Dynamic Network Slicing for User Mobility Support in 5G Networks

Diogo M. Gonçalves[1,2], Luiz F. Bittencourt[1], Edmundo R. M. Madeira[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (Unicamp)
Campinas, SP, Brazil
[2]Instituto Tecgraf – PUC-Rio – Rio de Janeiro, RJ, Brazil

diogomg@lrc.ic.unicamp.br, {bit, edmundo}@ic.unicamp.br

Abstract. *Network slicing has been a key supporting technology for resource management in modern computer networks, from 5G to 6G and beyond. This paper summarises the thesis developed by the first author at the IC/UNICAMP. This work presents new resource management mechanisms for the dynamic network slicing allocation problem to support mobile users in 5G/edge computing infrastructures. A network simulator was developed to evaluate the proposed solutions. The simulator is available as free software and has been adopted as the validation environment by the scientific community. The results of this work indicate the positive impact of slices' dynamic resource allocation in dealing with demand variations triggered by mobile users.*

1. Introduction

5G networks aim to provide network resources for different user groups. The network must accommodate applications and services with diverse and sometimes conflicting requirements in parallel to serving users with different characteristics, such as mobility. Virtualisation technologies allow 5 G's physical resources to be virtualised and presented as virtual networks. That process, named network slicing, creates a one-demand virtual network (slice) for each user group. Due to demand variations caused by user mobility, the 5G infrastructure needs to update its resource distribution among services and slices dynamically. Some challenges arise in that scenario, such as defining which resources should be allocated for each slice and when and where to perform a service migration to keep the service as close as possible to its users. Everything is made in parallel, considering mobility management requirements and other trade-off scenarios. Figure 1 illustrates one scenario of dynamic slicing with bandwidth reallocation triggered by user mobility.

Based on that scenario, this work developed new resource management solutions for two relevant challenges related to the network slicing context: network slicing allocation and service migration. This thesis raised seven research questions, which were answered during the development of this work. The contribution of this work provides a foundation for further understanding the strengths and weaknesses points of network slicing technology for user mobility support.

The remainder of this document is presented as follows: Section 2 defines the problem statement and research questions. Section 3 summarises the contributions of this PhD thesis. Section 4 introduces MobFogSim, the simulator developed as the validation environment for the proposed solutions. Section 5 presents the proposed allocation approach, the proposed delay allocation model, the validation scenarios, and its results. Section 6 presents the outcomes in terms of scientific production. Finally, Section 7 presents the conclusions and future directions.

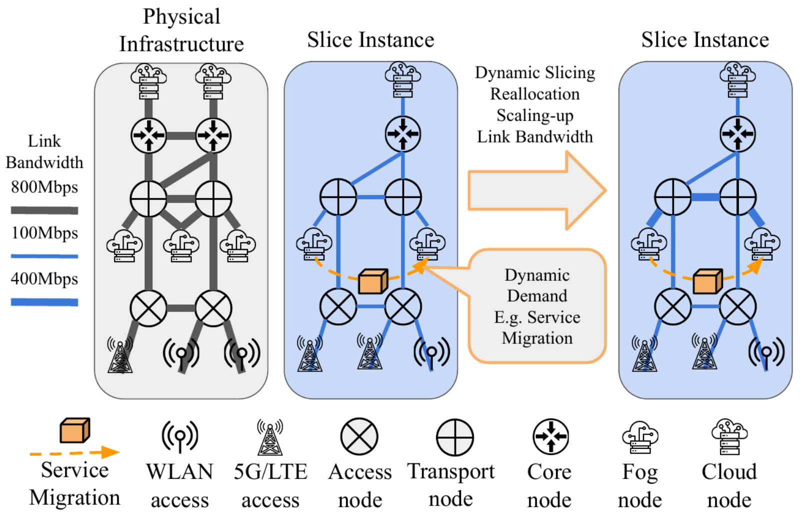

Figura 1. Example of the proposed bandwidth reallocation solution for dynamic slicing. Mobile users trigger traffic variations in the network, *e.g.*, service migration. Then, the slice requires a resource reallocation to scale up its link bandwidth in the transport network.

