2018 IEEE 32nd International Conference on Advanced Information Networking and Applications

End-to-End Efficient Heuristic Algorithm for 5G Network Slicing

Amal Kammoun[1,2], Nabil Tabbane[1], Gladys Diaz[2], Abdulhalim Dandoush[3] and Nadjib Achir[2]
[1] MEDIATRON Laboratory, University of Carthage, Sup'Com, Tunisia
[2] L2TI, Paris 13 University, Sorbonne Paris Cite, France
[3] ESME Sudria, France
{amal.kammoun, nabil.tabbane}@supcom.tn
{gladys.diaz, nadjib.achir}@univ-paris13.fr
abdulhalim.dandoush@inria.fr

***Abstract*— Coupling Software Defined Networking (SDN) and Network Function virtualization (NFV) has proved to be a promising paradigm for flexible resource provisioning in future networks. Network slicing is a very recent methodology that can be used in this paradigm for accommodating new services with wide different requirements over the same physical network. In this work, we propose a mathematical formulation of an optimization problem for an end-to-end (E2E) network slices deployment for different 5G-based use-cases. Each use case such as video streaming, intelligent transport, e-Health and public safety, has its own availability, reliability and delay tolerance requirements. Then, in view of the fact that the optimization problem is NP-Hard, a low-cost and efficient heuristic algorithm has been proposed. Last, the efficiency of the proposed algorithm is validated through extensive simulation. Applying our algorithm improves the quality of service (QoS) afforded to the users.**

***Keywords*—5G; SDN; NFV; Network Slicing; Reliability; Availability**

I. INTRODUCTION

Nowadays, in almost every area of expertise, people and industries are on top of a digital transformation due to the critical users' requirements and also to the emergence of new use cases [1]. Those new transformations make the mobile network operators rethink their actual implemented systems and strategies. In fact, the major dilemma of the existing technologies is the lack of flexibility, extensibility and also scalability because of the strong coupling between data plane and control plane and also to the vertical implementation of network's software. Thus, today's networks architectures are not able to efficiently implement some use cases and to satisfy critical requirements in terms of latency, reliability and availability [2].

Fifth Generation Mobile Network (5G) presents promising technologies aiming to meet new challenges that cannot be raised by actual network's implementations [3]. Unlike previous technologies, such as LTE and UMTS, 5G will not only enhance the network system but will also provide an end-to-end infrastructure that will provision all emergent use cases and will respond to stringent users' requirements [4]. In fact, the vision of the 5G system is to migrate towards a horizontal approach characterized by open software functions that run on logical and virtualized resources.

The key technologies for the 5G vision are the Software-Defined Networking (SDN) [5] and the Network-Function Virtualization (NFV) [6]. Those two paradigms are complementary and they will together achieve 5G goals and ensure high flexible networks. In fact, the SDN separates the control plane from the underlying infrastructure and moves the control logic to a centralized controller. In his turn, NFV decouples Network Functions (NFs) from hardware resources to be implemented on virtual machines which are running on top of servers, switches or routers. The controller will then program network's elements and steer flows through network functions. Thus, SDN and NFV provide network's programmability and break vertical software integration.

Moreover, in order to achieve the highest performance and flexibility for the network, 5G aims to break-down the "one size fits all" approach to move to the "one size per service" approach [7]. After the virtualization and softwarization of the network, the network slices can be implemented [8]. In fact, 5G proposes the network slicing technology that provides for each type of service a dedicated network slice [9]. Using this technology, slices will be provision on-demand and each slice will tailor only functions and services that the business model needs.

For the network slicing technology, challenges of the admission control and resources allocation arise. Our work is related to this context. In fact, we are interested in the deployment of E2E network slices for 5G use cases. For this purpose, we first introduce a 5G use cases classification based on requirements in terms of reliability, availability and latency. Then, we present our heuristic algorithm which aims to implement the slice on the most convenient resources in order to satisfy the user requests. Target resources have to be chosen according to the requirements of different use cases. We focus on this work on the reliability, availability and latency requirements.

