

Network Slicing in 5G Mobile Communication: Architecture, Profit Modeling, and Challenges

Mohammad Asif Habibi, Bin Han, and Hans D. Schotten

Abstract—Efficient flexibility and higher system scalability call for enhanced network performance, better energy consumption, lower infrastructure cost, and effective resource utilization. To accomplish this, an architectural optimization and reconstruction of existing cellular network is required. Network slicing is considered to be one of the key enablers and an architectural answer of communication system of 2020 and beyond. Traditional mobile operators provide all types of services to various kinds of customers through a single network, however, with the deployment of network slicing operators are now able to divide entire network into different slices each with its own configuration and specific Quality of Service (QoS) requirements. In a slice-based network, each slice will be considered as a separate logical network. In this way, the infrastructure utilization and resource allocation will be much more energy and cost efficient in comparison to traditional network. In this paper, we provided a comprehensive discussion on concept and system architecture of network slicing with particular focus on its business aspect and profit modeling. We thoroughly discussed two different dimensions of profit modeling, so called *Own-Slice Implementation* and *Resource Leasing for Outsourced Slices*. We further addressed open research directions and existing challenges with the purpose of motivating new advances and adding realistic solutions to this emerging technology.

I. INTRODUCTION

Fifth Generation (5G) communication system is expected to satisfy services, consumers, and business demands of beyond 2020. Besides supporting of massive number of user equipments, increased amount of network traffic, and enhanced QoS requirements of telephony/data services, 5G system provides support to a variety of vertical industries such as health-care, manufacturing, automotive, logistic, energy, environment, construction, and so on. These industries call for different use-cases and various QoS requirements, therefore, *one-size-fits-all* network architectural approach is no longer efficient for the utilization of 5G and beyond technologies. In order to efficiently accommodate vertical use-cases along with increased user demands over the same network infrastructure, 5G system requires architectural optimization and reconstruction with respect to current deployment. Network slicing can be one of the appropriate solutions to meet the above mentioned requirements, since it allows operator to partition network in a structured, elastic, scalable and automated manner.

Deployment of network slicing enables the operation of multiple logical networks over a single physical infrastruc-

ture in order to reduce total cost, decrease energy consumption, and simplify network functions in comparison to one network for different use-cases/business scenarios. In a slice-based network, each slice should be provided with its own specific characteristics and will be considered as a single logical network. In this way, the infrastructure utilization and resource allocation will be much more energy and cost efficient in comparison to traditional network.

Implementation of network slicing over 5G communication system arises many technical challenges, which are expected to be solved. There are also some business and economic issues (e.g. total cost, revenue, etc.) that need for significant optimization and re-designing in order to cope with new network architecture. On the other hand, the demand for broadband multimedia services has been increasing explosively, with this ongoing trend, the revenue of mobile network operators will soon be exceeded by Capital Expenditure (CAPEX) and Operating Expenditure (OPEX) required to operate the infrastructure. Therefore, total cost, expected revenue, and resource allocation in the context of network slicing are seemed to be very interesting research topics, which need for further discovery.

In this paper, we provide a comprehensive discussion on concept and system architecture of network slicing with particular focus on its business aspect and profit modeling. We thoroughly discuss two different dimensions of profit modeling i.e. *own-slice implementation* and *resource leasing for outsourced slices*. We further address challenges and open research directions in order to add new contributions and provide realistic solutions to existing problems in the field of network slicing implementation of 5G system. The reminder of this paper is structured as of following: Sec. II reviews state-of-the art in the field of network slicing. In Sec. III, we discuss concept and system architecture of network slicing for 5G mobile communication system. Sec. IV presents business aspect and two dimensions of profit modeling. Following that, we provide future research directions in Sec. V. Finally, a summary of this work is given in Sec. VI.

II. RELATED WORK

The concept of network slicing has exclusively been studied in literature. For instance, authors in [1] explained a detailed end-to-end framework of network slicing implementation in 5G communication system. The paper deals with deployment of vertical and horizontal slicing over the air-interface, Radio Access Network (RAN), and Core Network (CN). It further focuses on how to horizontally

slice both computation and communication resources to/from virtual computation platforms in order to improve scalability, enhance device capability, and increase end user experience.