2. Problem Statement and Research questions

Although the strength characteristics of network slicing, the slice orchestration required to achieve such benefits is a complex case. Some of these challenges have been investigated in this thesis. This work focuses on the *dynamic slicing allocation problem*. In particular, this work proposes the use of dynamic network slicing as a supporting technology for improving solutions for the service placement and migration problem in the context of user mobility support in 5G/edge computing infrastructures. The relevance of edge resource management and user mobility support, and network slice management in edge remain applicable from 5G to 6G and beyond. A simulation environment is also required to validate these proposed solutions. A brief discussion about these challenges and the research questions that are investigated in this thesis is presented below.

1. *Is dynamic slicing a suitable solution to deal with traffic variation triggered by mobile users in edge computing environments?* One of the main challenges in network virtualisation, known as Virtual Network Embedding Problem, is properly assigning the physical resources (node and links) to a virtual network. Furthermore, the slice instances may need to be reshaped to serve the new requirements due to spatial and temporal demand variations over time. However, resource reallocations introduce a trade-off between performance and computing overhead. Further research on the dynamic resource allocation of network slices in realistic user mobility scenarios is a relevant topic that needs to be addressed.
2. *What is the impact of resource sharing on the performance of network slicing for mobile users?* A network slice can be fully or partially isolated from other slice instances, with its resources either shared or dedicated. Sharing idle resources, especially in far-edge devices (e.g., vehicles), could enhance the pool of available resources for slice users. While several studies focus on isolation for security, this work instead emphasises improvements in resource utilisation.
3. *What is the impact of user mobility in network slicing management?* Mobile devices such smartphones, vehicles, and drones require continuous connectivity while

moving across different locations. In such scenarios, seamless mobility management strategies must be developed. These strategies include, among others, seamless handover, service migration, and dynamic resource allocation approaches to ensure load balancing. Although some studies evaluate the performance of network slicing, most do not consider user mobility.

4. *Can dynamic network slicing improve the performance of the service placement and migration process?* The placement of network functions (NFs) or services plays a strategic role in meeting latency requirements and ensuring load balancing. Furthermore, due to demand variations caused by factors such as user mobility, the placement of NFs may need to be reconfigured to achieve a required performance. In such scenarios, NFs and services are continuously reallocated to different nodes, a process known as service migration. Dynamic bandwidth slicing might enhance the efficiency of the service migration process.
5. *Can dynamic slicing be applied in vehicular-cloud environments?* Intelligent Transportation Systems (ITS) can take advantage of connected vehicles to increase traffic rate and safety and decrease energy consumption and pollution on the roads. Once vehicles' idle computing power could be shared among nearby vehicles, that pool of resources could form a Vehicular Cloud. Although network slicing has been evaluated in different contexts, there are not many works evaluating how network slice handles such dynamic scenarios;
6. *What is the required time to perform network slice operations such as resource allocation and slice reconfigurations and how it impacts the network performance?* In network resource management, resource reallocations incur computational and time costs, which directly affect the performance of solutions. In this context, there is a trade-off between the reallocation cost and the benefits of the new distribution. Additionally, the time between the slice reconfiguration request and the moment the slice becomes operational should be considered in performance evaluations. A model that estimates such reconfiguration delays and evaluates the impact of this overhead on slicing performance is not yet available in the literature;
7. *Is there a validation environment to simulate network resource management scenarios which includes user mobility, network slicing and service migration?* Although the advantages of the real testbed, economic limitations are presented as one of the main restrictions in adopting that environment. Based on that context, simulation and emulation tools raise as a feasible environment to validate these scenarios. A simulator that offer such features is not available in the literature.

The scope of this thesis includes all the cited challenges described above. The solutions required for this context need to simultaneously choose the edge nodes to place the users' services considering latency, throughput, and load balancing aspects while deciding when to perform service migrations due to demand variations caused by mobile users. In such a context, network slicing is used as a supporting technology for those migrations by improving the bandwidth availability. The solutions for the traffic variation should decide the shape of those slices, *i.e.*, which nodes and links compose each slice, as well as the amount of resources allocated, while deciding the frequency in which one slice will be reshaped. The contributions of this thesis are discussed below.

3. Contributions

In the face of the challenges present in the 5G infrastructures introduced above, this thesis evaluates the performance of dynamic network slicing and develops new mechanisms to support dynamic network slicing as a solution for high dynamic scenarios, in particular, for user mobility support. This work proposes solutions combining dynamic slicing allocation, service placement, and migration to improve mobile users' QoS. To achieve the main objective, five minor objectives combined compose this thesis. In summary, this work presented:

1. One heuristic solution which uses dynamic network slicing as a supporting technology to improve the service migration process;
2. One study of the dynamic network slicing as a supporting technology for user mobility support in edge computing infrastructures and vehicular networks;
3. A mathematical model which describes the overhead/expected time to allocate and reconfigure the network slices;
4. Resource management mechanisms for dynamic slicing allocation, which include solutions for defining the slice shaping, resource allocation and reallocation, and resource sharing scope;
5. One simulator for slicing validation in edge computing-based infrastructures, including support for user mobility and service migration.

