The Algorithmic Aspects of Network Slicing

Spyridon Vassilaras[1], Lazaros Gkatzikis[1], Nikolaos Liakopoulos[1], Ioannis N. Stiakogiannakis[1], Meiyu Qi[1], Lei Shi[2],
Liu Liu[2], Mérouane Debbah[1], and Georgios S. Paschos[1]

[1]Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies Co. Ltd.

[2]Huawei Network Research, China

Abstract—**Network slicing is a technique for flexible resource provisioning in future wireless networks. With the powerful SDN and NFV technologies available, network slices can be quickly deployed and centrally managed, leading to simplified management, better resource utilization, and cost efficiency by commoditization of resources. Departing from the one-type-fits-all design philosophy, future wireless networks will employ the network slicing methodology in order to accommodate applications with widely diverse requirements over the same physical network. On the other hand, deciding how to efficiently allocate, manage and control the slice resources in real-time is very challenging. This paper focuses on the algorithmic challenges that emerge in efficient network slicing, necessitating novel techniques from the communities of operation research, networking, and computer science.**

Index Terms—*Network Slicing; Software Defined Networks; Network Function Virtualization; Virtual Network Embedding; Next Generation Networks*

I. INTRODUCTION

Traditionally, cellular networks have been architected to support specific services, namely voice, messaging, and Internet access. However, wireless operators are now faced with the major challenge to support a number of diverse vertical industry applications in order to expand the wireless market. Thus, next-generation networks should simultaneously accommodate applications and services with requirements as diverse as ultra-low latency and high resilience for real-time control of critical systems or scalability to hundreds of thousands of connected devices towards the Internet-of-Things (IoT). Table I provides a summary of typical examples of such services which illustrate the wide diversity of their associated requirements. *Network Slicing* is a key enabling technology for this paradigm shift.

A. The Network Slicing Concept

The concept of Network Slicing is not a new one; it has been proposed in the context of distributed service architectures, e.g., content delivery networks, large scale distributed testbed platforms, and distributed cloud computing systems [1]. However, its introduction to *wireless* networks is quite recent. A *network slice* is a virtual network which is implemented on top of a physical network in a way that creates the illusion to the slice tenant of operating its own dedicated physical network. A virtual link between virtual nodes A and B with capacity β_{AB} can be realized as a multihop physical path with reserved bandwidth β_{AB} on all physical links constituting the path. A virtual node implements a specific network functionality as a

physical node (such as a router or a firewall) would do in a traditional network. Virtual links can be easily established with Software Defined Networking (SDN) routers. SDN allows the administrator to remotely configure the physical network in order to reserve on demand networking resources for the slice. Virtual nodes can be implemented as Virtual Network Functions (VNF) running on general purpose hardware forming a cloud infrastructure. Network Slicing requires a high degree of flexibility which was only made possible by the recent advent of Network Function Virtualization (NFV) and Software Defined Networking. On a physical network consisting of SDN routers and datacenters with NFV functionality, it is possible to rapidly instantiate and reconfigure slices with diverse and time-varying requirements.

Network Slicing offers a number of significant advantages that are particularly useful in the design of next generation wireless networks, namely:

1) **Slice isolation**: The complete isolation of slices allows for a simpler and more efficient design of each slice with the goal of meeting the requirements of the particular vertical applications and services offered by the slice tenant. In addition, network failure, overload, or security attacks in one slice will not affect the operation of other slices in the network.

2) **Simplified service chains**: In contrast to traditional cellular communications in which all services consist of the same functions, in Network Slicing each service may rely on a different subset of functions.

3) **Flexible VNF placement**: NFV introduces an additional degree of freedom regarding the placement of these functions on the network. Intelligent placement may improve network performance and reduce operating costs.

4) **Transparent slice management**: Subsets of the physical network resources might belong to different network domains (or even operators). Network Slicing provides an abstraction of the physical resources and makes slice management transparent to the slice tenant.

To illustrate the flexibility offered by Network Slicing, we consider 3 indicative applications in Fig. 1. The figure shows basic VNFs required for each application as well as the corresponding network domains where these VNFs can be placed. For example, traditional Voice and Broadband services require complicated control plane functionalities such as authentication

TABLE I
USE CASES [2]

Case	Application	Requirements
Broadband Access in dense areas	Open-air event, stadium	High traffic volume, throughput (up to 10Gbps), ms latency
Broadband Access everywhere	Minimum coverage everywhere	Guaranteed 50+ Mbps
High user mobility	Trains, vehicles, aircrafts, drones	Connectivity in 3D and at over 500km/h
Massive Internet of Things	Sensors, smart wearables and meters	Diverse RATs, low power, 1 million connections per km^2
Extreme real-time communications	Robotic control, autonomous cars	Sub-ms latency, reliability, mobility
Ultra-reliable communications	Smart grid, eHealth, public safety	Redundancy, ms latency

and mobility management, placed at the core cloud. IoT services could be implemented with a simplified control plane; for example, a smart meter service that monitors the energy consumption of houses, does not require mobility management functionality. Video delivery services can be optimized if user plane and caching functionality is available at the edge cloud, which reduces backhaul traffic and improves user experience. Network Slicing enables service-specific resource allocation, which leads to a simplified, smaller, and cost efficient network.