The remainder of this paper is organized as follows: Section 2 presents the related work. In the third section we overview the 5G use cases classification based on their reliability, availability and latency requirements. Section 4

1550-445X/18/$31.00 ©2018 IEEE
DOI 10.1109/AINA.2018.00065

presents our system model and the proposed heuristic algorithm. In section 5, simulations results of the proposed algorithm are presented. The last section concludes our work and gives perspectives.

II. RELATED WORK

In the literature, several algorithms for the virtual network resources embedding have been proposed in order to enhance the user's satisfaction, minimize the network deployment cost, improve resources utilization, etc. In [10] authors propose a heuristic algorithm for the admission control mechanism of user's requests. This Algorithm allocates dynamically network resources to each slice while maximizing the slice user's Quality of Experience based on RAN slice prioritization. This paper considers only the maximization of the user data rate when the network slice is created. However, for the emergent 5G use cases, additional parameters have to be guaranteed in order to satisfy the users' requests.

In [11] authors propose a unified approach based on the SDN paradigm that aims to optimize the provisioning of both virtual machines and network bandwidth. The proposed algorithm minimizes the cost of the over-provisioning of cloud resources when they are requested by the users. This work interests essentially on the optimization of the network bandwidth provisioning without considering other parameters that may impact the QoS like the availability, reliability and E2E latency.

In [12] authors study the problem of the network services decomposition and embedding. They propose two algorithms for the mapping of network service chains to the network infrastructure. The first proposed algorithm is an Integer Linear Program (ILP) while the second one is a heuristic algorithm. Both algorithms try to minimize the cost of the mapping while considering the requirements of the network service chaining and also the capabilities of the network infrastructure. The proposed embedding cost is based on the CPU, memory and storage of physical nodes, the cost of bandwidth in a physical link and the resources utilization rate. In this paper, authors consider only the bandwidth and the delay for the required QoS.

In [13] authors propose an algorithm for the virtual data center embedding problem which considers several constraints. The aim of their proposal is to minimize the network resource utilization rate while maximizing the reliability and also the revenue for the infrastructure providers. Regarding the 5G use cases requirements, other QoS indicators need to be considered.

In [14] authors propose two heuristic algorithms in order to solve the network slicing problem in the case of simultaneous processing and routing of a set of service requests. The aim of their work is to optimize the implementation of service function chains according to the link and node capacity constraints. No interest on the user's QoS requirements was made.

In summary, virtual networks embedding and resources allocation issues were studied for different contexts and several decision algorithms were proposed. They exploit multiple objective optimization techniques. However, regarding the emerging 5G use cases requirements, the decision of the network resources allocation needs to take into consideration several other constraints which guarantee the offered Quality of Service and Quality of Experience to users.

In our work, we present an optimization algorithm for the creation of a new slice in a given network according to a user request description. The purpose of our proposal is to select the best target network resources among available configuration scenarios in order to satisfy the slice requirements. Our proposed algorithm considers the values of the E2E availability, reliability and latency in order to provide the convenient slice to the users' requests.

III. STATE OF ART ON THE REQUIREMENTS OF 5G USE CASES

Several industries, research groups and standardization bodies such as Ericsson, NGMN Alliance and 3GPP have been focusing recently on the feasibility of some use cases of the 5G technology [1, 4, 15]. They have set the requirements that the 5G network has to achieve for the different use cases. Those use cases were classified according to their requirements into several categories. For instance, the Ericsson's white paper [1] presents three families for the 5G use cases which are i) massive machine type communication (mMTC), ii) critical communications, and iii) enhanced mobile broadband (eMBB).

This work addresses the optimization of slices deployment within the 5G context while taking into consideration the most important constraints for any use case. According to the literature, the key use case requirements are reliability, availability and latency. In the following, those parameters are defined.