4. MobFogSim: Simulation of Mobility and Migration for edge/fog Computing

One fundamental step in developing this thesis is using one reliable environment to validate the solutions for resource management developed in this work. Facing economic limitations in adopting the testbed approach, one simulator, named MobFogSim[J1[1] in table 1], has been developed for that purpose. Network simulators have been present as a suitable environment to validate resource management mechanisms in large-scale scenarios. In this context, some solutions have been proposed in recent works. However, none of them successfully address all scenarios covered by this thesis. In this context, we developed during this thesis a simulator which supports user mobility, network slicing, device-to-device communication, and service migration in the context of edge-to-cloud computing infrastructures.

4.1. Related work

This section aims to introduce the main network simulators related to edge computing environments available in the state of the art and present how MobFogSim is compared to them. In this context, among the 18 simulators considered as related work in this section, four characteristics were highlighted and discussed in this work: 1) Device mobility support, which is a common feature in 15 of 18 simulators; 2) device-to-device communication, which is offered in 7 simulators; 3) service migration, that is supported in 8 simulators; and 4) network slicing support, which seems in 4 simulators from that universe of 18 tools evaluated in this work. The list of simulators and their references are presented in the full version of the thesis. Besides the existence of some simulators with user mobility support, service migration support, device-to-device communication and network slicing support, MobFogSim is the only one that can combine these characteristics.

4.2. Contribution

The proposed simulator is unique in the state-of-art. MobFogSim presents several contributions to extending CloudSim and iFogSim [Gupta et al. 2017], two well-known and

[1] This article is a direct result of my thesis and the first authorship was shared due to the data collection carried out by Carlo.

well-validated simulators widely used for modelling and simulating cloud and edge computing networks. Figure 2 presents the logical architecture of MobFogSim and how it extends CloudSim and iFogSim. MobFogSim implements end-to-end network slicing in static and dynamic ways. Also, one realistic user mobility support, which enables the simulator to access real data from traffic datasets, was developed. Moreover, this work added the modelling of vehicular scenarios to the simulator, including the possibility for vehicles to behave as edge nodes that share their resources with nearby users.

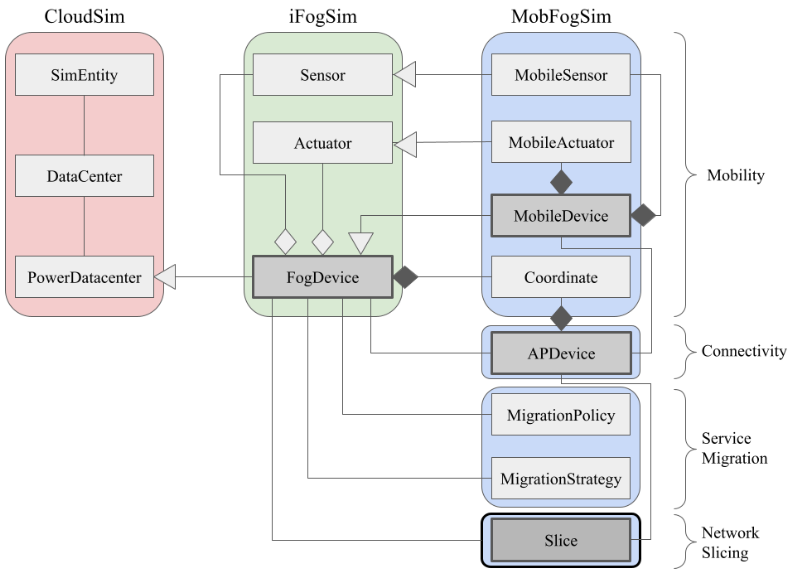

Figura 2. Architecture of MobFogSim, which extends Cloudsim and iFogSim.

All these enhancements make MobFogSim a complete tool for evaluating solutions regarding edge offloading, edge service migration, device mobility, and network slicing. Furthermore, the simulator was validated by comparing its results with testbed data. The experiments suggest that simulated and real results are equivalent. Further details are presented in Chapter 4 of the thesis. MobFogSim is available as Free Software.