For all these reasons Network Slicing has been recognized as a key element in the design of future wireless networks by the NGMN Alliance [2], the Open Networking Foundation (ONF) as well as many telecom vendors and wireless operators.
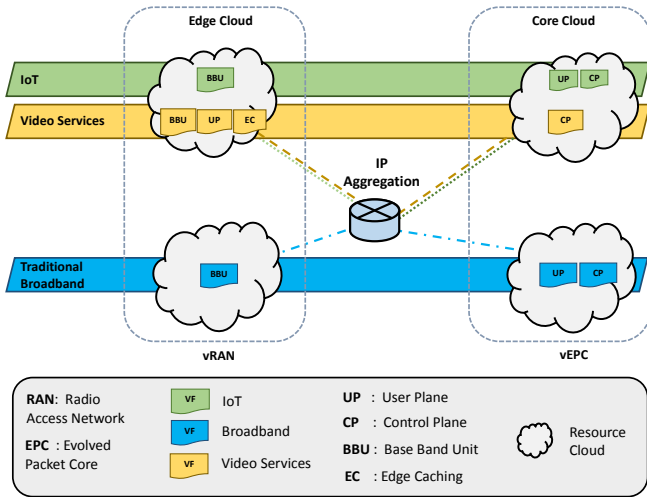

Fig. 1. Network slices supporting indicative applications with diverse requirements. Each slice consists of different VNFs which can be placed on different physical network domains.

B. A Wireless Network Slicing Architecture

Network Slicing has already been considered in late 4G and early 5G specifications. However this ongoing process has not yet produced a standardized architecture for Network Slicing. In order to discuss the involved algorithmic problems, we will focus on a generic architecture, explained below. Eventually, some of its aspects might differ from the future 5G Network Slicing standards, but we believe that the basic algorithmic components will remain the same.

The architecture we are considering is depicted in Fig. 2, showing the interactions of one (or more) physical *network operator* with multiple *enterprises* (e.g., OTT services, Virtual Network Operators). The role of the network operator is to provide the resources and to ensure the harmonic co-existence of different slices, while the role of enterprises is to place slice requests and then manage the provided slice. More specifically, in the lifecycle of slices, the following important operations take place: (i) planning of slice requirements (by enterprise), (ii) creation of slice (by network operator), (iii) intra-slice network management (by enterprise), and (iv) orchestration of different slices (by network operator). We detail them below.

For a new slice to be instantiated, an enterprise must first determine the required slice functionality and resources. It is envisioned that slice templates will be available for the most common types of services [3]. Thus, an enterprise may select the slice template that fits its purpose and parametrize it according to its needs. However, in order to be able to support novel services, a more flexible approach is also needed. In the extreme scenario, an enterprise could reserve processing, storage, and bandwidth resources at will, similarly to the way that resources are reserved in the cloud, and deploy on top only the necessary VNFs. In this case, the enterprise can determine the network topology (topological design) and the size of each component (dimensioning) via classical network planning tools, such as Netsim. The enterprises notify their slice requests via a specific interface, e.g., an intent-based language such as NEMO.

Upon receiving a slice request, the network operator faces the problem of embedding a concrete virtual network onto the physical network in an efficient way. This step involves decisions on placing and interconnecting several VNFs which can be formally expressed as a constrained optimization problem. The algorithmic considerations of this step are described in Section 2. Based on the solution to this problem the network operator creates the slice using SDN and NFV technology.

Once the slice has been created, the tenant (enterprise) can normally manage the provisioned slice resources in the same way it would manage a dedicated physical network. For this purpose, the network operator needs to expose a number of control functions to the enterprise controlling the slice.

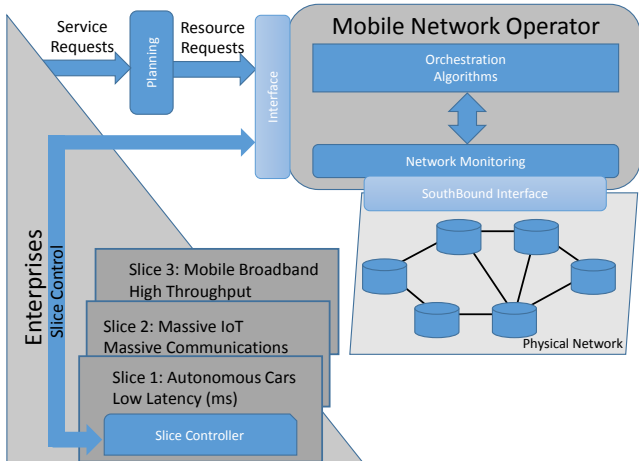In addition, from time to time, the enterprise may request

Fig. 2. A high level architecture for wireless Network Slicing.

to scale in or out the reserved resources, e.g., to address a traffic burst. The details of the complex interactions between the network operator and the enterprise during the lifetime of a slice are subject of ongoing development and standardization work [4] and are outside the scope of this paper. However, the presented abstraction level is sufficient for our discussion of dynamic Network Slicing algorithmic aspects in Section 3.

