these procedures must be tailored to admit as many slice requests as possible, improve network performance (e.g., maximize throughput and spectral efficiency, minimize interference), and enable advanced transmission technologies.

THE MNO TIER

Slice Request Generator (SRG): After having visualized information related to the underlying RAN via the DVI, MNOs formulate their request through the SRG by specifying, for example, which BSs should be included in the slice, which PoIs should be covered, the required number of resources, and minimum QoS levels, among others. The generated requests are then submitted through the SI. As remarked earlier, it is important that this process is autonomously executed by each MNO so that they can adapt their RAN slice to network changes and lower the computational burden on the IP. Moreover, by granting MNOs access to cumulative metrics such as overall network congestion, interference measurements, and total amount of resources allocated to the RAN slice, it is possible to achieve distribute RAN slicing [14]. This eliminates the need for any coordination mechanism among MNOs and preserves their privacy. In view of the above, the SRG gathers such cumulative information from the DVI and RAN RP, and formulates a new RAN slice request accordingly. Lastly, the generated request is submitted to the IP through the SI.

Control Abstraction (CA): The CA allows the MNO to define high-level control directives to optimize specific objective functions (e.g., transmission rate, latency, spectral efficiency) subject to one or more constraints (e.g., minimum data rate, maximum end-to-end delay and transmission power). This block operates as an abstraction layer hiding low-level network details — such as resources allocated to the RAN slice and MUs' position, number and generated traffic, among others — to the MNO.

Control Framework (CF): This block transforms MNOs' control directives defined in the CA (e.g., maximize spectral efficiency while guaranteeing a minimum data rate to each MU) into low-level resource allocation policies to serve MUs. The CF consists of the following elements.

MU Database: This database — akin to the home subscriber server (HSS) already used in current 4G cellular networks — contains information on MNOs' subscribers such as position, QoS requirements, and identities, and addressing. MNOs use it to determine which BSs should be included in their slice, and how many resources are needed to serve MUs and satisfy their QoS requirements.

RAN Resource Pool: As soon as the MNO is granted one or more RAN slices, this block collects information on networking resources assigned to each slice (e.g., operational frequencies, RBs, and antennas, among others). This way, MNOs are kept aware of which resources can be utilized to provide network services to mobile subscribers.

Automated Transmission Controller (ATC): This block provides the MNO with an abstraction layer that hides all low-level network details (e.g., MU position and CQI, employed modulation, resources allocated to the RAN slice, and power budget at each BS). This is achieved by converting high-level directives into optimization algorithms that optimally allocate RAN slice resources and meet MNOs' requirements and directives [15]. It follows that any MNO may efficiently and automatically control RAN slices without any in-depth knowledge of the underlying physical network as well as resource allocation and optimization algorithms.

MULTI-BS COORDINATION: A CASE STUDY

To show the effectiveness of our approach, we have integrated the components in Fig. 4 within a MATLAB simulator. We consider a cluster of eight BSs deployed in Boston, Massachusetts, extracted from the OpenCellID database (https://opencellid.org), as shown in Fig. 5.

We consider an LTE deployment where RAN slicing is performed at the RB level [13]. Each BS has 50 RBs, and MNOs distributively submit RAN slice requests through the SI, aiming at minimizing the network congestion and cost per slice [14]. Requests are collected by the SRC and converted into effective slicing strategies by the SCM, which slices network resources so that interference is minimized and multi-BS coordination is maximized [13]. MNOs whose slice requests have been accepted receive the list of RBs assigned to their slice, and maximize the slice throughput by scheduling MU downlink transmissions through the CF. MUs are uniformly distributed within the area shown in Fig. 5, and BSs transmit at the same output power level.

In Fig. 5 we show how RAN slicing policies computed by our framework not only enforce slice isolation, but also enable 5G-related transmission technologies such as CoMP and provide better signal-to-interference-plus-noise ratio (SINR) compared to other solutions. Specifically, Fig. 5 shows the percentage of RBs that can be leveraged for multi-BS wireless transmissions, the *5G-enabled RBs*, for different slicing enforcement strategies. We compare our proposed approach, which maximizes the probability that an RB can be used to serve MUs via two (or more) BSs in close proximity [13], to greedy (i.e., RBs are first allocated to those MNOs that have been assigned the largest amount of RBs) and first come first served (FCFS) (i.e., RBs are sequentially allocated to MNOs according to the temporal order of their RAN slice request submission).

We observe that the number of 5G-enabled RBs is almost constant under the proposed approach, but it decreases with the number of MNOs when using greedy and FCFS enforcement policies. To better understand the importance of such a result, in Fig. 5 we also show the average SINR of MUs. Since the proposed approach generates a high percentage of 5G-enabled RBs, it (i) reduces interference and (ii) enables CoMP transmissions, which both result in higher SINR values. On the contrary, greedy and FCFS produce fewer 5G-enabled RBs, causing the SINR to decrease when more MNOs submit slice requests, and thus more slices are instantiated on top of the same resources. Overall, the proposed framework generates up to 95 percent of 5G-enabled RBs, which eventually results in an SINR improvement up to 120 percent compared to other approaches.