Once the validation environment is developed and validated, further studies of network slicing performance can be carried out on MobFogSim. Some of those studies, which include the development and evaluation of resource policies for dynamic network slicing, are presented in the following section of this document. Further details can be seen in the full version of the thesis. Two journal papers[J1 and J2 in Table 1], and two conference papers[C2 and C4 in Table 1] present the features of MobFogSim.

5. Dynamic Network Slicing for Mobile Users

This work presents network slicing as a supporting technology for service migration in the edge of the network. In this context, network slicing is used to prioritize key groups of users during the migration process. Furthermore, dynamic network slicing improves resource management. Two allocation approaches are considered, static and dynamic, and different scenarios were evaluated in terms of available and demanded resources.

5.1. Related work

This section introduces the main related work in the state of the art. Several works propose solutions based on Network Slicing for different problems related to resource management in network infrastructures. In the literature, some works propose network slicing resource management mechanisms that consider user mobility support in their solution. Researchers also study services, VNFs, and SFC placement and migration. However, most of those works do not assume any values for the computing overhead for the resource allocation nor evaluate the impact of that overhead in their scenarios. On the other hand, some works focus on evaluating the overhead of network operations in the context of network slicing but do not consider user mobility, service placement or migration in that context. Only a restricted group of works consider overhead costs of the slice operation in a user mobility scenario.

This work is one of the few works in the literature that evaluates the *overhead of dynamic slicing allocation due to traffic variations in user mobility scenarios*. To the best of our knowledge, this work is the only one that proposes dynamic bandwidth slicing allocation to improve service migration. Although the existence of works considering the impact of slice reconfiguration, the metric assumed as overhead is mostly related to monetary or computing costs. This work is the only one that considers the impact of the processing delay in committing one slice reconfiguration. Also, one mathematical model that describes the expected time to proceed with one slice reconfiguration is proposed.

5.2. Proposed approach

This thesis investigates the *dynamic slicing allocation problem*, focusing on a support traffic variation triggered by user mobility scenarios. Different factors can result in an increasing traffic variation. However, *this work assumes a service migration between two different physical nodes as a significant traffic variation event* in the evaluation scenario. In the assumed scenario, the network infrastructure adopts the *follow-me cloud concept*. Then, some services may be migrated from one source edge node to another to serve its mobile users as closely as possible. In this scenario, *a network slice can request a resource reallocation to scale up its link bandwidth to proceed with the migration process*.

In this scenario, once the NFV Management and Orchestration (MANO) system detects an increasing traffic variation, it can start the slice reallocation process. It is assumed that the first attempt to serve the slice is allocating resources not assigned to any other slice. If there are no available resources in the physical node, the MANO system can identify idle resources allocated in other slices and start the process of resource redistribution. In this case of underutilisation of network resources, idle resources are deallocated from other slices and assigned to that new one. Hence, idle bandwidth resources may be reassigned at runtime to other slices needing more bandwidth to meet users' requirements. It is worth noting that resource reallocation is not immediate, as it always incurs a reallocation time due to computing overhead.

5.3. Delay allocation model

Despite the benefits of dynamic network slicing, such as better resource utilisation and avoiding underutilisation and overcapacity scenarios, they come with one additional management burden: the significant increase in the complexity of the operations. In addition, slice modifications are not instantaneous, and resource allocations may need to wait for reassignments to occur. The impacts of overhead in slicing reallocations in terms of network performance and users' QoS have not been fully investigated in the literature.

This work presents a mathematical model that describes the expected time to deploy and modify a network slice instance in runtime. The model describes embedding costs required to allocate a functional slice instance, *e.g.*, VM and VNF allocation and boot-up, and reconfiguration costs related to runtime orchestration, *e.g.*, scale-up and scale-down. The delay allocation model proposed in this work is based on an adapted version of the slicing operations lifecycle defined by [Yilma et al. 2020]. Figure 3 illustrates the slicing deployment steps adapted for the scenario assumed in this work. The proposed model explicitly describes the Onboarding Process Delay (OPD) and the Deployment Process Delay (DPD). Further details regarding the proposed model are presented in the full version of this PhD thesis and in the published article[J3 in table 1].