- Reliability: The reliability is the ability to deliver a service correctly according to its requirement without interruption. The reliability rate required by each 5G use case depends on its classification. Some numerical values of reliability rates for different 5G use cases are presented in the Ericsson white paper [1]. For instance, the 5G network should provide a reliability rate that is equal or higher than 99.999% for the ultra-reliable communication use case. For the use cases which are non-reliability critical such as pervasive video, the reliability rate may be less than 99%.

- Availability: As the 5G will be used for the public safety and the communication of emergencies, the network has to insure a required level of availability. In fact, the availability is the ability to deliver a service when requested in order to fulfill the required functions.

- Latency: The latency metric depends on E2E delay and the data plane latency. 5G networks have to insure in general 10 ms E2E latency for the non-latency critical use cases and at most 1 ms for the ultra-low latency use cases [4].

TABLE 1. 5G USE CASES CLASSIFICATION

Use Case Family	**Example of use cases**	**Requirements: Latency, Reliability, and Availability**	**Latency$_{max}$**	**Service Availability Level (SAL)**	**Availability$_{min}$**	**Reliability$_{min}$**
Enhanced Mobile Broadband (eMBB)	Pervasive video, Dense Urban Society, Video streaming in stadium Ultra-low cost network High Speed train, Moving Hot spots, etc.	There is no accuracy need for the reliability and availability for this use case family. For most use cases, the latency should be very low	Between 5 and 10 ms	SAL 3 : Available in normal conditions only	95%	95%
Massive IoT	Smart wearables, Sensor networks	This family is relatively tolerant to the delay and it requires a normal to high reliability and availability	10 ms			
Critical Communications	Ultra high reliability and ultra low latency use cases: Automated traffic control/driving, collaborative robots, remote object manipulation	Use cases of this category are characterized by a very low latency, a very high reliability and a high availability.	1 ms	SAL 2: Available in normal and overloaded conditions	99,9%	99,999%
	Ultra low latency use cases: Tactile Internet	Use cases of this category are characterized by a very low latency, a high reliability and availability.	1 ms	SAL 2: Available in normal and overloaded conditions	99,9%	99,9%
	Ultra high reliability and availability: eHealth (Extreme life critical), Public safety	Use cases of this category are characterized by a very high reliability and a very high availability with a low latency,	5 ms	SAL 1: Always available	99,999%	99,999%

Table 1 presents for each 5G use case family the required values of availability, reliability and also latency. It presents a classification of the 5G use cases according to their service availability level based on the ETSI recommendations [16].

IV. SYSTEM MODEL

A. Network slices deployment process

For the creation of network slices, we consider a network orchestrator that receives users' demands and according to their requirements it deploys a network slice that responds to their requests. Figure 1 presents the considered network slicing architecture.

A slice is composed of a set of *Virtual Network Functions* VNFs implemented on top of the virtual infrastructure. This composition of VNFs is adequately configured and chained according to the use case requirements. Each implemented slice is isolated from the other slices and has a dedicated network proceeding. The use of the network slicing concept allows the network system to provide as many network slices as requested over the same network infrastructure.

