

Modeling and optimization of 5G network design

Wesley da Silva Coelho

CNAM, Orange Labs

Chatillon, France

wesley.dasilvacoelho@orange.com

Abstract—Every second, a large amount of digital data is transported through infinite types of devices via cellular networks worldwide, and expectations are at a greatly accelerated growth with increasingly large requests. In few years, these networks could thereby reach their maximum capacities in terms of data transmission. To face these challenges, Network Slicing has been presented as a novel virtualized infrastructure for the new generation cellular network system. This technology now not only covers application-level abstraction but also physical and switching layers virtualization, with different radio access and link communication technologies. Hence, each service provider is to be able to deploy its communication services on top of logical networks, named Network Slices, specifically tailored to its technical requirements. The Device-to-Device communication mode is another approach presented as a promising alternative to traditional communication in cellular networks. This technology allows to reuse radio resources and to decrease the end-to-end latency of local communications. Consequently, the optimization of physical resources in cellular networks becomes crucial to better deploy virtual networks. The overall objective of our research is therefore to define and study the concept of device-to-device communication and network slice design in 5G systems, and propose mathematical models and innovative algorithms to solve the underlying optimization problems.

Index Terms—Network slicing, device-to-device, network design, optimization

I. INTRODUCTION

The expectations around the data flow on mobile networks are at a greatly accelerated growth with increasingly large requests. In few years, these networks could consequently reach their maximum capacities in terms of data transmission. To face all these challenges, 5G technology is posed by enabling the digitization of society and economic information. The idea behind the concept of 5G is that it does not correspond to a simple increase in data rate, as was the case for previous generations. With 5G, the aim is thereby to widen the diversity of users.

This technological evolution will touch the whole network environment, going from cellular and radio access to the application service architectures. This transition challenges network design since multiple resources and segments, historically managed independently from each other, are to be operated with continuity in networking and computing resource allocation and provisioning as a whole and unique service. In this context, different providers can be associated with different communication services running on the same physical network at the access, core, and application segments.

This work is supported by the french Agence Nationale de la Recherche (ANR), Project MAESTRO-5G ANR-18-CE25-0012.

Such communication services can therefore be of three classes: enhanced Mobile Broadband (eMBB; e.g., broadband everywhere and large-scale events), Ultra-Reliable Low Latency Communications (URLLC; e.g., online gaming and autonomous driving), and massive Machine Type Communications (mMTC; e.g., device-to-device communication and internet of things) - differing in the requirements, such as maximum latency, minimum availability, and bandwidth. To provide the necessary flexible provisioning, Network Function Virtualization [1], Software Defined Networking [2], and Network Slicing [3] technologies can be adopted to let the communication service provider deploy its services on top of logical networks.

Because of different bitrate and latency requirements, policies on radio access function splitting are also going to have an impact on the backhauling network dimensioning, and therefore on the placement of core network functions and on the configuration of edge computing application servers. Moreover, different policies for control versus data-plane function sharing and scaling are expected to be applied.

To overcome these challenges, *Network Slice* has been presented as a novel virtualized infrastructure model. This technology now not only covers application-level abstraction but also physical and switching layers virtualization, with different radio access and link communication technologies. Hence, each service provider is to be able to deploy its communication services on top of logical networks, named Network Slices, specifically tailored to its technical requirements.

The Device-to-Device (D2D) communication mode is another new approach presented as a promising alternative to traditional communication in cellular networks. This technology allows to reuse radio resources and to decrease the end-to-end latency of local communications. Then, D2D would allow a set of UEs geographically close to each other to establish direct D2D communications, or span multiple links (multi-hop D2D communications), to access a given service (e.g. video streaming or gaming) while ensuring the required service quality.

In this context, optimizing resources in cellular networks becomes crucial on backhauling network dimensioning, and hence on the placement of core network functions and the configuration of edge computing application servers. Moreover, different policies for control versus data-plane function sharing and scaling are expected to be applied. All these challenges must be overcome wisely and effectively, as the state of the system can change every second. In legacy technologies, such

as 3G and 4G, the entire network system was designed for approximately ten years of use, with small variations and slow evolution over the period. Contrarily, the 5G system is meant to be extremely flexible and still be able to offer a customized and complete virtual network in minutes for each Communication Service request. The overall objective of our research is therefore to define and study the concept of device-to-device and network slice design in 5G systems and propose mathematical models and innovative algorithms to solve the underlying optimization problems.

II. OUR CONTRIBUTIONS

In the course of the Ph.D thesis, we first overviewed the evolution of the mobile system, from 3G to 5G, and analyzed the entities appearing in new generation networks as well the main modeling aspects and technical constraints related to 5G systems and beyond. This work was partially published in the IEEE Communications Standards Magazine [4]. Then, we address the network slice design problem in different facets; Figure 1 summarizes our research directions and main contributions.

