Elastic Network Slice Composition with Traffic-Weighted Availability-Guarantee

Divya Khanure*, Riti Gour†, and Jason P. Jue*
*Department of Computer Science
The University of Texas at Dallas, Richardson, Texas 75080, USA
† Department of Aviation and Technology,
San Jose State University, San Jose, California 95192, USA

Email:{divya.khanure, and jjue}@utdallas.edu, riti.gour@sjsu.edu,

***Abstract*—Network slicing has helped in advancing the types of applications that could be supported by next-generation telecommunication networks. This is achieved by hosting Service Function Chains (SFCs) over the network slices that satisfy user demand. With that in mind, we have considered the problem of the composition of elastic network slices on physical infrastructures such that the availability requirements of the SFCs are satisfied. Instead of traditional availability, we utilize the concept of traffic-weighted availability to satisfy the SFC demands. We present a linear programming model to calculate the provisioning of the bandwidth resources over the slice paths while minimizing the total cost of provisioning. Our simulation results show that the approach in this work is effective when compared to the well-known dedicated protection mechanism.**

***Index Terms*—Network slicing, service function chains, traffic-weighted availability, and elastic slice.**

I. Introduction

The emergence of applications that need support for heavy traffic flows has been one of the driving forces of the next-generation telecommunications network. Traditional virtual networks were introduced to improve network utilization and efficiency. Relatively newer concepts like network slicing and network function virtualization (NFV) have only been adding to the benefits. These have helped to support a broad range of highly available, ultra-low latency applications for 5G networks. NFV is an enabler of network slicing, which helps host traditional network functions virtually and saves capital and operational costs. NFV also helps with making network reconfiguration simpler and faster.

Network slice composition problems have been studied in the past, majorly focusing on static slices with fixed requirements. Provisioning for this kind of network slice is easier compared to a situation with variable resource requirements, which is referred to as an elastic slice. In this work, we consider elastic network slices whose bandwidth requirements can change over time.

A key metric when deploying network slices is availability, which can be defined as the fraction of time that the slice is operational and providing the required resources. When composing elastic network slices, the network operator needs to ensure that performance metrics, such as availability, are guaranteed at all times, even when the bandwidth changes over time. If the availability is compromised when attempting to reallocate resources to accommodate the change in bandwidth, a penalty may be incurred by the network operator. This penalty may be a part of the service level agreement between the slice operator and the network operator.

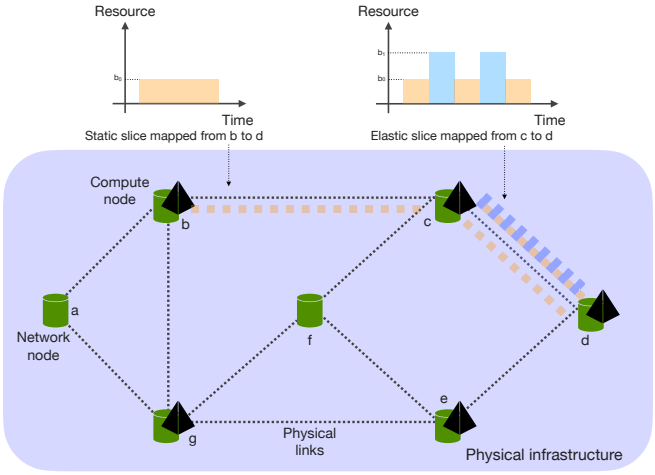

Fig. 1. Elastic network slice on the physical infrastructure.

The network slice consists of virtual computing resources (slice nodes) hosting VNFs of the SFCs and virtual links connecting these VNFs. Slice design involves topology, resource allocation, and mapping onto physical infrastructure.

In this work, we study an elastic network slice (shown in Fig. 1) composition problem that requires the mapping of slice nodes and links to the physical infrastructure. Each slice node hosts one or more VNFs, and each slice link provides bandwidth capacity for network flows between the nodes. When mapping slice resources over the physical network, proper dimensioning of computing and bandwidth resources is required. The bandwidth required by flows between the slice nodes may vary over time, and the resources allocated for the slice must satisfy the bandwidth availability requirements of the SFC while minimizing cost for the network operator.

If the availability of the VNFs and physical resources on a single path can satisfy the availability requirement of the SFC, then a single-path slice topology with the appropriate bandwidth allocation is sufficient to support the SFC. However, if the availability of a single path is less than the required avail-

979-8-3315-2096-0/25/$31.00 ©2025 IEEE

ability, then a multi-path slice topology must be utilized. Once the topology is constructed, we utilize the concept of traffic-weighted availability to determine the amount of bandwidth that is allocated on each path of the slice. Traffic-weighted availability is a metric that measures the long-term fraction of the required traffic supported by a specific slice. Its primary benefit lies in its ability to decrease the resources needed to meet a service's availability requirements. Unlike conventional availability metrics, traffic-weighted availability aims to reduce the allocation of unnecessary redundant resources, providing just the right amount of resources to meet demand.

