# Information Exchange to Support Multi-Domain Slice Service Provision for 5G/NFV

Alberto Solano
Transport & IP Networks
Telefónica I+D / Global CTIO Unit
Madrid, Spain

Luis M. Contreras
Transport & IP Networks
Telefónica I+D / Global CTIO Unit
Madrid, Spain

***Abstract*—Service providers will highly rely on network softwarization operational trends for addressing 5G verticals' market, leveraging on multi-domain cloud computing capabilities offered by different stakeholders. Consequently, multi-domain scenarios in which diverse providers form a federation will become more relevant for 5G and beyond. Due to the intrinsic limited footprint and infrastructure capacity offered by just a single provider, the progression towards these scenarios will require mechanisms to support an efficient orchestration of network slices over multi-provider assets. In this paper, we analyse the convenience of enabling timely exchange of information in terms of guaranteed latency and resource availability among various telecommunication providers participating in a federation to effectively ensure the deployment of 5G vertical services and reduce the service provision blocking probability.**

***Keywords*—5G Vertical services; Federation; Network Slicing; Multi-Domain Service Provision.**

## I. INTRODUCTION

Vertical customers are expected to fully exploit the capabilities of 5G networks through a more effective usage of the telecommunication providers' networks. The current approach to the exploitation of 5G networks consists in the dynamic instantiation of services that are logically composed by network functions, as enabled by network programmability and virtualization [1][2]. This is even fostered by the interconnection of 5G non-public or private networks, owned by the 5G verticals, to public networks, as the ones from operators [3][4]. The connectivity of those network functions, even spanning more than one administrative domain, provides a final communication service that is of interest for the vertical customer (e.g., media entertainment, driving assistance, eIndustry, etc.).

The deployment of network functions has been traditionally performed by means of monolithic boxes located at strategic Points of Presence (PoPs) within the telecommunication provider. Consequently, network functions have been tightly coupled to the telecommunication provider network topology, resulting in rigid service implementations. This fact has limited the fast introduction and delivery of new services due to the uncertainty about the future success of service offerings, as well as the evolution and scale of existing ones due to the static composition of services.

Vertical customers will require dynamic deployments with minimal operational impacts. Usually, vertical customers will follow a single contractor approach for serving their telecommunication needs, which basically means that the interaction will be direct with a unique service provider, acting as single-entry point for a service request. With such an approach, the responsibility for the entire provision of the service in the footprints and locations where the final service is intended to be provided out of the private is left to such single contractor. Consequently, the entire responsibility related to managing the operation and maintenance of functions composing the end-to-end service as well as the underlying infrastructure is generally outsourced to the telecommunication provider (in the case the vertical operates its own private network there could be exceptions to this respect). Note that in this paper 5G vertical service and network slice are treated as equivalent terms.

Commonly, because of global telecommunication market fragmentation at local, regional and international levels, there will probably be a mismatch between the vertical and the telecommunication provider footprints. This forces the single contractor to reach agreements with other providers (e.g., municipalities, cloud providers, other operators, etc.) for complementing its offerings, resulting in the need of performing multi-domain orchestration of services. As a result, it will be necessary to implement intelligent means for deciding which of the network functions composing a service can be deployed within the domain of the single contractor, and what others can be implemented in other administrative domains. The motivations for such kind of decisions can be multiple:

- Limitations on the coverage provided by the telecommunication provider acting as contractor: for example, the previously mentioned footprint mismatch.
- Geographic restrictions that could impose the need of deploying functions in different domains: for instance, the need of ensuring some guaranteed latency.
- Geolocation aspects that could force the deployment of certain functions in certain domains. A case for this could be the restrictions for the diffusion of a given content because of property rights or privacy protection for data requiring to be stored in a given location.
- Optimization and efficiency in the overall usage of resources for a given service. For example, the minimization of bandwidth consumption by placing some content on infrastructures of third parties closer to the vertical premises.
- Specialization in the provision of certain functions: for instance, a given function particular to a specific provider, or the provision of Internet transit by a given carrier.

Accordingly, multi-domain orchestration is becoming an increasingly relevant topic in the industry [5][6]. Several technical implications are arising, requiring the need of defining highly interoperable solutions in order to reduce cost

of integration and ensure interoperability in these multi-provider scenarios. This problem is related with the distributed virtual network mapping problem across multiple provider domains (see the references to this aspect included in the survey in [7]).