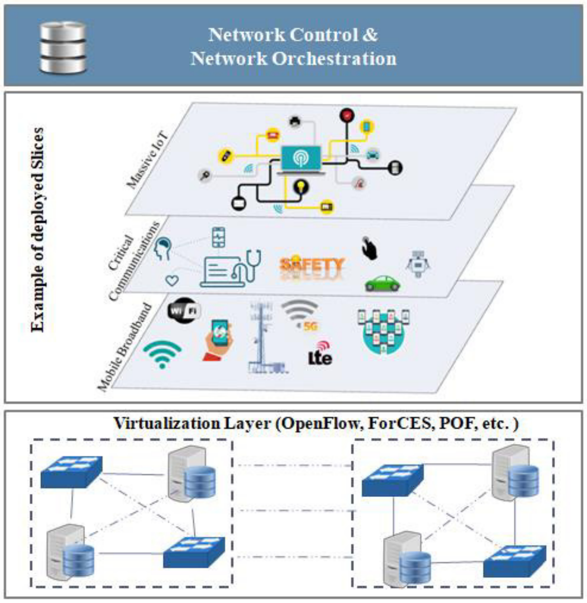

Fig. 1. Network Slicing Architecture

Clearly, the E2E slice's performance depends on the performance of all its constituent network resources and applications. Thus the E2E reliability and availability of the deployed slice depends on the reliability and availability of each constituent functional block. The combination of several components in the slice deployment is characterized by a serial dependency. In fact, all the constituent functional blocks need to be available and reliable at the same time in order to have the entire slice available and reliable. For example, the deployment of a VNF over the NFV Infrastructure (NFVI) is a serial dependency and its reliability and availability depends on the physical hardware, the hypervisor and the software of the VNF itself. In the considered architecture, we define a data base that contains the computed values of the reliability, availability, and latency of each network resource based on its previous behavior.

B. *Problem Formulation*

The virtualized infrastructure (VI) provides an abstraction of the physical network resources layer. It contains N Forwarding Elements (FEs) interconnected according to the VI topology and a set of K servers which are connected to the FEs. In order to create a slice, an interconnection between virtual network resources is required. Thus, for each slice a set of *v* Virtual Machines (VM), *n* nodes and *l* Links are associated ; S_i=<{VM_i}, {N_i}, {$Link_i$}>.

The Slice creation process is characterized as a set of M tasks to be executed with a deadline d_{thr}. A task is defined as the allocation of a network resource to the slice and the implementation of the required Virtual Network Function (VNF). Each task is characterized by its reliability, availability and latency requirements. We consider a binary variable $w_{(i,j)}$ to indicate if a the network resource NR_j is allocated to the slice $Slice_i$ or not:

$$w_{(i,j)} = \begin{cases} 1 \text{ if the resource } NR_j \text{ is allocated to the } Slice_i \\ 0 \text{ otherwise} \end{cases} \quad (1)$$

We suppose that all the forwarding elements, links and also servers have unlimited computational capacities to implement all the requested slices. In fact, we interest only on the reliability, availability and latency characteristics of those network resources.

$$NR_j \begin{pmatrix} \text{Reliability} = R_j \\ \text{Availability} = A_j \\ \text{Latency} = L_j \end{pmatrix} \quad (2)$$

When a new slice creation request (SCR) is received, the *Network Orchestrator* has to search for the target resources to allocate for this slice based on its service-level agreement (SLA). In fact, the required SLA of each slice is described based on the following parameters:

$$SCR_i \begin{pmatrix} \text{Reliability} = R_i \\ \text{Availability} = A_i \\ \text{Latency} = L_i \end{pmatrix} \quad (3)$$

In this work, our goal is the maximization of the availability and reliability for each created slice while considering the requirements of the user's demand. Therefore, the following constraints have to be considered:

- Availability constraint:

$$A_j \geq A_i \ \forall j \quad (4)$$

Constraint (4) guarantees that the availability of the selected network resource is equal or higher than the required user's availability.

- Reliability constraint:

$$R_j \geq R_i \ \forall j \quad (5)$$

Constraint (5) guarantees that the reliability of the selected network resource is equal or higher than the required user's reliability.

- Delay constraint:

$$\sum_j D_j \leq D_i \quad (6)$$

Constraint (6) ensures that the E2E delay from the source node to the destination node is lower than the maximum tolerant delay by the requested service.

- Objective function:

Our objective is the maximization of the E2E reliability (R) and E2E availability (A) of the deployed slice:

$$\text{Maximize } A + R \quad (7)$$

This problem is NP-Hard. Therefore, we introduce in the next part a heuristic algorithm which searches for the optimal solution by selecting the most suitable resources for the requested slice in a reasonable time.