Moreover, [2], [3], and [4] focused on deployment of network slicing over the RAN architecture of mobile networks. Authors in [2] analyzed network slicing in multi-cell RAN in order to support radio resource splitting among various slices. This paper further proposed four types of RAN slicing approaches along with their detailed comparison. However, authors in [3] explained how network slicing may impact various aspects of design and functions of RAN architecture of 5G mobile networks. The paper thoroughly covered RAN requirements for network slicing implementation. In [4], authors provided a comprehensive discussion on deployment of network slicing in heterogeneous Cloud-RAN (C-RAN) in order to improve throughput through computation and communication resource sharing.

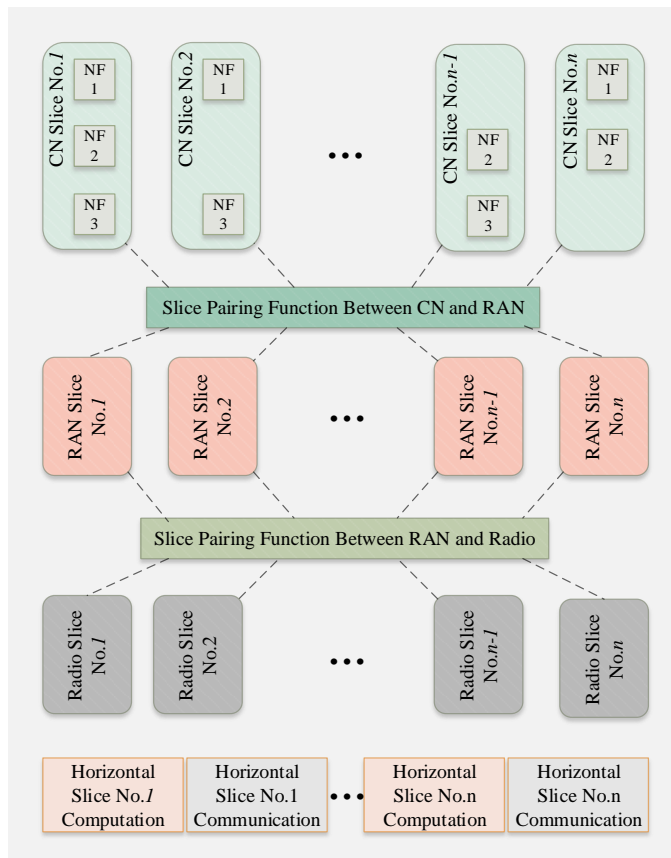


Fig. 1. Network Slicing System Architecture

Further, authors in [5] addressed network slicing related concepts i.e. resource allocation, virtualization technologies, orchestration process, and isolation function. The paper provided a comprehensive discussion on Software Defined Networking (SDN) and Network Function Virtualization (NFV) along with a deployment use-case (which considers network slicing using both NFV and SDN integration). The authors further demonstrated existing challenges and future research directions of network slicing implementation in 5G

communication system. Moreover, a comprehensive survey on architecture and further research directions of network slicing is also available in [6].

The Next Generation Mobile Networks (NGMN) introduced Network Virtualization Substrate (NVS), which allows infrastructure provider to control resource allocation of each of the virtual instance of an enhanced Node B (eNB) before customization of scheduling of each virtual operator within allocated resource. On the other hand, a heuristic-based admission control mechanism is introduced in [7], which dynamically allocates network resource to various slices in order to increase end user satisfaction considering specific requirements of each of the slice.

One of the main objectives of network slicing is to optimize profit modeling of traditional telecommunication networks. In order to increase overall revenue and decrease total network expenditure, a comprehensive study of business and economic dimensions of network slicing is required. Authors in [8] analyzed profit generated by different slices over the same network infrastructure and further modeled network resource management. Meanwhile, [9] dealt with the designing of an algorithm that allocates requests of network slices, which further maximizes total revenue of network infrastructure provider. However, in our paper, we first provide a comprehensive discussion on concept and system architecture of network slicing with particular focus on its business aspect and profit modeling. We further discuss two different dimensions of profit modeling i.e. own-slice implementation and resource leasing for outsourced slices in details. Moreover, we address open research directions with the purpose of motivating new advances and adding realistic solutions to this emerging technology.