A. Domain Creation Problem in Device-to-Device Communication

We defined and studied the concept of device-to-device communications and network slicing in 5G systems and propose mathematical models and innovative algorithms to solve the underlying optimization problems. In particular, we studied the Domain Creation problem, which is a routing and resource assignment problem arising in future 5G networks. We proposed two algorithms, exact and heuristic, to solve it. The exact approach is based on a node-arc ILP formulation enhanced by two families of valid inequalities that are used within a branch-and-cut framework. We also proposed a solving method based on a decomposition of the DCP into two sub-problems: the routing sub-problem and the resource allocation sub-problem. First, a significant impact of the cuts in strengthening the LP relaxation and reducing the computation time was observed. Despite a longer runtime for some instances using the proposed cuts, we could observe that the gap between the root solution and the final solution is always smaller when neighborhood cuts are applied. Using capacity cuts, this improvement is noted for half of the instances. However, the final size of the tree after finding the final solution using these cuts can be up to 76.40% smaller. Our experiments also showed that the proposed heuristic approach performs well, even on large instances with up to 2100 devices and 1500 service requests on 7-cell networks. It would be interesting and most probably very powerful to use it as a primal heuristic to boost the efficiency of an exact algorithm. Our results were presented at the International Network Optimization Conference and published in its proceedings [5].

B. Network Slice Design

We also introduced and studied the Network Slice Design Problem (NSDP) in 5G systems. We first proposed a

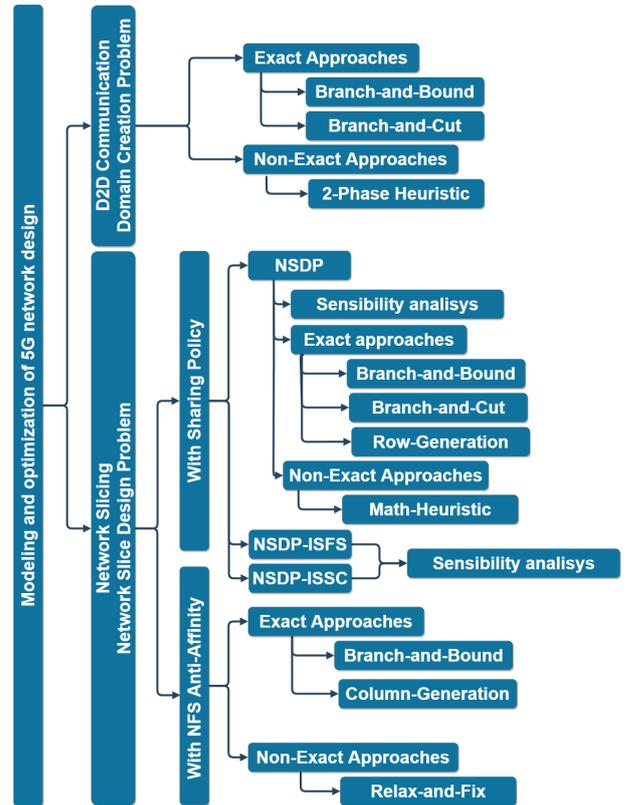


Fig. 1: Research direction and main contributions.

mixed-integer linear programming (MILP) formulation for the problem including novel *splitting*, *mapping* and *provisioning* constraints described in the published 5G standards documents [6]–[9]. Then, we modeled new variants and extensions of the problem: NSDP with intra-slice flexible splitting (NSDP-ISFS) and NSDP with inter-slice split continuity. We then provided several sensibility analyses regarding the impact of each proposed variant on the network. We demonstrated by simulation the impact of taking into full and partial consideration of the peculiar constraints rising from the standards. For instance, we reported numerical results showing that flexible splitting appears as a key factor to deal with heterogeneous requirements to deploy distinct communication services, leading to considerable network slice cost decrease. In our simulations, the number of NFSs needed to deploy the virtual networks could be reduced by up to 56% depending on which of the six proposed sharing policies is applied to each network slice. We also observed that different variants related to the flexible splitting have an important impact on the physical network; depending on the selected approach, the average load on physical links could be reduced by a factor of 3. Our results were presented at the International Conference on Network and Service Management and published in its proceedings [10].

In order to strengthen the linear relaxation of the proposed MILP, we proposed several classes of valid inequalities and integrate them into a Branch-and-Cut framework to solve the

problem. We further proposed several strategies to reduce the symmetries and the size of the model. Numerical experiments showed the efficiency of each approach on different instance classes. For instance, the proposed symmetry-breaking and lower bound-based constraints led to an important decrease in the final gap: from 30% to 1% in some instances. Also, our Branch-and-Cut algorithm could reduce the size of the Branch-and-Bound tree by 85% and outperformed the solver's Branch-and-Bound algorithm in all tests. Finally, our Row-Generation framework outperformed the Branch-and-Bound approach in all tests, especially in those with sparse graphs.