C. Aim and Objectives

The *Network Slicing* paradigm makes network management flexible and opens new horizons for network efficiency. At the same time, it raises a number of challenging issues which are receiving increasing attention from the research community. The goal of this paper is to provide directions on the most suitable algorithmic tools to address a multitude of Network Slicing problems.

To this end, in Section 2, we investigate the efficient Network Slicing problem and demonstrate that, for a given set of slices, orchestration at the operator level can be seen as a special instance of the Virtual Network Embedding (VNE) problem. We delineate several NS problem variants and extensions and identify open research topics. In Section 3, we analyze the envisioned mechanisms to deal with network requirements that vary over time and outline how existing network engineering techniques can be applied in the NS context. We conclude the paper in Section 4.

II. Optimal Network Slicing

The *Network Slicing* (NS) problem is a combined optimization problem of placing network functions over a set of candidate locations and deciding their interconnections. The mathematical problem can be formulated using the well-studied Virtual Network Embedding (VNE) problem–see [5] for a comprehensive survey.

A. Virtual Network Embedding

We are given a physical network G and a virtual network H, and we are asked to "embed" or "map" the virtual onto the physical network.

The physical network $G = (V, E, \boldsymbol{\beta}, \boldsymbol{c})$ has nodes $v \in V$ and links $e \in E$ characterized by capacities $\beta_v \geq 0$ and $\beta_e \geq 0$, respectively. In the online operation of the network, β can play the role of *residual capacity*, i.e., the remaining resource of the node or link after we take out the current utilization. Each node v and link e is also associated to a cost c_v and c_e, respectively. Depending on the application, costs may reflect congestion, preference in terms of operator agreements, load balancing, or real cost of operation.

The virtual network $H = (N, L, \boldsymbol{d}, M)$ with virtual nodes N and virtual links L, has capacity requirements d_n and d_l for each virtual node $n \in N$ and link $l \in L$ respectively. The usage cost is a linear function of the used capacity, i.e., the cost of a virtual node n with capacity demand d_n using a physical node v with cost c_v is $c_v d_n$. Each virtual node n can be embedded in exactly one physical node from a set of physical nodes M_n which is a subset of the set of all physical nodes V.

An embedding of H onto G, consists of mapping *i*) each virtual node $n \in N$ onto a physical node $v \in M_n$, and *ii*) each virtual link (m, n) onto a loop-free physical path, connecting the two physical nodes u and v to which the physical nodes m and n have been mapped. A *feasible embedding* is an embedding in which all link and node capacity constraints are satisfied, i.e., the sum of capacity demands of all virtual nodes embedded on a physical node is less than the capacity of this physical node and the sum of demands of all virtual links going through a physical link is less than the capacity of this link. The VNE problem is defined as finding the feasible embedding with the least cost.

The basic VNE problem is an ILP which is NP-hard as it can be proven by reduction to the *multi-way separator* problem. Even with a given virtual to physical node mapping, the problem of optimally allocating a set of virtual links to single physical paths reduces to the *unsplittable multicommodity flow* problem and therefore it is also NP-hard. Nevertheless, multiple heuristic approaches are available [5]. For example, one may decompose the original joint NS problem into a node embedding and an integral min cost multicommodity-flow problem (link embedding). In [6], the former is addressed via a heuristic based on availability of resources, whereas the latter via solving a continuous relaxation of the ILP followed by rounding.