III. NETWORK SLICING: CONCEPT AND SYSTEM ARCHITECTURE

Network slicing, in its simplest description, is to use virtualization technology i.e. NFV or SDN in order to design, partition, organize and optimize communication and computation resources of a physical infrastructure into multi logical networks for the sake of enabling of variety of services [1]. With deployment of network slicing, a single physical network infrastructure is sliced/partitioned into multiple virtual networks, which is called *Network Slice*. Each slice can have its own architecture, applications, packet and signal processing capacity, and is responsible for provisioning of specific applications and services to specific end users. Examples of network slices can be: a slice to serve remote control function of a factory, a slice serving for a utility company, a slice dedicated to provide emergency health services, and so on. A slice is consisted of Virtual Network Functions (VNFs), which are appropriately composed to support and build up services that are supposed to be delivered to the end users.

Network slicing deployment includes two main phases: creation and runtime [5]. In the slice creation phase, end user requests a slice from a network slice catalog, the tenant provides the slice immediately upon request. In the runtime phase, different functional blocks within each slice, which

are already created are now operating and providing service according to the end user's request. Each network resource i.e. NFV and functional block within a specific slice should have its own security mechanisms and must ensure operation within expected parameters in order to prevent access to unauthorized entities. This will lead to guarantee that faults or attacks occurring in one slice are confined to that given slice and will not propagate across slice boundaries. Slicing helps operator to provide new services and applications only by deployment of a slice instead of rolling out a new network, which leads to decrease CAPEX and saves time.

Network slices are operating on a partially shared infrastructure. This infrastructure is consisted of dedicated hardware i.e. network elements in the RAN and shared hardware i.e. Network Functions Virtualization Infrastructure (NFVI) resources. Network functions running on shared resources are usually instantiated in a customized manner for each slice, however, this approach cannot be applied to the network functions relying on dedicated hardware. Therefore, designing and identification of common functions is one of the key research directions in network slicing.

There are two different concepts and scenarios of using of network slicing in communication networks [10], slicing for the purpose of QoS and Slicing for the purpose of infrastructure sharing. Both dimensions of network slicing are described as of following:

- **Slicing for QoS:** The basic idea is to create various slices in order to offer different types of services to the end users, and to assure specific types of QoS requirements within specific slice. An example of this type of slicing can be a slice, which is created to provide service to a specific group of devices considering specific QoS requirements e.g. live video streaming, broadband connection to medical emergency response operation, and so on.
- **Slicing for Infrastructure Sharing:** The fundamental idea of this scenario of network slicing is to virtualize RAN domain of a wireless network, and further share it among various operators. There is a slice owner and a slice tenant. The owner gives the slice to a tenant based on an agreement. The tenant has overall control on both functions and infrastructure of that slice. This concept of network slicing leads to optimize network cost model for increasing the overall revenue, and meanwhile providing network scalability.

The purpose of network slicing in 5G mobile communication is to allow operators in order to share infrastructure among each other in flexible and dynamic manner, and to manage resource efficiently considering increased number of devices and massive amount of user traffic. However, a detailed discussion on objectives and motivation of network slicing implementation can be found in [10]. Network slicing helps mobile operators to simplify creation, configuration, and operation of network services. In order to efficiently allocate network resources, two-tier priorities are introduced in [7]. The first tier is *Inter-slice Priority*, which refers to

different priorities between various slices of a network. The priority of each of the slice is defined between owner and tenant of the slice. The second tier is *Intra-slice Priority*, which is referred to the priorities between different users of a single slice. These priorities are defined between users and service provider.