C. *Proposed Algorithm*

In order to determine the target resources for the slice creation, we have opted for the utility theory. In fact, using the utility function, our algorithm computes the score of each candidate network resource while taking into consideration input parameters as well as their weights. Then, the selection will be based on the score of each resource. Our algorithm will select the resources that have the highest utility score.

The utility function that has been retained is as follow [17]:

$$U(x) = 1 - e^{-\alpha x} \quad (8)$$

Where x is the decision vector and α is the corresponding weight (α>0).

In this work, the decision vectors that we implement in the utility function are the network resources availability, reliability and latency. The formula used to compute the score of the network resource NR_j while taking into account the requirements of the slice request *i* is the following:

$$SC(NR_j) = \left(1 - e^{-\alpha_A \times A_j}\right) \times \left(A_i \otimes A_j\right) \times \left(1 - e^{-\frac{\alpha_L}{L_j}}\right) \times \left(L_i \otimes L_j\right) \times \left(1 - e^{-\alpha_R \times R_j}\right) \times \left(R_i \otimes R_j\right) \quad (9)$$

Where

- $(A_i \otimes A_j) = \begin{cases} 1 \text{ if } A_j \geq A_i \\ 0 \text{ otherwise} \end{cases}$
- $(R_i \otimes R_j) = \begin{cases} 1 \text{ if } R_j \geq R_i \\ 0 \text{ otherwise} \end{cases}$
- $(L_i \otimes L_j) = \begin{cases} 1 \text{ if } L_j \leq L_i \\ 0 \text{ otherwise} \end{cases}$
- α_A: Weight of the decision criterion A_i.
- α_R: Weight of the decision criterion R_i.
- α_L: Weight of the decision criterion L_i.

This utility function is an increasing function. Therefore, the variation of utility values is proportional to the changes in variable values. Also, it distinguishes between ascending and descending criteria. For example, the availability and reliability are ascending criteria that has to be maximized to better meet the user satisfaction, while the latency is a descending criterion to minimize.

The chosen function returns values normalized to [0, 1] domain. For decision criteria we set $\alpha_A = \alpha_L = \alpha_R = \frac{1}{3}$ which means that equal weight is given for the different performances criteria.

After computing different scores of network resources between the source node and the destination node, the algorithm will select the one with the best score value. This process enables us to determine the best network slice deployment scenario while considering network resources characteristics.

We present in the following the proposed algorithms that search for the slice implementation scenario that satisfies the user request. The first proposed algorithm determines at first the list of all possible paths from the source to the destination. In fact, the network control system has a complete view of all the network resources. Therefore, for each requested slice it determines the different implementation scenarios. Then *Algorithm1* determines within a deadline execution d_{thr} the convenient available network slice implementation scenario between the source node and the destination node. This scenario is selected while executing the second proposed algorithm. In fact, the second algorithm computes the score of each selected path and the score of each resource belonging to this path. When the *Algorithm2* finds a path that has a non-null score, this scenario will be retained. In fact, according to the equation (9), a path has a non-null score only if its *reliability*, *availability* and *latency* characteristics correspond to the slice requirements.

Algorithm1 Selection of the Slice Network Resources

```
src = Source node of the $Slice_i$
dst = Destination node of the $Slice_i$
Initialize S = Null; S is the retained network path assigned to the $Slice_i$
P = $F_{Controller}$ (src, dst); This function restitutes the list of all possible paths from src to dst computed by the controller.
While ($S = Null$ && $d < d_{thr}$) do
    S = ScoredPath (src, dst, P, score); score is an input/output argument for ScoredPath function. It is used to compute the score of the retained path.
done
If score <> 0 then
    confirm S as the target resources for the $Slice_i$
    return S
else
    return " request timeout"; that means $d >= d_{Thr}$ or score=0
End if
```