B. Network Slicing

To decide how to create a slice we may equivalently study the embedding of an appropriate virtual network, reflecting the required slice components. We define the slice $H' = (N, L, \boldsymbol{d}, M)$, where $L, \boldsymbol{d}$ are links and their capacity requirements as before, while the virtual nodes N represent VNFs. As a result, the connectivity of H' also describes how the different VNFs are connected. In this case, the location constraint sets
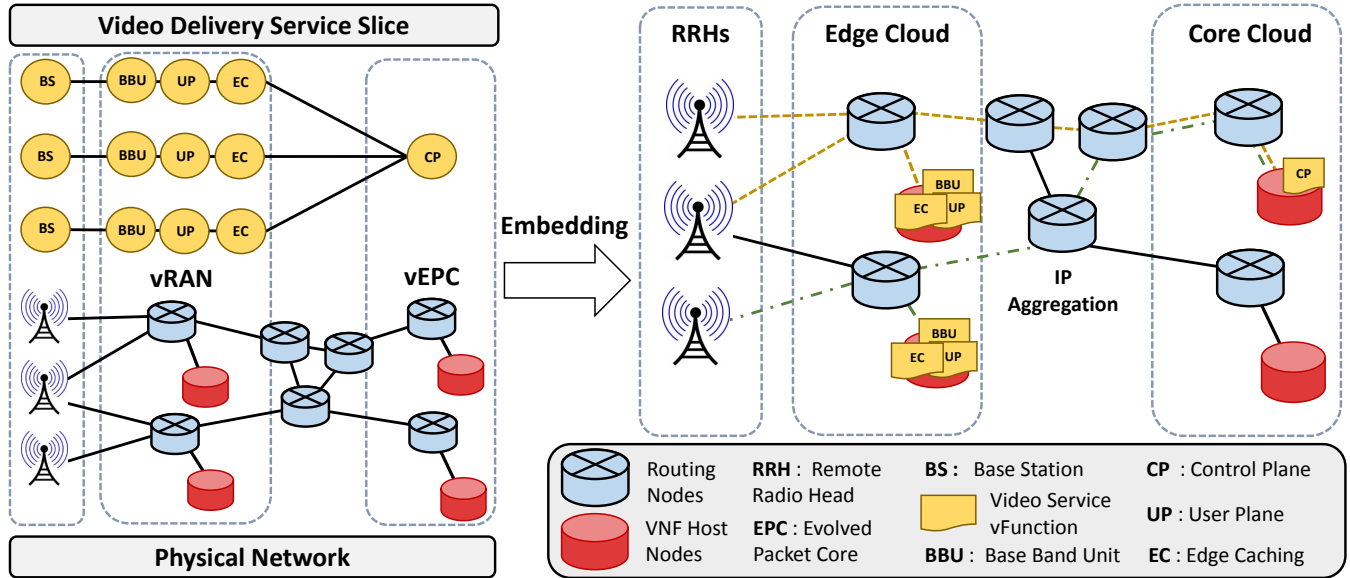
Fig. 3. An example of embedding a video delivery slice on a simplified wireless network. Here, CP and UP refers to one or more VNFs at the corresponding plane.

M_n can be used to capture both the capabilities of physical nodes to run a specific VNF and the location requirements of the applications and users of the network. The *Network Slicing* problem is to find the feasible slice embedding with the least cost. By appropriately selecting the embedding sets M_n and virtual links, costs, capacities, and demands, any NS problem can be represented as an extended VNE problem.

In Figure 3 we illustrate some key aspects of the network slicing problem. We focus on the implementation of a video delivery slice over a mobile wireless network. A graph representation of the virtual network, the substrate physical network and the resulting embedding of the VNFs are depicted.

The specific slice consists of four types of VNFs (Base Band Unit, Edge Caching, User Plane and Control Plane) and the corresponding communication links.[1] Supporting such a video delivery service, requires the implementation of one service chain per Base Station (BS) that terminates at the core network. Multiple VNFs may run on the same or different physical nodes, as long as connectivity is ensured and capacity constraints are satisfied. Notice that out of the several possible interconnection paths, the optimal one has to be selected.

The embedding of the 3 BSs is trivial, since each one has to be associated with a specific Remote Radio Head. BBU, UP and EC VNFs can be embedded on any of the physical resource nodes in the vRAN area, while the CP VNFs only on the vEPC resource nodes. This basic example illustrates that the optimal VNF placement and interconnection depends on all problem

parameters including physical node and link costs, capacities and communication requirements.

C. Basic NS as an optimization problem

From the above discussion, it is evident that the basic NS problem is a constrained *optimization* problem. In its simplest form this is a VNE-type of problem, in which we have to jointly decide i) the optimal placement of VNFs at resource nodes and ii) the necessary link capacity reservations for their interconnection, under additive link and node capacity constraints so that the overall resource utilization cost is minimized. Such problems, though NP-hard, can be cast as Integer Linear Programs (ILP). Thus, they can be solved using standard ILP solvers like CPLEX[TM] or Gurobi[TM].

Next, we quantify the potential cost benefits of optimized network slicing with a simple numerical example. Consider the physical network of Fig. 4. The physical link and node capacities/costs are shown in parentheses, while also the slice resource requirements are given per service category. For example, the video service (type II) has small bandwidth requirement between user plane and control plane functionality, due to edge caching which ensures that video service is obtained directly at the edge for popular videos. Similarly, IoT slices may rely on a simplified control plane and hence require a smaller bandwidth. Different network functions can be embedded either on vRAN, or on vEPC (the embedding options are given also in the figure), or on both in the special case of UP for video services. To quantify the benefit of network slicing, we embed these slices onto the physical network by solving the NS optimization problem.