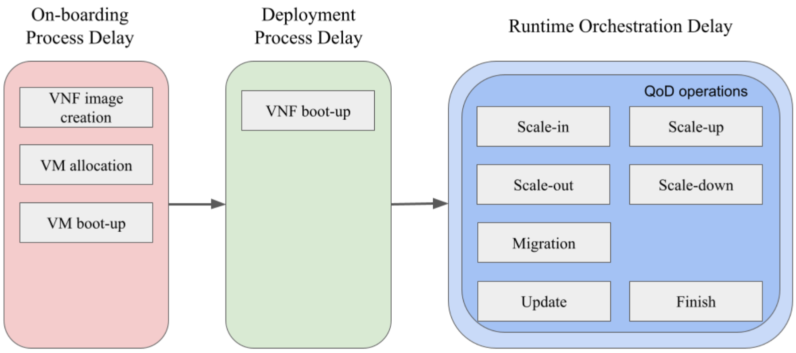

Figura 3. Slicing Operations lifecycle. Adapted from [Yilma et al. 2020]

Based on values provided by the proposed mathematical model, some experiments present how the reconfiguration overhead on network slicing impacts the network performance. In this work, the slice reconfiguration is based on dynamic bandwidth demands on the transport network triggered by service migrations.

5.4. Validation setup

In this experiment, as the demand for each network slice changes over time, the NFV Management and Orchestration (MANO) system can successfully change the resources assigned for the slices to support that demand variation. In this scenario, only idle resources can be reallocated to the other slice if that slice requires it. The required time to process that operation is assumed to be 2, 30, and 60 seconds. The simulation setup is defined as follows: two slices, named Slice 1 and Slice 2, receive 50% of the bandwidth available in the physical links at the beginning of the simulation. Three different levels of user demand were evaluated. Slice 1 was evaluated, serving 10%, 20%, and 40% of the 80 users from the simulation. Slice 2 served 90%, 80%, and 60%.

In this experiment, 16 edge nodes uniformly placed on a 5kmx5km map. 64 access points provide the connection between the users and the infrastructure. Based on the testbed setup defined in [J1 in Table 1], a 74 Mbps bandwidth and 3.47 ms latency link connect the edge nodes. 30 groups with 80 users (2400 in total) are considered in this work. Each simulation used one of these groups. The user mobility was based on data of vehicles from Luxembourg database [2]. The simulation time is limited to one hour.

5.5. Performance evaluation and the impact of allocation delay

The experiments introduced in this section present the impact of different delays in the reconfiguration process. The results are presented considering a 95% confidence interval based on the average of 30 simulation runs.

[2]Codeca, L., Frank, R., and Engel, T. (2015). Luxembourg SUMO Traffic (LuST) Scenario: 24 hours of mobility for vehicular networking research. In IEEE Conference on Vehicular Networking (VNC).

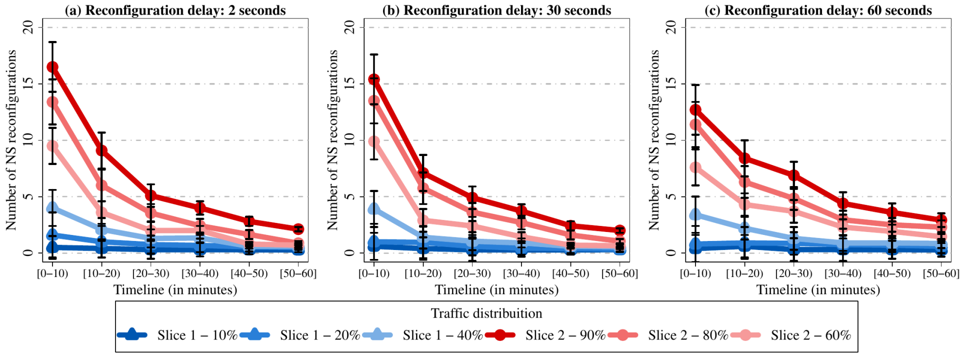

Figura 4. Timeline of bandwidth reallocations grouped by reallocation delay.