Algorithm2 ScoredPath (input src, input dst, input P, input/output score)

```
L = Null; L is the retained network resources path for the $Slice_i$.
c1 = card1(P); The function card1 computes the cardinality or the number of possible paths from src to dst.
i = 1
PS = 0; the score of the $Path_i \in P$ ( each path of the list P is referenced as $Path_i$ )
score = 0; score of the retained path.
While (i<=c1 && PS=0) do
    c2 = card2($Path_i$); The function card2 computes the cardinality or the number of network resources of the $Path_i$..
    v = 1
    for j = 1 .. c2 do
        Compute the score $SC(NR_j)$ of the network resources $NR_j$ of the $Path_i$ using the equation (9).
        $v = v \times SC(NR_j)$ ; compute the path score.
    end for
    PS = v;
    If (PS <> 0) then
        L = $Path_i$; the retained path.
        score = PS; score of the retained path.
    End if
    i++
done
Return L
```

V. NUMERICAL RESULTS

In this section, we present simulation results of our proposed utility-based algorithm (UBA) compared to the following mono-criterion algorithms:

- Delay-based algorithm (DBA): the choice of network resources is based on the delay parameter. The aim of this algorithm is the minimizing of the E2E delay.
- Availability-based algorithm (ABA): the slice is embedded on the most available resources.
- Reliability-based algorithm (RBA): the selection of the network resources for the requested slice is based on the reliability parameter. The most reliable network resources are selected.

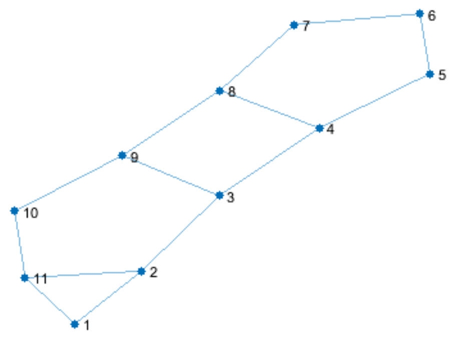

Fig. 2. The Abilene topology (11 nodes, 28 directed links)

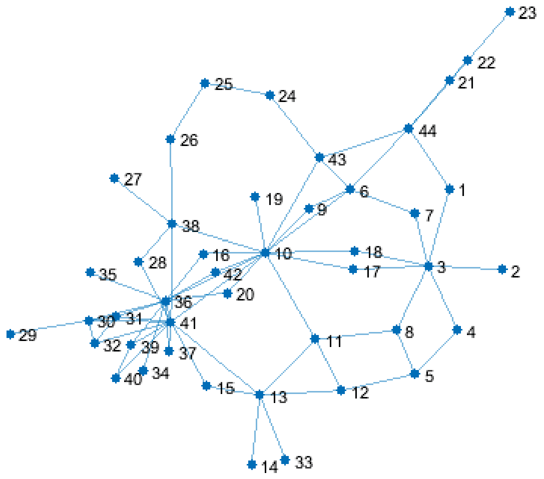

Fig. 3. The Geant topology (44 nodes, 142 directed links)

We have evaluated our algorithm using MATLAB for the three use cases families presented in Table 1. Two scenarios of real network topologies were considered: i) the Abilene topology as a small network scenario and ii) the Geant topology as a large network scenario. Those two topologies are obtained from the Internet Topology Zoo project [18]. The reliability, availability and delay of network resources are generated randomly.

Figure 4 presents the score of the slices deployment in the case of the small network scenario. We have evaluated the score for the enhanced mobile broadband, massive IoT and also critical communications use cases which were presented in section III. This score has been evaluated by applying our proposed UBA algorithm as well as the ABA, RBA and DBA algorithms. Simulation results show that the total score of the slice deployment process for each use case family is enhanced when our algorithm is applied compared to the other considered algorithms. In fact, our algorithm takes into consideration the three network performance criteria (availability, reliability and latency) for the network resources selection while the mono-criterion algorithms consider only one network performance indicator which reduces the score of the deployed slice.