In our previous work [1] we formulated a slice composition problem for allocating an appropriate amount of bandwidth resources over slice links. The objective was to meet the availability requirement while minimizing the cost of establishment. This work included slice topology composition for the static slices. Unlike our previous work, this work focuses on elastic network slices, and how we can provision appropriate resources to satisfy the availability requirement.

The rest of the paper is organized as follows. In Section II we present the related background literature. In Section III we present the problem statement and the availability analysis in detail. Section IV discusses our algorithm for slice composition, which includes path selection and bandwidth provisioning along those paths. In Section V we discuss the effectiveness of our approach through simulations. We conclude the paper in Section VI.

II. Background Literature

Network slice composition and resource allocation problems are well studied for next-generation telecommunication networks [2]–[7]. Several recent works have also discussed elastic network slices and their embedding on the physical infrastructure [8]–[11]. In [8], the authors present an elastic slice problem with reinforcement learning to make admission control decisions. In [9] and [11] the authors present a slice admission architecture in which a forecasting agent predicts the future resource usage of currently active elastic slices for the next time window. This information is used to allocate the required resources to each slice. An admission control problem using a reinforcement learning (RL) approach is also discussed, and the decision is made on whether to accept or reject the incoming slice requests. The difference between our work and the above works is that we have an additional availability constraint that needs to be satisfied before making an admission control decision to accept or reject the slice.

The work in [10] focuses on an approach that uses a digital twin of the 5G network in predicting network traffic to improve network resource utilization and reliability. The prediction results from the approach are used in provisioning elastic network slices. The metrics of performance are Root Mean Square Error and Mean Absolute Percentage Error in [10], while ours is traffic aware availability and cost.

III. System Model

We consider a physical network represented by a graph $G = (V, E)$, where V is a set of nodes and E is a set of physical edges. The set of nodes is further classified into V_n and V_c, where V_n is a set of network nodes that are responsible for the bandwidth flows across the network, and V_c is a set of compute nodes which provide computing capabilities for the virtual functions. We assume that network nodes are always available to support the slice request and we denote the availability of the compute nodes by a_v, $\forall\ v \in V_c$. The capacity of compute nodes is denoted by s_v, $\forall\ v \in V_c$. Each edge $e \in E$ has an availability value a_e, bandwidth capacity b_e, and length d_e.

For an SFC request i we are given the source s_i and destination d_i along with the variable bandwidth requirement b_r^i where r signifies the bandwidth level. We specify that there are two bandwidth levels for a request, represented by $r = 0$ which is b_0^i, for the base requirement, and $r = 1$ which is b_1^i for the peak requirement. The availability requirement of the SFC is given by A_{req}. We utilize the concept of traffic-weighted availability which gives the long-term fraction of traffic that is supported by the slice, given the bandwidth allocated on each path of the slice, and the availability of physical infrastructure resources and deployed VNFs [1].

The request should maintain the availability requirement during all bandwidth allocations (base and peak). We assume that the SFC requires a set of VNFs, $M = \{m_1, m_2, m_3, .., m_n\}$ which are mapped to compute nodes, and the computing requirement of each function m_i is a function of the bandwidth of the traffic flow, and is given by r_{m_i} in units of CPU cycles per unit time. An individual VNF has an associated availability a_{m_i} and cost C_m.

An ideal case for slice composition would be to have a single path slice topology that can support the variable bandwidth requirements while simultaneously satisfying the availability requirement for the SFC request. Traditionally, we have seen problems in literature with availability-guaranteed approaches for static slices. However, with elastic slices, the challenge is to provision resources such that the slice can maintain its availability while accommodating a change in bandwidth requirement without having to construct a new topology or change the utilized paths on the physical infrastructure. When one path cannot satisfy the bandwidth and availability requirements, we resort to constructing a multi-path slice topology between the source s_i and the destination d_i.

Let K denote the number of paths in the slice topology. We assume that paths are calculated such that there are enough computing resources to host the VNFs and enough bandwidth resources to accommodate the fraction of flows on the links of each path. We define a parameter y_{rk}^i, $\{y_{rk}^i : \mathbb{R}^+ | 0 \leq y_{rk}^i \leq 1\}$ which represents the fraction of requested bandwidth b_r^i that will be allocated on path k. The total capacity of bandwidth resources allocated for a slice operating at bandwidth level b_r^i on path k is given by:

$$b_{rk}^i = y_{rk}^i \cdot b_r^i. \tag{1}$$

We accommodate the possibility of over-provisioning the resources to meet the availability requirement. Thus, it is possible that the sum of b_{rk}^i over all the K paths may exceed the required bandwidth b_r^i.

The availability of each path depends on the availability of the underlying infrastructure components along the path. Furthermore, since paths are not necessarily disjoint, the availabilities may be correlated. We define $\mathbf{x} = (x_1, x_2, ..., x_K)$ as a vector of binary indicators, x_k, where x_k = 1 if path k is available, and x_k = 0 if path k is unavailable. We denote the set of all possible states as χ, and the fraction of time that the system is in state $\mathbf{x} \in \chi$ is denoted as $\pi_{\mathbf{x}}$. The values of $\pi_{\mathbf{x}}$ depend on the availability of the underlying infrastructure resources onto which the slice is mapped and are calculated using the equations in our previous work [1].