This paper focuses on one of those technical implications, aiming at analysing the benefits of providing an exchange of information in terms of guaranteed latency and resource availability in different data centers in the various telecommunication providers participating in a federation. The objective of this exchange of information is to improve the ratio of finally deployed services at the cost of revealing some internal details of the providers participating in the federation, such as achievable latency and available resources. The main contributions of the paper are the analysis and simulation of the blocked number of services in a federation of providers with and without information exchange, and the proposition of how to populate such information among domains. Section II describes a 5G/NFV service provider environment supporting multi-domain orchestration at the service level. Section III details the information that is proposed to be shared between telecommunication providers participating in a federation for an effective end-to-end deployment of services. Section IV presents a case analysis by simulating the service blocking rate, depending on the level of information available between the administrative domains for the placement of network functions. Finally, Section V provides some concluding remarks and future lines of work.

## II. 5G/NFV Multi-Domain Service Provison

5G/NFV is expected to create new opportunities for the multi-domain service provision provide through the use of network slicing, orchestration and federation [7].

Multi-domain network Slicing aims at accommodating different services over a multi-provider network infrastructure. From the 5G/NFV perspective, this infrastructure can be perceived as a collection of virtual resources that are used to deploy services. Thanks to the flexibility introduced by NFV, services can be divided into several Virtual Network Functions (VNFs) that can be deployed in a convenient manner throughout the collection of resources available, forming a network slice. These resources can be located in two different types of administrative domains:

- *Local domain*: Also known as the entry-domain. It refers to the telecommunication provider acting as single contractor of the vertical customer for its service deployment.
- *Overflow domains*: These domains complement the offerings of the local domain by federating to it. These domains will host the VNFs that are not deployed in the local domain due e.g. to the limited infrastructure that it owns or to any other constraint.

Service orchestration capabilities in the local domain coordinate the provision of services across domains by creating, managing and deploying network slices by considering the overall virtual resources in the federation.

When orchestrating services, each VNF can be considered as an atomic unit. That is, each VNF is taken as an integral set of resources and connectivity assets. When federating a service, defined as a composition of network functions, a partitioning of a network-graph is performed by assigning different service VNFs to distinct administrative domains, and allocating their supportive set of resources in a given datacentre. In order to keep VNF's atomicity, the set of resources for a given VNF cannot be partitioned but be allocated in the same datacentre.

## III. Information Exchange For A Multi-Domain Scenario

This section deals with the exchange of information among administrative domains to produce better informed decisions at the time of allocating network functions across domains.

### A. Population of available resources and latency per domain

In a multi-domain scenario the local domain will interact with several different overflow domains to provide an efficient orchestration of network slices and to overcome the mismatch between 5G verticals' service expectations and entry provider owned resources.

In the approach here proposed, each overflow domain participating in the federation will advertise information about resources and latency towards the other participants in the federation. Thanks to populating such information, the local domain can decide which VNFs can be allocated to any of the overflow domains, if needed. However, if the local domain has not a priori information about guaranteed latency and resource availability, it could try to allocate a VNF to an overflow domain whose data centers do not fully comply with the VNF latency constrains or do not have enough available resources to deploy the VNF.

In case there is no exchange of information among administrative domains, the local domain can only rely on its own data centers and assume that a best effort approach will be taken by the overflow domains when orchestrating a multi-domain network slice. That is, the VNFs that cannot be directly allocated in the local domain are passed over some overflow domain with the expectation of being deployed in such other domain. Eq. (1) describes the accounting of available resources at the time $t$ when no information exchange is in place, which basically corresponds to the available resources in the local domain.

$$R_t = \sum_{n=1}^{N} \left(C_{DC_n} - U_{t-1}\right), U_{t-1} < C_{DC_n} \quad (1)$$

Where

- $R_t \in \mathbb{R}^+$ represents the available resources for federating allocating VNFs at the time $t \in \mathbb{R}^+$ without communication support
- $C_{DC_n} \in \mathbb{R}^+$ corresponds to the maximum capacity of the local domain for $n \in \mathbb{N}$ of $N \in \mathbb{N}$ data centers that complies with the latency constraints of the service
- $U_{t-1} \in \mathbb{R}^+$ are the utilized resources at the time $(t - 1) \in \mathbb{N}$.

In this situation, the overflow domains may deploy or not the allocated VNFs depending on their resources availability and the latency that they can provide, but that fact is unknown at the time of request from the local domain because of the

lack of information. This case is represented by $U_{t-1} \geq C_{DC_n}$, i.e., there are no more resources available in the local domain.