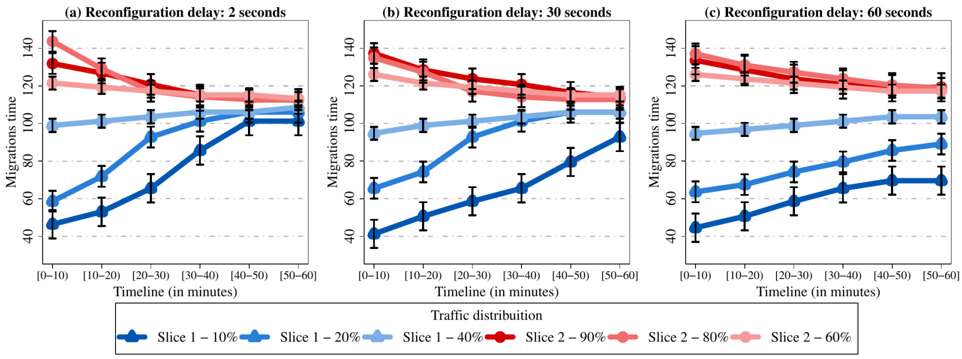

Figura 5. Timeline of service migration time.

Figure 4 presents the number of reconfiguration requests by each slice granted by the infrastructure. It is assumed that each slice in the simulation received 50% of the physical resources at the beginning of the simulation. Despite the equal resource distribution, different workload levels were assigned to the slice. The results are grouped by the different values of reconfiguration delay adopted in the simulation. Figure 4 (a) presents the number of reconfigurations considering 2 seconds between the reconfiguration being requested by the slice and the reallocation being completed. Figure 4 (b) considers 30 seconds and Figure 4 (c) assumes 60 seconds.

The results show that the lower the reconfiguration delay, the higher the number of reconfigurations the infrastructure can provide along the slice life cycle. The number of reconfigurations can differ by up to 40% if compared to 2 and 60 seconds as values for the reconfiguration delays. The simulations point out that, considering a reconfiguration delay of 2 seconds, Slice 2 with 90% of the users had, on average, 24.5 accepted requests in the first 10 minutes of simulations. In the same scenario, but considering 60 seconds as reconfiguration delay, that number falls 49%, going to 12.7 at the same period in the simulation. These results show that every evaluated workload scenario presents a reduction in reconfigurations as the reconfiguration delay increases. Furthermore, the number of reconfigurations tends to fall and be more stable as the timeline of the slice lifecycle goes to its end in a scenario with the lowest reconfiguration delay.

The reallocation delay impacts the resource management performance and might lead the slices to present resource overutilisation and underutilisation. It may impact the performance of services. Figure 5 presents the implications of the dynamic bandwidth reallocation on the service migration process.

In slices with low reconfiguration delay, as presented in Figure 5 (a), the average duration of service migrations quickly converges to 116 seconds for all slices. That tendency is slower in the remanding scenarios. In the scenario with 30 seconds or more as reconfiguration delay (Figure 5 (b) and (c), the slices do not even converge their resources to a stable level. It impacts the duration of their migrations. In those scenarios, after 60 minutes of simulations, the performance of the service migrations in their slices is still not balanced. In such scenarios, some slices perform service migrations better than the average performance, *e.g.*, slices with fewer users (Slice 1), while the more significant number of users perform worse than the average (Slice 2).

As identified in Figure 5, the slices' reconfiguration delay impacts the performance of some infrastructure processes, *e.g.*, service migrations. As the infrastructure does not provide the proper number of migrations, it may leave the services in edge nodes that may no longer fulfil their service requirements. In that scenario, the infrastructure may present difficulty in providing seamless mobility management strategies that face demand variations. This experiment evaluating the impact of different values on the slice reconfiguration delay indicates that lower values lead the slices to better and more stable resource distributions. Consequently, those slices provide a better environment to serve users' service requirements.

6. Scientific Production, Awards, and Grants

The contributions of this work are focused, but not limited to, on proposing solutions for dynamic network slicing to support user mobility scenarios in fog/edge infrastructures. During the development of this thesis, the proposed content, which includes mathematical models, simulation scenarios, algorithms, and software, was peer-reviewed and published in relevant journals and conferences. The contribution of this thesis generated, as first author[3], three journal papers, four conference papers, and a tutorial[4]. In addition, a book chapter was produced as co-author. Recently, this work was selected for presentation in the Dissertation Digest Session of the IEEE/IFIP NOMS 2025[5]. Further details regarding the publications can be found in complementary files submitted with this paper and in the full and final version of this PhD thesis, available online [Gonçalves 2024].

In addition, the development of this thesis presents as an outcome a free software simulator named MobFogSim[6] which have been available for community use. Currently, to the best of our knowledge, **MobFogSim has more than 150 citations, and at least 27 of those works report that simulator as its validation environment**. It includes 21 conference and journal papers, 2 undergraduate theses, 2 master's and, 2 PhD theses. The list of works which use MobFogSim is available in the full version of this PhD thesis.