---

[1]For simplicity we group together as a single VNF and denote by UP and CP a number of required VNFs in the User and Control Plane respectively which are not already shown as stand-alone VNFs
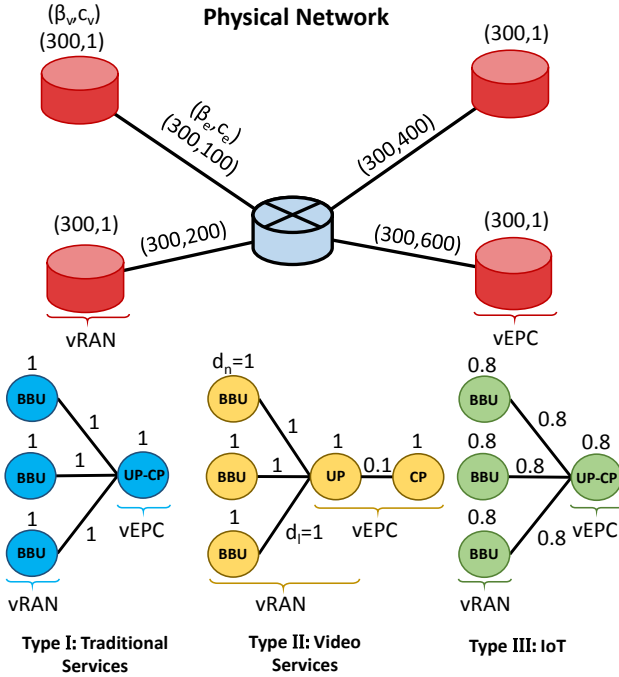
Fig. 4. The 3 types of services and the physical network used in the simulation

In Fig. 4 we depict the obtained network cost of utilized resources according to the optimal slicing[2]. The baseline assumes that all three services are treated according to the traditional mobile broadband approach, as Type I. To perform the comparison, we create two possible mixes of traffic representing future projections, a) one representing medium market penetration in which slices of types I, II and III are requested in proportion 2:1:1, and b) one where slicing has become the de facto standard and the requests are in proportion 1:2:2. Then we scale up the slice resource requirement. Cost savings of up to 25% and 40% can be achieved through slicing in the two scenarios. Interestingly, the improvement is not monotonically increasing with network load. In some situations of high load, the vRAN resources are exhausted and cannot accommodate further caching functionality, leading to a reduction of relative cost benefit. This indicates that the slicing flexibility is better exploited under proper dimensioning of the network.

For real life network scenarios, exact ILP solution approaches are impractical as they require immense computational and storage capabilities. In addition, the time requirement to instantiate a new slice might be in the range of seconds or less, whereas solving such an optimization problem exactly would require tens of minutes. For this reason, as already mentioned, heuristic approaches are used to provide near-optimal solutions to the basic NS/VNE problem in a short time.

[2]By setting all costs equal to 1, one may optimize resource utilization

D. Variants of the Network Slicing problem

A number of variants and extensions to the basic NS problem are of great practical interest for next generation wireless networks.

Survivability constraints. In this case, redundant physical resources (nodes or paths) must be reserved to protect the slice from physical node or link failures. Many types of survivable slices can be considered: protection against single or multiple failures, correlated failures with Shared Risk Groups, different types of protection (1+1, 1:1, shared backup, etc), and different recovery schemes. Any efficient algorithm which solves the basic NS problem using a shortest path algorithm (such as Dijsktra's algorithm) as a subroutine, could be extended to solve the 1+1 protection problem by substituting this subroutine with an efficient algorithm solving the 1+1 shortest pair of paths problem [7]. On the contrary, any protection scheme with shared backup introduces significant complexity making efficient algorithms for the basic problem inadequate for the extended problem.

QoS constraints. The feasible embedding must also satisfy some QoS constraints. For example each physical element (link or node) is associated with a fixed latency and each virtual link can tolerate up to a maximum end-to-end latency. For each virtual link embedded to a given physical path, the sum of the individual latencies of the nodes and links forming the path must be less than the maximum virtual link latency. A number of other QoS constraints, such as maximum jitter, maximum probability of packet loss, etc., can be applied in isolation or in combination, substantially increasing the complexity of the problem. As in the survivability case, an efficient algorithm for the basic problem can be easily extended to the case of a single additive QoS constraint (such as a maximum delay constraint) by using an efficient algorithm for constrained min cost path computation.

Optical network constraints. As a large part of the network infrastructure of the future will be optical, taking optical network constraints into account is of significant importance. Optical network constraints are related to wavelength continuity and physical layer impairments (PLIs). PLIs cause notoriously difficult problems in path selection because they create non-linear degradation to the optical signal. PLIs are also known to cause cross-wavelength signal interference. This can lead to the extremely complicated situation where one virtual network embedding can cause quality degradation to other embeddings and therefore impose additional constraints to the NS problem. Multi-layer (electrical and optical) NS has also received attention in the literature. In all such cases, the most promising approach is to decompose the problem into many independent subproblems [8].