When introducing the exchange of information in terms of guaranteed latency and resource availability that the overflow domains can ensure, the local domain can perform informed decisions when passing over VNFs to other domains. The available resources in this case increases, as reflected in Eq. (2), because of the knowledge of the available resources of the overflow domains, which can be considered by the local domain on its decision process.

$$R_t = \begin{cases} \sum_{k=1}^{K} \sum_{n=1}^{N_k} C_{DC_n}^k - R_{t-1}^k, U_{t-1}^k < C_{DC_n}^k \\ 0, \quad U_{t-1}^k \geq C_{DC_n}^k \end{cases} \quad (2)$$

Where $K$ overflow domains containing $N_k$ datacenters that complies with the latency constraints of the service are taken into consideration.

When there are no more resources available in the local or overflow domains, i.e., $U_{t-1}^k \geq C_{DC_n}^k$, then the system cannot serve the request then blocking service provision.

### B. BGP protocol as support for the information exchange

The interconnection of different networks, each of them representing a different administrative domain (i.e., distinct Autonomous Systems managed by different providers), is commonly performed by leveraging on the Board Gateway Protocol (BGP) protocol [9].

BGP has a huge capability to scale and has undoubtedly contributed to the success of the Internet by facilitating a standardized manner of interconnecting networks. However, the network slice concept will intensively stress the network capabilities because of the dynamicity introduced in the deployment and lifetime of services [10]. This can be expected also to occur in the multi-domain environment. As a result BGP must count with new mechanisms to support them, together with some other aspects identified as evolutionary paths of the protocol [11].

Board Gateway Protocol Link-State (BGP-LS) [12] is the mechanism by which Interior Gateway Protocol (IGP) link-state and traffic-engineering information is collected from the local domain and is shared with other domains using the BGP protocol. This is achieved by using a new Network Layer Reachability Information (NLRI) encoding format. NLRIs are used to advertise link, node, and prefix information in the form of parameters and attributes. It is defined as a set of Type/Length/Value (TLV) fields.

This paper proposes to define also as TLVs fields information related to guaranteed latency supported and resource availability, advertising that between the various telecommunication providers participating in a federation. This same approach has been taken in [13] to populate traffic engineering performance metrics such as link bandwidth or delay.

Table I presents the proposed values for the parameters to be exchanged between domains in a federation. The proposed parameters can be encoded as TLV fields and sent within each Link-State NLRI updates. As a result, and thanks to the BGP-LS protocol, all the telecommunication providers participating in the federation can receive up-to-date information about the latency and availability that the datacentre of other providers in the federation can provide. This information is crucial to decide where to deploy the VNFs outside the local domain, avoiding blocked-service provision due to not fulfilling the latency contains or not having enough vacancy in the data centers of an overflow provider.

TABLE 1: PROPOSED VALUES

| | Type | Length | Comment | Reference value |
|---|---|---|---|---|
| **Guaranteed Latency** | int | 24 bits | Max measured link delay value (in ms) over a configurable interval | Max Unidirectional Link Delay [13][14] |
| **CPUs availability** | int | 32 bits | Number of available virtual CPUs | vmCurCpuNumber [15] |
| **RAM availability** | int | 32 bits | Memory size | vmCurMem [15] |
| **HDD Disk availability** | int | 32 bits | Virtual storage size | vmStorageAllocatedSize [15] |

The following section proposes a case analysis on the service blocking probability that assess the previous equations and studies, with simulated data, how often a service is blocked when supporting or not the exchange of communication.

## IV. CASE ANALYSIS: SERVICE BLOCKING PROBABILITY

This section describes a case analysis on service blocking probability with the aim of assessing the benefit of introducing the exchange of information between administrative domains for the provision of network slices.

### A. System structure definition

In order to set a simulation scenario for the case analysis, a system structure is first proposed. This system structure represents the interaction between the multiple stakeholders that interact with each other to provide services to tenants. It is composed by three modules that oversee different functions in the case analysis, as indicated in Figure 1.