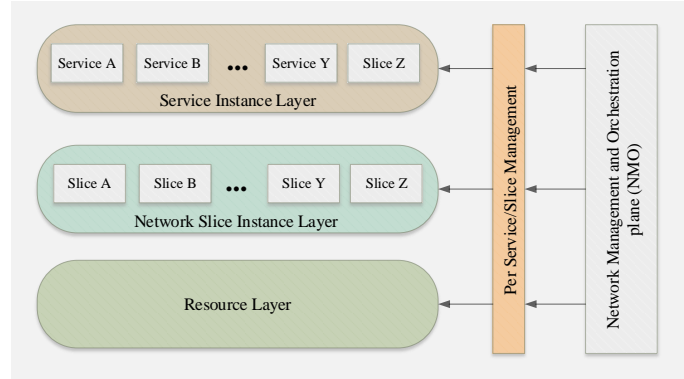


Fig. 2. Network Slicing Management Architecture

Slicing can be deployed in two dimensions over 5G communication networks, *Vertical Slicing* and *Horizontal Slicing*. Vertical slicing enables vertical industries and services, and focuses on CN. However, horizontal slicing improves system performance and increase end user experience, and mainly deals with RAN architecture. We define in details what is understood by both types of slicing as of following [1]:

- **Vertical Slicing:** The development and deployment of vertical slicing has already started in late Fourth Generation (4G) and early 5G, and is mostly focusing on the core domain of mobile networks. Mobile broadband networks are sliced vertically in order to serve vertical industries and applications in a more cost efficient manner. It segregates traffic of vertical industries from the rest of general broadband services of mobile network, which leads to simplify traditional QoS engineering problems.
- **Horizontal Slicing:** Increased number of user equipments and massive amount of traffic generated at the edge of mobile network expand network slicing from core domain to the RAN and air interface, which is called horizontal slicing. It is designed to accommodate new trends for scaling of system capacity, enabling of cloud computing, and offloading of computation devices at the edge of mobile networks. Horizontal slicing enables resource sharing among nodes and devices of a network. For example high capable network devices/nodes share their resources such as communication, computation, and storage with low capable network devices/nodes, which leads to enhance overall network performance.

Both vertical and horizontal slicing are independent from each other. End-to-end traffic flow in a vertical slice is transited between CN and user devices. While in a horizontal slice, it is usually transited locally between two ends of a

slice e.g. between a portable device and a wearable device [1]. In vertical slicing, each of the nodes of a network deploys similar functions among slices, however, in a horizontal slice new functions could be added and created at a network node.

Fig. 1 shows the concept and system architecture of network slicing. The architecture consists of CN slices, RAN slices, and radio slices. Each slice in CN is built from a set of Network Functions (NFs), some NFs can be used across multiple slices while some are tailored to a specific slice. There are at least two slice pairing functions, which connect all of these slices together. The first pairing function is between CN slices and RAN slices, and the second pairing function is between RAN slices and radio slices. The pairing function routes communication between radio slice and its appropriate CN slice in order to provide specific services and applications. The pairing function between RAN and CN slices can be static or semi-dynamic configuration in order to achieve required network function and communication. The mapping among radio, RAN and CN slices can be 1:1:1 or 1:M:N, it specifically means that a radio could use multiple RAN slices, and a RAN slice could connect to multiple CN slices.

End-to-end slicing architecture shown in Fig. 1 represents logical decomposition of network slicing, and takes specific network domain functions i.e. CN and radio network domains into account. From the operational perspective, NGMN defines that network slicing concept is consisted of three layers: *Service Instance Layer*, *Network Slice Instance Layer*, and *Resource layer*. Each of these three layers are described below and shown in Fig. 2. [11].

- **Service Instance Layer:** It represents end user and business services, which are expected to be supported by the network. Each service is represented by a *Service Instance*. These services can either be provided by the network operator or by a third party.
- **Network Slice Instance Layer:** A network operator uses a *Network Slice Blueprint* in order to create a *Network Slice Instance*. The network slice instance provides the network characteristics required by a service instance. The network slice instance may be shared across multiple service instance, which are provided by a network operator. The network slice instance can be consisted of none, one or more *Sub-network Instances*, which may be shared by another network slice instance. *Sub-network Blueprint* is used to create a sub-network instance to form a set of network functions, which runs on the physical/logical resources.
- **Resource Layer:** The actual physical and virtual network functions are used to implement a slice instance. At this layer, network slice management function is performed by the resource orchestrator, which is composed of NFV Orchestrator (NFVO), and of application resource configurators.