Distributed operation. A centralized NS algorithm has complete knowledge of the physical network and all slices that need to be embedded. In some cases, this centralization of information might not be possible. Examples could include slices that traverse multiple operators and multiple SDN domains, or
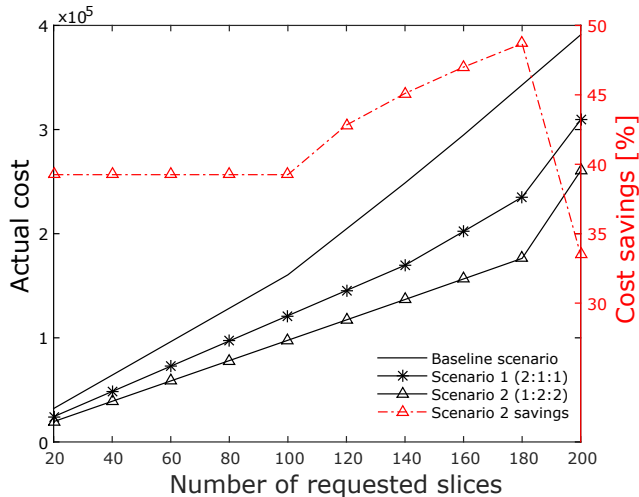
Fig. 5. Quantifying the benefits of network slicing for heterogeneous services in resource utilization. Whereas the actual costs are increasing with load, the relative cost savings, as depicted by the red line, are improving up to the point where the vRAN resources are exhausted.

involve scalable designs with multiple SDN controllers. In these situations, a distributed slicing algorithm is required. Auction algorithms assuming a cost-utility model for the various actors in NS have been proposed for centrally supervised or fully distributed VNE and NS (see [9] and references within).

NFV specific constraints. The virtualization of network functions introduces novel characteristics that do not appear in traditional networking. For example, the bandwidth requirement of a flow may change as it passes through a VNF due to compression/decompression, or processing at VNF may introduce additional node latency which is a function of the overall VNF load. Although it has been shown that such constraints can be incorporated in an ILP NS formulation [10], deriving efficient NS algorithms that address those issues is an open research topic.

E. Unique challenges of Network Slicing

A number of important requirements are unique to network slicing and have not been addressed in the literature, to the best of our knowledge. These requirements are divided in the following broad categories:

End-to-end constraints. In the QoS routing literature, QoS constraints (such as maximum allowable end-to-end latency) are typically imposed per path and the possibility of a constraint across many paths is not considered. In NS, it is reasonable to consider an end-to-end latency requirement for a service chain, which is a latency constraint across a tandem of paths.

Heterogeneous requirements. In traditional network embedding, extensions such as QoS and survivability constraints are uniformly applied to the entire network. In NS, it is envisioned that different slices will have different requirements as they will serve diverse end-user applications. Even in the same slice, some links might require an increased degree of

protection or QoS compared to others. For many solution approaches extending the algorithm to the heterogeneous requirements case is not straightforward.

Multitenancy and non-linear resource utilization. Multitenancy has been extensively studied in the literature of data center virtualization, since running multiple virtual machines on the same physical server introduces a non-negligible resource utilization overhead. In practice, due to contention for the shared physical resources, the utilized computational capacity on a virtual node is more than the sum of the resources allocated to each virtual node. This type of resource utilization models can be captured with the use of non-linear functions. Similarly, if statistical multiplexing of traffic is desired, this can be modeled with non-linear utilization functions on the physical links. Clearly, departing from linear cost functions complicates the solution of the resulting optimization problems considerably.

Slice fairness. A challenging problem is to decide how to split resources among different competing slices. In classical network flows, this challenge is resolved by the use of fairness metrics. However, the basic notion of fairness must now be generalized to involve entire slices instead of flows, and allow the consideration of multiple resources, like bandwidth, processing power, and memory, which leads to the problem of *slice fairness*. Some directions to resolve slice fairness include formulating the NS with convex utilities, and using the concept of multi-resource fairness [11] from the related literature of cloud computing.

III. Dynamic Network Slicing

Communication networks are inherently dynamic: a widely known example is the diurnal fluctuation of network flow that follows human activity. Other phenomena may also lead to time-varying slice requirements. Cultural and sports events, service attacks and server downtime, variability of wireless channels, time-varying cost of virtual resource at different locations, failures of optical links, etc. *The primary goal of dynamic Network Slicing is to allow the network operator to reconfigure and migrate the slices in order to match the network variability.*

Dynamic Network Slicing is designed to support elastic networks, i.e., networks that change shape over time.[3] In elastic networks, the classical resource reservation explained in the previous section may not be enough, since the slices may need to scale in and out over time, or even change shape. Below we explain how to use monitoring, and online optimization to provide effective dynamic Network Slicing.

A. Network State Monitoring and Prediction

To correctly drive the resource management decisions in the dynamic setting, the network operator needs to maintain an accurate depiction of resources that are currently used and

---

[3]For example, the elastic Content Distribution Networks (elastic CDNs) by Akamai Technologies and Juniper Networks use virtualized caching servers which can be installed at different locations possibly closer to the user demand.

their availability over time. A challenging problem therein is to model the resource utilization as a time evolving process.

Current network monitoring tools provide a static view of utilization for each network resource (link and node capacities, VNF capabilities, etc.). In a dynamic environment, the impact of embedding a particular slice on future slice requests is unclear. For example, consider the case of optimally embedding a slice based on static information. Once a new slice request arrives, a reconfiguration of the old slice might be needed due to resource contention. However, slice reconfiguration comes at a cost and hence a prediction mechanism of future resource utilization can be helpful.