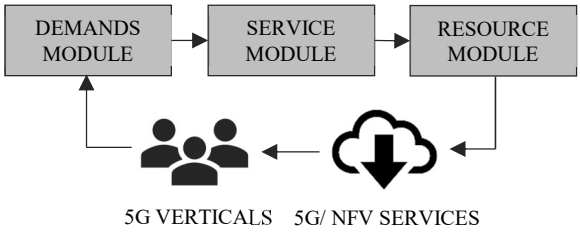


Figure 1: System modules

The *Demands Module* generates the service demands (i.e., the vertical network slices) and takes care of the service characterization. The *Service Module* performs the partitioning of the network-graph for a given service by assigning VNFs to an overflow domain complying with the latency constrains, when no resources are left in the local domain. In addition, this module calculates the resources that are required from each domain to make the network slice deployment possible. Finally, the *Resource Module* takes care of allocating the required resources to the data centers of each domain.

*B. Case analysis scenario stakeholders definition*

One possible scenario has been defined for illustration purposes. The case analysis scenario gathers three types of inputs: services or network slices (as requested by 5G verticals), data centers (for resources enabling the deployment of VNFs) and administrative domains (for telecommunication providers).

Services are defined by describing the VNFs in terms of computing capacity (CPU), memory (RAM), and storage (HDD) resources, as well as bandwidth and latency needs. In order to characterize the system behaviour, three different kind of services are considered, as indicated in Table 2.

TABLE 2: NETWORK SLICE SERVICE PARAMETRIZATION

| | CPUs | RAM [GB] | Disk [GB] | BW [Gbps] | Latency [ms] |
|---|---|---|---|---|---|
| **Service 1** | C | R | D | B | L |
| **Service 2** | 5C | 5R | 5D | 5B | 5L |
| **Service 3** | 10C | 10R | 10D | 10B | 10L |

Data centers are defined according to their resource capacity in terms of CPUs, RAM, storage (HDD), bandwidth capacity and their guaranteed latency. Three data centers sizes are considered: small, medium and large. The three of them are defined as a function of the Service 1 parameters defined in Table 2.

Concerning the number of resources allocated to each of the data center types, Table 3 describes the characterization of each of them.

TABLE 3: DATA CENTER CHARACTERIZATION

| | CPUs | RAM [GB] | Disk [GB] | BW [Gbps] | Latency [ms] |
|---|---|---|---|---|---|
| **Small** | 25C | 25R | 25D | 25B | L |
| **Medium** | 50C | 50R | 50D | 50B | 2L |
| **Large** | 300C | 300R | 300D | 300B | 5L |

Concerning the number of domains in the federation, three of them are considered: the local domain plus two overflow domains. When deploying a service in an overflow domain, a latency increment of $L$ ms is added to account the physical latency between domains, which has some implications in the consideration of data centres for the allocation of the VNFs, attending to the characterization in Table 2.

- *Small datacentre*. When federating a service in a small datacentre, the minimum latency that can be ensured is $L + L = 2L$. Therefore, Service 2 ($2L \leq 5L$) and Service 3 ($2L \leq 10L$) can be allocated to these data centers.
- *Medium datacentre*: When federating a service in a medium datacentre, the minimum latency that can be ensured is $2L + L = 3L$. Therefore, Service 2 ($3L \leq 5L$) and Service 3 ($3L \leq 10L$) can be allocated to one of these data centers, as well.
- *Large datacentre*. When federating a service in a large datacentre, the minimum latency that can be ensured is $5L + L = 6L$. Therefore, only Service 3 ($6L \leq 10L$) can be allocated to a large datacentre.

Accordingly, Service 1 can only be deployed in the local domain and this domain must have at least one small datacentre. Service 2 can be deployed in the local domain or in any overflow domain containing a small or medium datacentre. Ultimately, Service 3 can be deployed to any datacentre in any domain.

Finally, Table 4 presents the distribution of data centers considered in the analysis across the three proposed domains. The reduced number of data centers in Table 4 has been selected to force a service-blocking situation for both the local domain and the overflow domains, in order to compare the effectiveness of the information exchange. This service-blocking occurs when a data center is fully occupied and cannot deploy any further VNF, then forcing the service (i.e., the network slice) to be blocked.