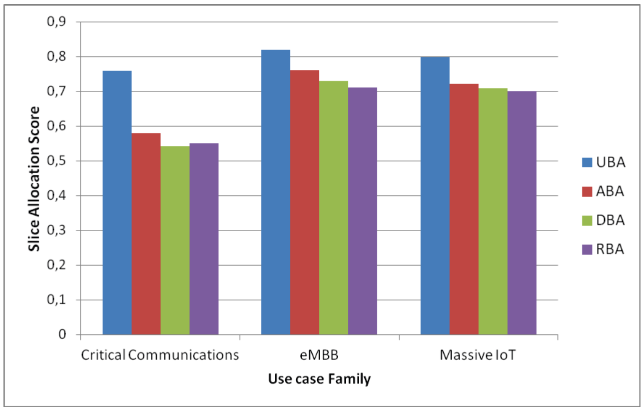

Fig. 4. Slice Deployment Score in Abilene topology

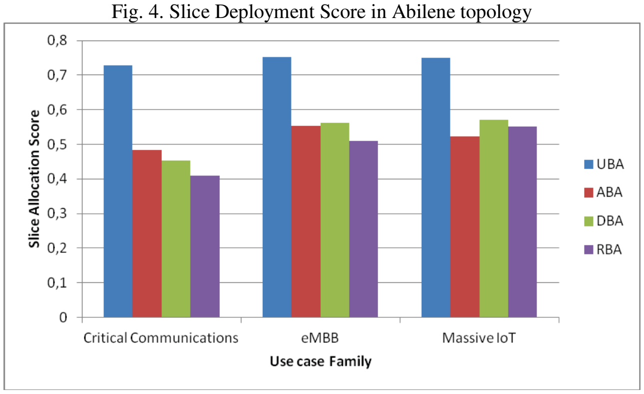

Fig. 5. Slice Deployment Score in Geant topology

Figure 5 presents the E2E score generated by the different algorithms for the large network topology. Similarly to the small network scenario, simulation results show that the score of the slice deployment when our proposed algorithm is applied is higher than the score generated by the other mono-criterion algorithms. Moreover, for the other algorithms, the total slice creation score decreases with the increase of the number of network resources while we maintain a high score when we apply our proposed algorithm.

For the two network topologies, the eMBB and the massive IoT use cases families have the highest utility score for all the applied algorithms. In fact, those two families are characterized by a flexible requirements compared to the critical communications use case family which is characterized by stringent requirements. Simulation results demonstrate that the UBA increases the utility of the deployed slices for all the use case families as it selects the network resources based on the three network performance indicators.

Therefore, our algorithm is efficient even in a large scale network. The selection of the best slice creation scenario among available ones is insured when our proposal is applied. In fact, the three network performance indicators are considered in the UBA algorithm, while for the other algorithms only one indicator is considered and the others are ignored. Thus, our algorithm provides a better QoS for the deployed slices.

VI. CONCLUSION

This paper proposes a utility-based slice deployment algorithm (UBA) that aims to maximize the QoS for the deployed slices while considering the use case requirements as well as the network resources characteristics in terms of *availability*, *reliability* and *latency*. In our scenario, we consider a network control and orchestration module that creates slices according to users' demands. Our objective is to maximize the slice deployment score by selecting the most convenient network resources. Simulation results show that the proposed algorithm is able to select the most convenient network slice deployment scenario while maximizing the total deployment score. As future works, we will consider the management of the deployed slices in order to maintain the required QoS. Furthermore, we will try to better investigate the network performances by handling the interference between the several deployed slices.

REFERENCES

[1] ERICSSON WHITE PAPER, 5G SYSTEMS. https://www.ericsson-.com/assets/local/publications/white-papers/wp-5g-systems.pdf, Jan. 2017. Accessed: 2017-09-18.