Moreover, to address the time complexity related to the proposed exact approaches, we proposed an open-access framework based on a Math-heuristic to address the underlying optimization problem. The overall idea of the proposed approach relied on decomposing the NSDP into several sub-problems and sequentially solve them while encompassing control-plane and data-plane separation and novel mapping and decomposition dimensions influencing the placement and interconnection of slices. Numerical experiments showed the efficiency of our approach on different instance classes, which could attain near-optimal solutions in a competitive runtime. Comparing it to a mixed-integer linear programming formulation, the proposed Math-Heuristic could reduce the average runtime and the final gap by up to 78% and 90%, respectively. Moreover, our approach could reduce the congestion on the physical network, better balancing the data flow while considering all technical constraints. For instance, the average load on physical links and physical nodes could be reduced by 16% and 35%, respectively. This work was accepted as a full paper at the International Teletraffic Congress [11].

Finally, we also studied another variant of the NSDP, where dedicated network functions are deployed to each network slice. Different strategies are then proposed in order to efficiently solve the optimization problems related to the problem. First, we propose a compact formulation and an extended one. To solve the former, we proposed a Relax-and-Fix heuristic that relies on repetitively solving the proposed related ILP with only a few integer variables and fixing or relaxing most of the remaining integer and binary ones. In our simulations, the average gap could be reduced from 82.8% to 4.31% (resp. 48.4% to 5.77%) on medium-big (resp. medium) instances for instance. Moreover, while ILP could not find any feasible integer solution within 1-hour runtime for any big-size instance, the average runtime and final gap were respectively 907 seconds and 5.2% when Relax-and-Fix was applied on the same instances.

C. Ongoing Work

To address the exponential number of variables in the aforementioned extended formulation, a column generation-based framework is proposed. To this propose, we start with the solution provided by the Relax-and-Fix algorithm while our pricing sub-problems are based on the compact formulation proposed to address the problem. The results of our simulation applying this approach will be soon presented.

III. PERSPECTIVES

On a practical note, a tough but interesting extension of our work related to the DCP is to include users' mobility and temporal aspect in radio resource assignment. Indeed, managing handovers is highly expected in real scenarios, which might potentially imply an important reconfiguration of the created domains in every short time period (i.e., minutes or even seconds). Also, other variants of the NSDP can be studied (e.g., applying Multi-Access Edge Computing) and a service-aware function objective might be proposed. In other words, different parameters might alternatively be optimized regarding the service related to each network slice (e.g., latency over capacity). Finally, an extension of our work on network slicing might also include availability constraints in order to ensure all technical constraints that are expected in the Service Level Agreements.

It is also expected to embed the proposed column generation-based algorithm into a branch-and-price framework to ensure the optimal solution for each instance. On the other hand, in order to address the time complexity related to the column generation-based algorithm proposed to the NSDP-DNF, heuristic-based pricing algorithms might be proposed in order to attain near-optimal solutions in a competitive runtime. Finally, applying clustering algorithms as pre-processing might also be a powerful approach to be applied in a distributed parallel programming, in which each thread might be responsible for solving smaller NSDP or DCP instances. For instance, such clustering algorithms can be based on geographical zone or even on service affinity.

REFERENCES

- [1] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, "Network function virtualization in 5G," *IEEE Comm. Mag.*, vol. 54, no. 4, pp. 84–91, 2016.
- [2] T. Chen, M. Matinmikko, X. Chen, X. Zhou, and P. Ahokangas, "Software defined mobile networks: concept, survey, and research directions," *IEEE Comm. Mag.*, vol. 53, no. 11, pp. 126–133.
- [3] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Comm. Mag.*, vol. 55, no. 5, pp. 94–100, 2017.
- [4] W. d. S. Coelho, A. Benhamiche, N. Perrot, and S. Secci, "Network function mapping: From 3G entities to 5G service-based functions decomposition," *IEEE Communications Standards Magazine*, vol. 4, no. 3, pp. 46–52, 2020.
- [5] A. Benhamiche, W. da Silva Coelho, and N. Perrot, "Routing and resource assignment problems in future 5G radio access networks," *Network Optimization INOC 2019*, p. 89, 2019.
- [6] 3rd Generation Partnership Project, "3GPP TR 38.913 V14.3.0: Study on scenarios and requirements for next generation access technologies," 2017.
- [7] —, "3GPP TR 28.801 V15.1.0: Study on management and orchestration of network slicing for next generation network," 2018.
- [8] —, "3GPP TS 23.501 V15.4.0: System Architecture for the 5G System," 2018.
- [9] —, "3GPP TR 38.801 V14.0.0: Study on new radio access technology," 2017.
- [10] W. da Silva Coelho, A. Benhamiche, N. Perrot, and S. Secci, "On the impact of novel function mappings, sharing policies, and split settings in network slice design," in *International Conference on Network and Service Management*, 2020.
- [11] —, "A Math-Heuristic for network slice design," in *33rd International Teletraffic Congress*, Avignon, France, Aug. 2021.