TABLE 4: NUMBER OF DATA CENTERS PER DOMAIN

| | Small | Medium | Large |
|---|---|---|---|
| **Local domain** | 3 | 0 | 0 |
| **Overflow domain 1** | 0 | 1 | 0 |
| **Overflow domain 2** | 0 | 0 | 1 |

For all the domains, the allocation of VNFs to their data centers depends on three rules:

- *First rule*: On the assumption that the local domain complies with the latency and resource availability constrains, the VNFs are allocated to the local domain. The overflow domains remain unused.
- *Second rule*: When the first rule is not fulfilled, VNFs will be assigned to one of the overflow domains. Here we observe two possibilities depending on the use or not of the exchange of information proposed in this paper.
    - *Without information exchange*: In this case the local domain is not aware of the state of the overflow domains and doesn't know in advance whether the VNFs that are intending to be federated could be deployed. One overflow domain will be randomly selected.
    - *With information exchange*: In this case the local domain knows beforehand whether the VNFs that are intending to be federated could be deployed and exactly where can be deployed in the case it is possible.
- *Third rule*: In any case, either in local or overflow domain, the resources of a given data center are totally consumed before attempting to deploy VNFs in another datacentre of the same kind.

After applying the three rules, if it is not possible to deploy all the VNFs of a network slice, the service is reported as blocked. As described only technical constraints have been taken into consideration on the deployment decision (resource availability, performance, etc). Other constraints such as the deployment cost could be included, but are not in the scope of the paper.

*C. 5G Vertical service demand and lifetime definition*

The definition of the 5G vertical service demand and duration is done in accordance with the values in Table 5.

There, the arrival rate and the lifetime for each of the services defined in Table 5 can be found.

TABLE 5: SERVICE ARRIVAL RATE AND DURATION

| | Arrival rate | Minimum lifetime [H] | Maximum lifetime [H] | Mean lifetime [H] |
|---|---|---|---|---|
| **Service 1** | 10 | 96 | 240 | 168 |
| **Service 2** | 25 | 18 | 30 | 24 |
| **Service 3** | 50 | 0.5 | 4.5 | 2.5 |

Network slice request arrivals are characterized by a Poisson distribution. This probability distribution gives the probability of occurrence of events $P(t;\lambda)$ in a fixed interval of time $T$ (one year in this analysis) provided that these events occur with a known constant rate $\lambda$ and with independency of the time $t$ since the last event occurred, as indicated by Eq. (3) :

$$P(t;\lambda) = \frac{e^{-\lambda}\cdot\lambda^{t}}{t!} \tag{3}$$

*w*here

- $\lambda \in \mathbb{R}^{+}$ is the frequency of service requests based on the type of service as proposed in Table 5.
- $t \in \mathbb{R}^{+}$is the time of arrival.

Service lifetime is characterized by a truncated Gaussian distribution [16] to ensure that the lifetime values are distributed between the minimum and maximum lifetime values indicated in Table 5 for each of the services. It is characterized in Eq. (4) [17] with mean $\mu$ and standard deviation $\sigma^2$ that lies within the interval (a, b), with $-\infty < a \leq b < \infty$.

$$P(x;\mu,\sigma,a,b) = \begin{cases} \dfrac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right)} & a \leq x \leq b \\ 0 \quad Otherwise \end{cases} \tag{4}$$

where

- $\mu \in \mathbb{R}^{+}$ is the mean which corresponds to the mean slice lifetime obtained from Table 5
- $\sigma^2 \in \mathbb{R}^{+}$ is the standard deviation which is set to 1 in order to ensure a small deviation from the mean.
- $\phi(\xi) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}\xi^2\right)}$ is the probability density function of the Gaussian distribution.
- $\Phi(\xi) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\xi}{\sqrt{2}}\right)\right)$ is the probability density function of the cumulative Gaussian distribution.

### *D. 5G/NFV Multi-Domain Service Provision analysis*

The scenario setting described in the previous sections has served as baseline for running a simulation of the federation behaviour. This scenario tries to be as generic as possible, including for such a purpose three services that represent different resources' load as well as lifetime and arrival rate.. The simulation has been developed in MATLAB, by considering 10,900 events (i.e., service or network slice requests) generated during a one-year timeframe. Of those events, 1,605 belong to Service 1, 3,595 to Service 2 and 5,700 to Service 3.

As it was stated in Section II, VNFs are considered an atomic unit. When a VNF of a service cannot be deployed, the entire service is not deployed, and it is accounted as a blocked service. This situation can occur due to two possible situations:

1. *Latency constrains*: Depending on the latency that each of the datacentre types can guarantee, and the incremental latency of $L$ ms due to the fact of deploying the service in an overflow domain, the final latency could be greater than the minimum required latency for a given service. In this case, the service cannot be deployed.
2. *Occupation constrains*: When the previous constrains is overcome, the occupation constrains must be analysed. In order to deploy a VNF in a particular data center, it has to have enough resources for the deployment. Otherwise, the service is also blocked.