CONCLUSIONS

We discuss a unified framework for operator-to-waveform 5G RAN slicing, which allows control from the MNO's selection of base stations and maximum number of users to the waveform-level scheduling of resource blocks. Simulation results show that our framework slices RAN physical resources so that 95 percent of available RBs can be used to perform coordination-based 5G transmission technologies (e.g., MIMO and CoMP), which allows for up to 120 percent improvement in terms of SINR experienced by mobile users.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant CNS-1618727 and the Office of Naval Research under grant N00014-17-1-2046.

REFERENCES

[1] P. Rost *et al.*, "Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks," *IEEE Commun. Mag.*, vol. 55, no. 5, May 2017, pp. 72–79.

[2] F. Restuccia, S. D'Oro, and T. Melodia, "Securing the Internet of Things in the Age of Machine Learning and Software-Defined Networking," *IEEE Internet of Things J.*, vol. 5, no. 6, 2018, pp. 4829–42.

[3] F. Restuccia and T. Melodia, "Big Data Goes Small: Real-Time Spectrum-Driven Embedded Wireless Networking Through Deep Learning in the RF Loop," *Proc. IEEE INFOCOM*, 2019.

[4] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From Network Sharing to Multi-tenancy: The 5G Network Slice Broker," *IEEE Commun. Mag.*, vol. 54, no. 7, July 2016, pp. 32–39.

[5] I. Afolabi *et al.*, "Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies, and Solutions," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 3, 2018, pp. 2429–53.

[6] B. Bangerter *et al.*, "Networks and Devices for the 5G Era," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 90–96.

[7] GSMA, "Network Slicing: Use Case Requirements"; https://tinyurl.com/y3bb3x4h, accessed Dec. 21, 2019.

[8] X. Zhou and L. Chen, "Demand Shaping in Cellular Networks," *IEEE Trans. Control of Network Systems*, vol. 6, no. 1, Mar. 2019, pp. 363–74.

[9] I. Parvez *et al.*, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Commun. Surveys & Tutorials*, vol. 20, no. 4, 2018, pp. 3098–3130.

[10] J. Ordonez-Lucena *et al.*, "Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, May 2017, pp. 80–87.

[11] A. Blenk *et al.*, "Survey on Network Virtualization Hypervisors for Software Defined Networking," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 1, 2016, pp. 655–85.

[12] C. Marquez *et al.*, "How Should I Slice My Network? A Multi-Service Empirical Evaluation of Resource Sharing Efficiency," *Proc. ACM MobiCom*, 2018, pp. 191–206.

[13] S. D'Oro *et al.*, "The Slice Is Served: Enforcing Radio Access Network Slicing in Virtualized 5G Systems," *Proc. IEEE INFOCOM*, Apr. 2019.

[14] S. D'Oro *et al.*, "Low- Complexity Distributed Radio Access Network Slicing: Algorithms and Experimental Results," *IEEE/ACM Trans. Networking*, vol. 26, no. 6, 2018, pp. 2815–28.

[15] Z. Guan *et al.*, "WNOS: An Optimization-Based Wireless Network Operating System," *Proc. ACM MobiHoc*, 2018, pp. 241–50.

BIOGRAPHIES

SALVATORE D'ORO [M'16] is an associate research scientist at the Institute for the Wireless Internet of Things, Northeastern University, Boston, Massachusetts. His research interests include game theory, optimization, and learning, and their applications to 5G systems and the Internet of Things.

FRANCESCO RESTUCCIA [M'16] is an associate research scientist at the Institute for the Wireless Internet of Things, Northeastern University. His research interests lie in the modeling, analysis, and experimental evaluation of wireless systems, with applications to pervasive and mobile computing and the Internet of Things.

TOMMASO MELODIA [F'18] received his Ph.D. degree from the Georgia Institute of Technology in 2007. He is the William Lincoln Smith Chair Professor and the director of the Institute for the Wireless Internet of Things at Northeastern University. He is the director of research for the PAWR Project Office. His research focuses on modeling, optimization, and experimental evaluation of wireless systems. He is an Associate Editor of *IEEE Transactions on Wireless Communications*, *IEEE Transactions on Mobile Computing*, and *IEEE Transactions on Biological, Molecular, and Multi-Scale Communications.*

Greedy and FCFS produce fewer 5G-enabled RBs, causing the SINR to decrease when more MNOs submit slice requests, and thus more slices are instantiated on top of the same resources. Overall, the proposed framework generates up to 95 percent of 5G-enabled RBs, which eventually results in an SINR improvement up to 120 percent compared to other approaches.