Typically, prediction mechanisms rely on the use of historical data. In our context, techniques from machine learning can be used to exploit the raw monitored data from the SDN controllers and derive predictions for the impact of new slices into the network [12].

B. Online Optimization

The methodology of dynamic Network Slicing creates an arena of online optimization problems. To optimize the use of available resources and meet the time-varying slice requirements, the network operator needs to constantly optimize the slice resource allocation, while deciding to admit or not new slices. This befalls to the area of online optimization, where powerful algorithmic tools exist, such as the stochastic network optimization and the domain of online competitive algorithms [13]. The *online Network Slicing* problem is related to other classical online problems such as (i) the online minimum cost multicommodity problem, (ii) the online network embedding problem, (iii) the online VNF placement problem, (iv) the online packing problem, (v) the online facility location problem, and different variations of them. Below we consider methodological approaches for optimization problems of Network Slicing with an evolutional character.

A typical way of solving online problems is by using the offline optimization counterpart in two phases: quick assignment and readjustment [14]. The quick assignment phase exploits quick but suboptimal algorithms to decide how to embed the slices one by one. These assignments bring the system to a suboptimal configuration. Then the reconfiguration algorithms resolve the global offline resource optimization and the slices are reconfigured into a well-performing configuration. Tradeoffs between cost and frequency of reconfigurations have been studied in [14] for the case of minimum cost multicommodity flow problem.

When the global optimization is difficult (as is most cases in Network Slicing), the mentioned approach spends considerable amount of time in suboptimal configurations. To alleviate this phenomenon, a promising direction is to utilize the technique of Bandwidth Calendaring [15]. The idea applies to slices with either predictable behavior or requested with an early prebooking. The network operator can construct a calendar of slice demands and resolve a time-expanded NS optimization problem covering a large period of time, e.g., a week or more.

When calendaring construction is not possible, the online optimization problems can be solved directly by online algorithms. The classical approach in this realm is to model the problem in the adversarial setup, meaning that the algorithms fight against the worst possible scenario of demand arrivals [13]. Although this typically results in conservative algorithms, it also provides very strong results in the form of algorithms with competitive ratios.

IV. CONCLUSION

Network slicing is a novel methodology for provisioning resources in the upcoming wireless networks. This paper illustrates numerous algorithmic challenges of slice optimization, which also represent promising research directions. Tools from the operation research, theoretical networking, and computer science are envisioned to provide network optimization algorithms and control techniques which can give practical answers to these challenges.

REFERENCES

[1] F. Esposito, I. Matta, and V. Ishakian, "Slice Embedding Solutions for Distributed Service Architectures," *ACM Computing Surveys (CSUR)*, vol. 46, no. 1, pp. 6:1–6:29, Oct. 2013.

[2] NGMN Alliance, "NGMN 5G White Paper," *Next Generation Mobile Networks Ltd., Frankfurt am Main*, Mar. 2015.

[3] N. Nikaein, E. Schiller, R. Favraud, K. Katsalis, D. Stavropoulos, I. Alyafawi, Z. Zhao, T. Braun, and T. Korakis, "Network Store: Exploring Slicing in Future 5G Networks," *10th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch)*, Sep. 2015.

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

[5] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, "Virtual Network Embedding: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1888–1906, Apr. 2013.

[6] R. Mijumbi, J. Serrat, J.-L. Gorricho, and R. Boutaba, "A Path Generation Approach to Embedding of Virtual Networks," *IEEE Transactions on Network and Service Management*, vol. 12, no. 3, pp. 334–348, Jul. 2015.

[7] M. R. Rahman and R. Boutaba, "SVNE: Survivable Virtual Network Embedding Algorithms for Network Virtualization," *IEEE Transactions on Network and Service Management*, vol. 10, no. 2, pp. 105–118, Jun. 2013.

[8] S. Peng, R. Nejabati, and D. Simeonidou, "Impairment-Aware Optical Network Virtualization in Single-Line-Rate and Mixed-Line-Rate WDM Networks," *OSA Journal of Optical Communications and Networking*, vol. 5, no. 4, pp. 283–293, Apr. 2013.

[9] F. Esposito, D. D. Paola, and I. Matta, "On Distributed Virtual Network Embedding With Guarantees," *IEEE/ACM Transactions on Networking*, vol. 24, no. 1, pp. 569–582, Feb. 2016.

[10] B. Addis, D. Belabed, M. Bouet, and S. Secci, "Virtual Network Functions Placement and Routing Optimization," *IEEE 4th International Conference on Cloud Networking (CloudNet)*, Oct. 2015.

[11] A. Ghodsi, M. Zaharia, B. Hindman, A. Konwinski, S. Shenker, and I. Stoica, "Dominant Resource Fairness: Fair Allocation of Multiple Resource Types," *8th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, Apr. 2011.

[12] L. Nie, D. Jiang, L. Guo, and S. Yu, "Traffic Matrix Prediction and Estimation Based on Deep Learning in Large-scale IP Backbone Networks," *Elsevier Journal of Network and Computer Applications*, vol. 76, no. C, pp. 16–22, Dec. 2016.