Figure 2 presents the cumulative blocked services for each of the service types in the case of exchanging and not exchanging latency and resource availability information among domains in the federation, during one year in the conditions described. Table 6 summarizes the absolute number of blocked services.

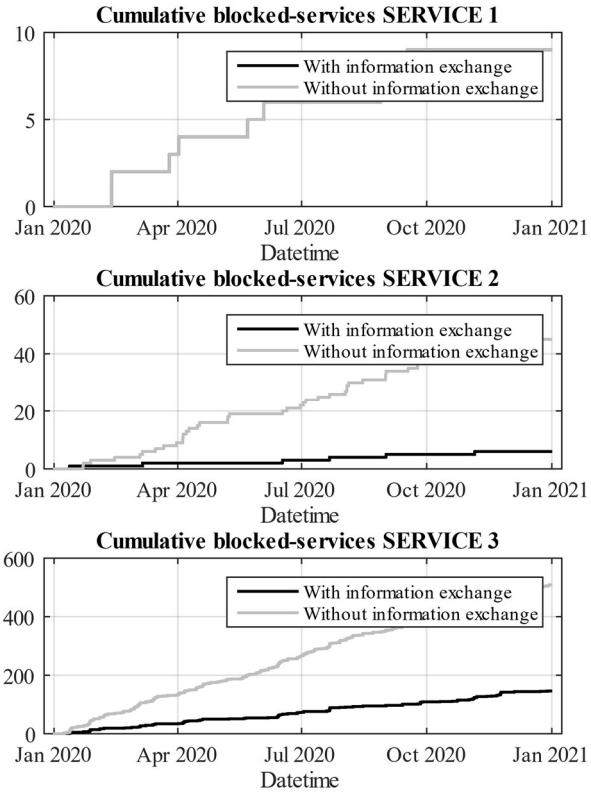


Figure 2: Cumulative blocked-services along simulated time per typoe of service

Because of the latency constrains defined for each kind of service, Service 1 cannot be essentially federated and as a consequence the results are the same in both cases. So the impacts of enabling information exchange among domains

can be perceived on the results obtained for Service 2 and 3. In this respect, the results obtained in Table 6 lead to the conclusion that the exchange of information in terms of guaranteed latency and resource availability highly decrease the ratio of blocked services. This is specially the case for sudden demands represented by very frequent services demanding a large number of resources during short term, such as Service 3.

TABLE 6: ACCUMULATED BLOCKED SERVICES

| | With information exchange | Without information exchange | Improvement in percentage |
|---|---|---|---|
| **Service 1** | 9 | 9 | 0% |
| **Service 2** | 6 | 45 | 86.66% |
| **Service 3** | 246 | 509 | 51.66% |

This reduction in the number of blocked services will have direct impact on the incomes of the single contractor of the vertical customer, but also for the federation itself, since non-blocked services, making use of resources from overload domains, also imply the generation of incomes for those domains, incomes that otherwise are lost. The reduction will depend on the characteristics of service, reaching in some cases an improvement of 86.66% in the best case.

## V. CONCLUDING REMARKS

Vertical industries are in the process of consume providers' network in the form of network slices in a more agile way with the introduction of 5G and the generalization of NFV and SDN as enablers of the network softwarization trend. Multi-domain scenarios, such as the federation of service providers will become common in order to accomplish such demands from 5G verticals. These multi-provider settings will benefit of the up-to-date interchange of information regarding the availability of resources and latency constrains for a proper partition of service graphs supporting the network slices.

A barrier for such interchange from a provider perspective is the exposure of internal information to other providers, which can be conflicting. Proper levels of abstraction can mitigate such situation. A potential approach could be to handle different kinds of abstractions at topological level as a function of the level of trustness among providers, associating the corresponding resources available. For instance, [18] defines white, black and grey topologies as abstract type levels of topological information exposure among providers, with higher or lower degree of exposition. Similar approach could be taken here, by associating resource information to those types of topologies. Since multiple providers will form such federations, it is mandatory to define standardized mechanisms for that in order to minimize integration costs and ensure interoperability. It is proposed to leverage on Link-State NLRIs updates sent over BGP-LS as a valid alternative for implementing such information interchange, due to the extensive usage of BGP in such interconnection settings (i.e., between Autonomous Systems).