Network Management and Orchestration (NMO) plane shown in Fig. 2 provides orchestration and management functions of above mentioned three layers. NMO functions

need to allow for the orchestration and management in a per-slice level.

IV. NETWORK SLICING FOR PROFIT MODELING

Besides the improved scalability and flexibility provided by network slicing, it also worths to discuss business dimension of applying of network slicing from operator's perspective. In traditional networks, the cost of an operator in terms of CAPEX and OPEX is high in reference to the total generated revenue. There are several reasons for such higher costs, where the most important one is the network resource underutilization.

In traditional network architectures, the network operator provides an unspecified network resource pool for general utilization of all applications. However, depending on the performance requirement, various applications can have highly specified characteristics of resource utilization. Therefore, under traditional architectures, it is often to reserve excessive amount of certain network resources for some use-case with only slight demands on them. A lot of network resources are thus wasted, leading to a low network utilization rate.

In contrast, with deployment of network slicing, operators are able to efficiently analyze operational cost and revenue generated from each individual slice in the entire network. Based on revenue analysis, operators can allocate specified network resource bundles to different slices, which makes the resource management much more structured, flexible, and efficient. It is concluded that the very same network can be utilized to seamlessly provide more and better services i.e. generate more revenue without any increase in CAPEX. Therefore, network provider needs new algorithms in order to cope with new architecture and maximize revenue. To accomplish this, a deep review of the telecommunication regulatory framework has to be made. Innovative ways of pricing, new grounds for cost sharing and standardized solutions, which provide the required support for interoperability in multi-vendor and multi-technology environments, must be studied as well. Moreover, inter-operator network sharing and cooperative slicing concepts can be efficiently implemented by optimizing the network cost model for increasing the overall revenue, and meanwhile providing network scalability.

Network cost models in traditional networks are typically built based on CAPEX and OPEX, which are estimated according to the number of base stations (BSs), the transmission power and the traffic volume. For sliced networks, this methodology is not appropriate any more. As each resource can be shared by several network slices, and the slicing scheme varies from one resource to another. Hence, OPEX cannot be generally estimated for the entire physical network. A novel slice-oriented cost model is therefore needed. As we already discussed, every slice is specified for a particular pre-defined service, which covers a group of use cases with similar demands for the QoS. Therefore, a slice can be identically defined by a set of KPI requirements.

As 5G networks are supposed to support network slices implemented by both traditional mobile network operators

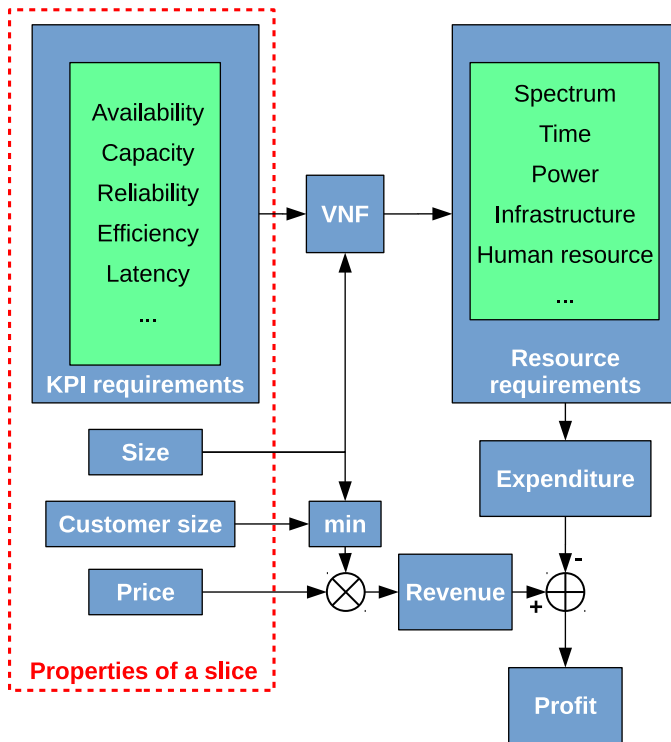


Fig. 3. Mapping chain that converts properties into profit

(MNOs) with own network infrastructure/resources and virtual MNOs (tenants), which rents infrastructure and network resources from traditional MNOs to implement and deliver services, two different business modes are active in 5G networks, and hence also two different profit models.