[2] N. Omnes, M. Bouillon, G. Fromentoux and O. L. Grand, "A programmable and virtualized network & IT infrastructure for the internet of things: How can NFV & SDN help for facing the upcoming challenges," 2015 18th International Conference on Intelligence in Next Generation Networks, Paris, 2015, pp. 64-69.

[3] The European 5G Annual Journal, 2017, [Available online: https://bscw.5g-ppp.eu/pub/bscw.cgi/d204796/Euro%205G%20-Annual%2_Journal%202017-web.pdf].

[4] "5G White Paper" February 2015 [online] Available: https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf.

[5] D. Kreutz, F. M. V. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky and S. Uhlig, "Software-Defined Networking: A Comprehensive Survey," in Proceedings of the IEEE, vol. 103, no. 1, pp. 14-76, Jan. 2015.

[6] R. Mijumbi, J. Serrat, J. L. Gorricho, N. Bouten, F. De Turck and R. Boutaba, "Network Function Virtualization: State-of-the-Art and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 236-262, Firstquarter 2016.

[7] M. R. Sama, X. An, Q. Wei and S. Beker, "Reshaping the mobile core network via function decomposition and network slicing for the 5G Era," 2016 IEEE Wireless Communications and Networking Conference, Doha, 2016, pp. 1-7.

[8] Paul, M., Schallen, S., Betts, M., Hood, D., Shirazipor, M., Lopes, D., Kaippallimalit, J., - Open Network Fundation document "Applying SDN Architecture to 5G Slicing", April 2016; www.opennetworking.org/images/stories/downloads/sdnresources/technical-reports/Applying_SDN_Architecture_to_5G_Slicing_TR-526.pdf.

[9] Vassilaras et al., "The algorithmic aspects of network slicing," in IEEE Communications Magazine, vol. 55, no. 8, pp. 112-119, 2017.

[10] M. Jiang, M. Condoluci and T. Mahmoodi, "Network slicing management & prioritization in 5G mobile systems," *European Wireless 2016; 22th European Wireless Conference*, Oulu, Finland, 2016, pp. 1-6.

[11] J. Chase, R. Kaewpuang, W. Yonggang and D. Niyato, "Joint virtual machine and bandwidth allocation in software defined network (SDN) and cloud computing environments," 2014 IEEE International *Conference on Communications (*ICC), Sydney, NSW, 2014, pp. 2969-2974.

[12] Sahel Sahhaf, Wouter Tavernier, Matthias Rost, Stefan Schmid, Didier Colle, Mario Pickavet, Piet Demeester, Network service chaining with optimized network function embedding supporting service decompositions, In Computer Networks, Volume 93, Part 3, 2015, Pages 492-505, ISSN 1389-1286.

[13] Xuemin Wen, Yanni Han, Bing Yu, X. Chen and Zhen Xu, "Towards reliable virtual data center embedding in software defined networking," MILCOM 2016 - 2016 IEEE Military Communications *Conference*, Baltimore, MD, 2016, pp. 1059-1064.

[14] N. Zhang, Y. F. Liu, H. Farmanbar, T. H. Chang, M. Hong and Z. Q. Luo, "Network Slicing for Service-Oriented Networks Under Resource Constraints," in IEEE Journal on Selected Areas in Communications, vol. 35, no. 11, pp. 2512-2521, Nov. 2017.

[15] "3gpp tr 22.891, feasibility study on new services and markets technology enablers, stage 1, rel-14, V14.2.0, 2016-09."

[16] ETSI Industry Specification Group (ISG) NFV. Etsi gs nfv-rel 001 v1.1.1 : Network functions virtualisation (nfv); resiliency requirements, January 2015.

[17] L. Suciu and K. Guillouard. A Hierarchical and Distributed Handover Management Approach for Heterogeneous Networking Environments. In 3rd International Conference on Networking and Services, 2007.

[18] Available at http://www.topology-zoo.org/.