This paper has presented a method for evaluating the service blocking of VNF chains for 5G vertical services in multiple provider scenarios. A tool has been developed for assisting on simulation of multi-domain scenarios. More realistic and detailed characterization of services, topologies and associated parameters is left for future work. Additional drivers, such as deployment costs will be included (e.g., following auction based allocation as in [19]). Further work will also be oriented towards the economic impacts due to the reduction of the service blocking thanks to the interchange of relevant resource and latency information, as well as the proposition in standardization fora of the identified mechanisms for populating this kind of information among providers in an abstract but yet meaningful way.

## ACKNOWLEDGEMENTS

This work has been partly funded by the project 5GROWTH (Grant Agreement no. 856709).

## REFERENCES

[1] ETSI, "Network Transformation; (Orchestration, Network and Service Management Framework)", 1st edition, October 2019. Available at: https://www.etsi.org/images/files/ETSIWhitePapers/ETSI_White_Paper_Network_Transformation_2019_N32.pdf

[2] 5G-PPP, "Vision on Software Networks and 5G", version 2.0, January 2017. Available at: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP_SoftNets_WG_whitepaper_v20.pdf

[3] 5G-ACIA, "5G Non-Public Networks for Industrial Scenarios", July 2019. Available at: https://www.5g-acia.org/fileadmin/5G-ACIA/Publikationen/5G-ACIA_White_Paper_5G_for_Non-Public_Networks_for_Industrial_Scenarios/WP_5G_NPN_2019_01.pdf

[4] J. Ordonez-Lucena, J. Folgueira, L.M. Contreras, A. Pastor Perales. "The use of 5G Non-Public Networks to support Industry 4.0 scenarios", *IEEE Conference on Standards for Communications and Networking (CSCN)*, Granada, Spain, October 2019.

[5] R. Guerzoni, et al. "Analysis of end-to-end multi-domain management and orchestration frameworks for software definedinfrastructures: an architectural survey", *Transactions on Emerging Telecommunications Technologies*, Vol. 28, Issue 4, 2017.

[6] C.J. Bernardos, A. Rahman, J.C. Zuniga, L.M. Contreras, P. Aranda, P. Lynch, "Network Virtualization Research Challenges", RFC 8568, April 2019.

[7] F Esposito, I Matta, V Ishakian, "Slice embedding solutions for distributed service architectures", ACM Computing Surveys (CSUR), 2013.

[8] T. Taleb, I. Afolabi, K. Samdanis, F.Z. Yousaf, "On Multi-Domain Network Slicing Orchestration Architecture and Federated Resource Control", *IEEE Network*, Vol. 33, Issue 5, Sept.-Oct. 2019.

[9] Y. Rekther, T. Li, S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[10] L.M. Contreras, D.R. López, "A Network Service Provider Perspective on Network Slicing", *IEEE Softwarization*, January 2018. Available at: https://sdn.ieee.org/newsletter/january-2018/a-network-service-provider-perspective-on-network-slicing

[11] W.J.A. Silva, D.F.H. Sadok, "A Survey on Efforts to Evolve the Control Plane of Inter-Domain Routing", *Information*, Vol. 9, 2018.

[12] H. Gredler, J. Medved, S. Previdi, A. Farrel, S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, March 2016.

[13] L. Ginsberg, S. Previdi, Q. Wu, J. Tantsura, C. Filsfils, "BGP - Link State (BGP-LS) Advertisement of IGP Traffic Engineering Performance Metric Extensions", RFC 8571, March 2019.

[14] L. Ginsberg, S. Previdi, S. Giacalone, D. Ward, J. Drake, Q. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", RFC 8570, March 2019.

[15] H. Asai, M. MacFaden, J. Schoenwaelder, K. Shima, T. Tsou, "Management Information Base for Virtual Machines Controlled by a Hypervisor", RFC 7666, October 2015.

[16] N.T. Thomopoulos, *Probability Distributions with Truncated, Log and Bivariate Extensions*, Springer, 2018.

[17] S. Kortum, "Normal Distribuion," Universidade de São Paulo, November 2002. Available at: https://edisciplinas.usp.br/pluginfile.php/2028147/mod_resource/content/0/Normal truncada.pdf

[18] D. Ceccarelli, Y. Lee, "Framework for Abstraction and Control of TE Networks (ACTN)", RFC 8453, August 2018.

[19] F. Esposito, D. Di 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