- **Own-Slice Implementation:** When a MNO implements a slice with its own network infrastructure and resources, it holds the full a priori knowledge about the slice, including the VNF scaling characteristic, the implementation cost, the customer demand for the slice service and the service charge rate. Therefore it can estimate the revenue and expenditure generated by the slice for an arbitrary certain size, as shown in Fig. 3. By appropriately scaling the sizes of different slices, the MNO is able to optimize the overall network profit under the limit of its resource pool, which is an optimization problem in resource allocation. This business mode and profit model has been discussed in [8].
- **Resource Leasing for Outsourced Slices:** Alternatively, a MNO can also make revenue by leasing its network resources in predefined bundles to tenants for their slice implementation. Different resource bundles can be provided with respect to the slice characteristics such as service type and performance elasticity. In this case, the leases can only be requested by the tenants while the MNO makes the decision if to accept or to reject the request. Once a lease is confirmed the MNO cannot flexibly scale or terminate the tenant-slice. Nevertheless, by setting up an appropriate decision mechanism according to the statistical knowledge

of requests, the MNO can still optimal the long-term overall revenue under the limit of its resource pool. An example of this business case an the optimization problem has been reported in [9].

Detailed discussion and comparison of both of these two dimensions of cost modeling is out of scope of our paper. However, it can further be developed along with below mentioned research directions in the future.

V. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

Network slicing is still at its early stage in terms of development, therefore, further enhancement and studies are needed to become a mature technology and thus be adopted over various domains of emerging 5G system. Despite significant advantages that network slicing brings to 5G system, there are also some challenges that are arisen. In this section, we are going to identify main research problems and future directions in the field of network slicing that are needed to be investigated for its full implementation over all parts of 5G system [1] [5] [6].

1-Business and Profit: The integration of various slices - each provides service at different stages and targets specific end users, and partnership between several operators through infrastructure sharing lead to create new challenges for total network investment, service level agreement between owner and tenant, and expected generating revenue. Considering such an integrated business-oriented approach, new economical strategies and profit modeling should be extensively analyzed and developed in order to meet 5G network requirements. To reach this goal, a deep study of existing telecom regulatory framework has to be conducted. New innovative ways of pricing, cost of infrastructure sharing, service level agreement between the owner and tenant of slice, and expected generating revenue should be addressed and furthermore standardized.

2-Security: Existing and some of proposed open interfaces in network slicing, which support network programmability lead to bring new potential attacks to softwarized networks. These concerns are arising major barriers on the way to deploy 5G networks, and call for an extensive study and development of multi-level security framework consisted of both policies and mechanisms, dynamic threat detection, user authentication, accounting management, and remote attestation.

3-Management: Despite efficient dynamicity and higher scalability that network slicing brings to 5G system, network management and orchestration in multi-tenant scenarios are considered to be major concerns. In order to dynamically assign network resources to different slices, the optimization policy that manages resource orchestrator should deal with situations where resources demands are vary. To accomplish this goal, i) an effective cooperation between slice-specific management functional block and resource orchestrator is required, ii) all policies are needed to be automatically validated, and iii) computationally design all resource allo-

cation algorithms and conflict resolution mechanisms at each abstraction layer.

4-Performance: The 5G communication system is composed of several virtual networks, different radio access technology, and various QoS requirements over the same infrastructure. When network slices are deployed, network performance analysis and QoS measurement can become more challenging and complicated tasks. Therefore, an intensive study is required to provide solutions for dynamic performance measurement and network analysis considering both time and cost.

5-Standardization: Network slicing standardization process is still in its initial phase and is mostly focused on vertical slicing. There are wide range of studies being conducted on network slicing by various research projects including NGMN, 5G NORMA, Co-Funded framework, WWRF, 3GPP, and 5GPPP. In all these projects, network slicing is considered as one of the most fundamental requirements of 5G communication system. The current status of network slicing development is dealing with concept, system architecture, requirements at different network subsections, and the impact of slicing on network architecture. Above all these various research directions of network slicing, a comprehensive global standardization is required to be deployed in emerging 5G system. Full standardization process of network slicing is expected to be finalized in the context of Release 15 and beyond of 3GPP.