[13] G. Even, M. Medina, and B. Patt-Shamir, "On-Line Path Computation and Function Placement in SDNs," *B. Bonakdarpour, F. Petit (eds.) Stabilization, Safety, and Security of Distributed Systems, Springer Lecture Notes in Computer Science*, vol. 10083, Nov. 2016.

[14] S. Paris, A. Destounis, L. Maggi, G. S. Paschos, and J. Leguay, "Controlling flow reconfigurations in SDN," *35th Annual IEEE International Conference on Computer Communications (INFOCOM)*, Apr. 2016.

[15] L. Gkatzikis, S. Paris, I. Steiakogiannakis, and S. Chouvardas, "Bandwidth calendaring: Dynamic services scheduling over Software Defined Networks," *IEEE International Conference on Communications (ICC)*, May 2016.

BIOGRAPHIES

Spyridon Vassilaras [M'05–S'15] is a principal research engineer at the Huawei France Research Center since Dec. 2014. His main research interests are in network performance evaluation and optimization. He received the Dipl.Eng. degree from the National Technical University of Athens (1995) and the M.S. and Ph.D. degrees from Boston University (1997, 2001). From 2003 till 2014 he was a researcher and professor at Athens Information Technology. He has participated in 9 EU funded projects and published numerous research papers.

Lazaros Gkatzikis [S'09–M'13] is a senior research engineer at the Huawei France Research Center. Previously, he held research-related positions at KTH Royal Institute of Technology (Sweden), Technicolor (France), CERTH-ITI and University of Thessaly (Greece). His research interests include network optimization, game theory, and performance analysis. He received his Ph.D. degree in computer engineering and communications from the University of Thessaly, Volos, Greece.

Nikolaos Liakopoulos received his B.S. in Physics and M.S. in Control and Computing from National Kapodistrian University of Athens, Greece, in 2012 and 2015. Since 2016 he has joined Huawei FRC, in Network Control and Resource Allocation Team and is working towards his PhD in collaboration with Eurecom and UPMC. His research focuses in distributed and centralized control for wireless networks and design of algorithms. Specifically, in Software Defined Wireless Networks and 5G Network Slicing.

Ioannis Stiakogiannakis [S'10–M'14] is a senior research engineer at Huawei. His experience has been on the design, development and evaluation of algorithms for communication networks, such as 5G, LTE, elastic optical networks, and software defined networks. He is an IEEE Wireless Communications Professional, certified with IEEE WCET. He received the Dipl.-Ing. and Dr.-Ing. degrees from the School of Electrical and Computer Engineering, National Technical University of Athens in 2007 and 2011 respectively.

Meiyu Qi received her B.Sc. in Mathematics from University of Tianjin (China) in 2005. Then she obtained her M.S. in Operation Research from the same university in 2007. She finished her doctoral thesis at University of Augsburg (Germany) in 2011. In the same year, she joined Huawei as a networking algorithm researcher in the Transmission Technologies Research Department. Her research expertise focuses on mathematical programming algorithmic / meta-heuristic algorithmic applications in optical and IP networks. Since 2013, she works on SDN-related algorithm research.

Liu Liu received the M.Sc. and Ph.D. degrees in communication and information systems at the University of Electronic Science and Technology of China in 2011 and 2015, respectively. He was a visiting scholar in computer science and engineering at SUNY at Buffalo from 2012 to 2014. He joined Huawei as a research engineer in 2015. His research interests focus on network planning and optimization, uncertainty optimization, approximation algorithms, and cloud computing.

Lei Shi got his master of Computer Science from University of Sichuan (China) in 2005. In the same year, he joined Huawei as an engineer in the WDM product line, and in 2007 became a networking algorithm researcher in the Transmission Technologies Research Department. His research expertise focuses on mathematical programming algorithmic / meta-heuristic algorithmic applications in IP, optical and microwave networks.

Mérouane Debbah [S'01–M'04–SM'08–F'15] entered the Ecole Normale Suprieure de Cachan (France) in 1996 where he received his M.Sc. and Ph.D. degrees respectively. Since 2014, he is Vice-President of the Huawei France R&D center and director of the Mathematical and Algorithmic Sciences Lab. His research interests lie in fundamental mathematics, algorithms, statistics, information & communication sciences research. He is an IEEE Fellow and a WWRF Fellow. He has managed 8 EU projects, 24 national and international projects and received more than 16 best paper awards for his research contributions.

Georgios S. Paschos [S'01–M'06–SM'15] is a principal researcher at Huawei Technologies, Paris, France, leading the Network Control and Resource Allocation team, since Nov. 2014. Previously, he held positions at VTT (Finland), CERTH-ITI and University of Thessaly (Greece), and MIT (USA). He received his Dipl.Eng. from Aristotle University of Thessaloniki (2002), and his Ph.D. degree from the University of Patras (2006). He has received two best paper awards and serves as an associate editor for IEEE/ACM Trans. on Networking and as a TPC member of IEEE INFOCOM.