This research was awarded a grant from the Emerging Leaders in the Americas Program issued by the Government of Canada for six months of research at the University of Waterloo, Canada. Also, this research was awarded a grant from the UNICAMP's Teaching, Research, and Extension Support Fund (FAEPEX) and INCT Future Internet for Smart Cities (InterSCity) for three months of research at Cardiff University, UK.

[3]Diogo and Carlo share the first author role in [J1 in Table 1], as highlighted in that paper. Diogo's contributions: the design and implementation of MobFogSim, the simulation design and analysis of its results. Carlo's contributions: the design and implementation of the testbed, data collection and analysis of its results. Both authors contributed to writing and revising the paper.

[4]https://cloudnet2024.ieee-cloudnet.org/tutorials

[5]https://noms2025.ieee-noms.org/call-dissertation-digest

[6]https://github.com/diogomg/MobFogSim. Last accessed: February 10th, 2025

Tabela 1. List of publications with quality indicators, including Qualis classification, impact factor, H5-index, and the number of citations in Google Scholas(GS).

Cat.	Ref	Title	Qualis	IF	H5	Citations GS
Journal	J1	MobFogSim: Simulation of mobility and migration for fog computing. **Simulation Modelling Practice and Theory**, 2020.	A1	3.5	83	160
	J2	End-to-end network slicing in vehicular clouds using the MobFogSim simulator. **Ad Hoc Networks**, 2023.	A1	4.4	110	8
	J3	Overhead and performance of dynamic network slice allocation for mobile users. **Future Generation Computer Systems**, 2024.	A1	6.2	164	3
Conference	C1	Análise da predição de mobilidade na migração de aplicações em computação em névoa. **SBRC**, 2019.	B2		8	4
	C2	Dynamic network slicing in fog computing for mobile users in MobFogSim. **IEEE/ACM UCC**, 2021.	A3		13	17
	C3	Fatiamento dinâmico de redes em computação em névoa para usuários móveis. **SBRC**, 2021.	B2		8	2
	C4	Alocação de fatias de rede fim-a-fim para usuários móveis utilizando o simulador MobFogSim. **SBRC**, 2022.	B2		8	1
			Total number of citations			195
Tutorial	T1	MobFogSim: Simulation of Mobile Applications in Edge Computing Environments. **IEEE CloudNet**, 2024	A2		12	
Dissertation Digest	DD1	Dynamic Network Slicing for User Mobity Support in 5G networks (to be presented). **IEEE/IFIP NOMS**, 2025	A2		22	

7. Conclusion and Future Directions

In conclusion, this thesis provides as legacy two main contributions: 1) It provides a comprehensive analysis of the strengths and weaknesses of static and dynamic resource allocation approaches for network slicing. The contribution includes the development of an allocation heuristic, a mathematical model for estimating allocation delays, and extensive performance evaluations across various scenarios which focus on user mobility support, and 2) It developed an open-source simulator as a validation environment for the evaluated scenarios. The simulator offers for the research community an alternative to the challenges of building real testbed infrastructures. This tool has already seen positive adoption. These contributions open multiple directions for further research in this topic. In particular, machine learning-based management and orchestration systems can enhance the efficiency and automation of network operations throughout the slice life cycle. The findings of this thesis can serve as valuable input for fine-tuning such ML models.

Acknowledgements

This work is part of the INCT of the Future Internet for Smart Cities (CNPq 465446/2014-0, CAPES 88887.136422/2017-00, and FAPESP 2014/50937-1). This work was funded by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, grant 2021/05253-0 by São Paulo Research Foundation (FAPESP), and grant 382766/2022-9 by CNPq.

Referências

Gonçalves, D. M. (2024). Dynamic network slicing for user mobility support in 5g networks. *Universidade Estadual de Campinas. Phd Thesis.*

Gupta, H., Vahid Dastjerdi, A., Ghosh, S. K., and Buyya, R. (2017). iFogSim: A toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Software: Practice and Experience*, 47(9):1275–1296.

Yilma, G. M., Yousaf, Z. F., Sciancalepore, V., and Costa-Perez, X. (2020). Benchmarking open source NFV MANO systems: OSM and ONAP. *Computer communications*, 161:86–98.