6-Access Network Virtualization: Core network slicing has been already investigated, and there are numbers of available literature, which focus on efficient slicing of this particular subdomain. However, one of the main challenges for further network virtualization lies on the RAN of 5G system. As 5G network is composed of multiple access technologies, therefore, it is vital for RAN virtualization solutions to be able to accommodate these various technologies. This presents an additional comprehensive challenge on the RAN, since it is unclear so far, whether multiple access technologies can be multiplexed over the same hardware or each will need its own dedicated hardware. The answers to these questions can further be investigated and provided by conducting specific studies and various research on this particular topic.

VI. CONCLUSIONS

In this paper, we have provided a comprehensive discussion on concept and system architecture of network slicing in 5G mobile communication with particular focus on its business aspect and profit modeling. We have thoroughly discussed two different dimensions of cost modeling, called own-slice implementation and resource leasing for outsourced slices. We have further addressed existing challenges and open research directions in the field of network slicing for the sake of adding new contributions and providing realistic solutions to the problems.

In future, we intend to extend various dimensions of this paper. In the system architecture aspect, we are interested

to work on the development of horizontal slicing implementation over RAN architecture and end user devices. On the profit modeling side, we are interested to design and develop realistic models of resource cost and service revenue with detailed parameters. Despite above mentioned two research directions, there are also some other very important research challenges that are thoroughly described in the previous section of this paper.

ACKNOWLEDGMENT

This work has been performed with supports of the H2020-MSCA-ITN-2015 project 5GAuRA. The authors would like to acknowledge the contributions of their colleagues. This information reflects the consortia's view, but the consortia are not liable for any use that may be made of any of the information contained therein.

REFERENCES

- [1] Q. Li, G. Wu, A. Papatthanassiou, and U. Mukherjee, "An end-to-end network slicing framework for 5g wireless communication systems," 2016.
- [2] O. Sallent, J. Perez-Romero, R. Ferrus, and R. Agusti, "On radio access network slicing from a radio resource management perspective," *IEEE Wireless Communications*, vol. PP, no. 99, pp. 2–10, 2017.
- [3] I. da Silva, G. Mildh, A. Kaloxylou, P. Spapis, E. Buracchini, A. Trogolo, G. Zimmermann, and N. Bayer, "Impact of network slicing on 5g radio access networks," in *2016 European Conference on Networks and Communications (EuCNC)*, June 2016, pp. 153–157.
- [4] Y. L. Lee, J. Loo, and T. C. Chuah, "A new network slicing framework for multi-tenant heterogeneous cloud radio access networks," in *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEEES)*, Nov 2016, pp. 414–420.
- [5] J. Ordonez-Lucena, P. Ameigeiras, D. Lopez, J. J. Ramos-Munoz, J. Lorca, and J. Folgueira, "Network slicing for 5g with sdn/nfv: Concepts, architectures, and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 80–87, May 2017.
- [6] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, May 2017.
- [7] M. Jiang, M. Condoluci, and T. Mahmoodi, "Network slicing management and prioritization in 5g mobile systems," in *European Wireless 2016; 22th European Wireless Conference*, May 2016, pp. 1–6.
- [8] B. Han, S. Tayade, and H. D. Schotten, "Modeling Profit of Sliced 5G Networks for Advanced Network Resource Management and Slice Implementation," in *The 22nd IEEE Symposium on Computers and Communications (ISCC)*, July 2017.
- [9] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, and X. Costa-Perez, "Optimising 5g infrastructure markets: The business of network slicing," in *The 36th IEEE International Conference on Computer Communications (INFOCOM)*, 2017.
- [10] M. Richart, J. Baliosian, J. Serrat, and J. L. Gorricho, "Resource slicing in virtual wireless networks: A survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 462–476, Sept 2016.
- [11] N. Alliance, "Description of network slicing concept," *NGMN 5G P*, vol. 1, 2016.