If there is no blocking, the total bandwidth available to a SFC request operating at bandwidth level b_r^i in the state $\mathbf{x}$ is given by:

$$\sum_{k=1}^{K} y_{rk}^i \cdot b_r^i \cdot x_k. \tag{2}$$

If this value is greater than or equal to b_r^i, then the SFC is receiving 100% of its requested bandwidth, possibly with some redundancy for meeting availability requirements. If the value is less than b_r^i, then the fraction of requested bandwidth received by the SFC is given by:

$$\sum_{k=1}^{K} y_{rk}^i \cdot x_k. \tag{3}$$

If a SFC requests an increase in its bandwidth from base rate b_0^i to peak rate b_1^i, the request may be blocked if there are insufficient resources, and the SFC will continue to operate at its base rate b_0^i. In this case, the fraction of requested bandwidth received by the SFC is given by:

$$\sum_{k=1}^{K} y_{0k}^i \cdot \frac{b_0^i}{b_1^i} \cdot x_k. \tag{4}$$

The traffic-weighted availability for SFCs operating at the base rate is given by:

$$A_0^i = \sum_{\mathbf{x} \in \chi} min(1, \sum_{k=1}^{K} y_{0k}^i \cdot x_k) \pi_{\mathbf{x}}. \tag{5}$$

The traffic-weighted availability for SFCs operating at the peak rate is given by:

$$\begin{aligned} A_1^i = &\sum_{\mathbf{x} \in \chi} \min\Big(1, \sum_{k=1}^{K} y_{1k}^i \cdot x_k\Big) \pi_{\mathbf{x}} \cdot (1 - P_B) + \\ &\sum_{\mathbf{x} \in \chi} \min\Big(1, \sum_{k=1}^{K} y_{0k}^i \cdot x_k \cdot \frac{b_0^i}{b_1^i}\Big) \pi_{\mathbf{x}} \cdot (P_B), \end{aligned} \tag{6}$$

where P_B is the fraction of time that a request asks for a bandwidth b_1^i, but is only allocated bandwidth b_0^i.

IV. Slice Composition Problem

In this section, we explain how the slice is composed with the selection of paths and present the bandwidth allocation problem. Given the physical infrastructure $G(V, E)$, with physical links and their availability, distance, cost, and bandwidth capacities, compute nodes with their cost and computing capabilities and the slice request from s to d with variable bandwidth requirement, availability requirement, and set of VNFs and their computing requirements, we need to find the number of required paths K that are used in the slice topology and find the routing of K paths that have sufficient capacity and high availability. The next problem is to determine the allocation of bandwidth on each of the paths for both base rate and peak rate conditions such that the traffic-weighted availability requirements are met. The objective is to minimize the total cost of allocated resources.

A. Path Selection

The decision to utilize a single path for a network slice within a network depends on the availability of underlying physical resources. To assess if a single path is sufficient, it is essential to evaluate its overall availability. While the maximum-availability path may satisfy the requirements, it may not necessarily represent the most cost-effective option. Therefore, a trade-off between availability and cost must be carefully considered when selecting paths for network slices.

If a single path is not able to meet the availability requirement of the i^{th} SFC, then the i^{th} flow would need to be directed over two or more paths in the slice topology. For two paths, $\mathbf{x} = (x_1, x_2)$ denotes the availability state of paths p_1 and p_2 ($\mathbf{x} \in \{(0,0), (0,1), (1,0), (1,1)\}$). For a base request, we need to map p_1 and p_2 over the physical infrastructure such that: $y_{0p_1}^i(\pi_{(1,0)} + \pi_{(1,1)}) + y_{0p_2}^i(\pi_{(0,1)} + \pi_{(1,1)})$ and $\pi_{(1,1)} + y_{0p_1}^i \pi_{(1,0)} + y_{0p_2}^i \pi_{(0,1)}$ are greater than the required availability. If for a pair of paths p_1 and p_2, $\pi_{(0,0)} \leq 1 - A_{req}$ (denotes the state when both the selected paths are not available), then the pair is capable of satisfying the availability requirement for the base request.

We utilize the K shortest paths algorithm in which we calculate K shortest paths (not necessarily disjoint) from s to d using Yen's algorithm [12] considering the availability as the weight to compute the cost of the path.

B. Bandwidth Allocation

Given a set of paths, we formulate a linear programming (LP) problem that solves for the optimal amount of bandwidth $(y_{rk}^i \cdot b_r^i)$ for the given paths that results in the lowest cost while meeting the availability requirement. The description of the parameters and variables used are shown in Table 1. The objective of the LP is to minimize the cost:

$$\begin{aligned} \min \mathcal{C}^i = \sum_{k=1}^{K} \Big[\sum_{e \in p_k} \Big(y_{0k}^i \cdot b_0^i + \Big[y_{1k}^i \cdot b_1^i \cdot (1 - P_B) + \\ y_{0k}^i \cdot b_0^i \cdot (P_B) \Big] \Big) \cdot C_u + \sum_{m \in M} C_m \Big]. \end{aligned} \tag{7}$$

The problem is subjected to capacity, availability, and bandwidth constraints. We define the capacity constraint for each physical link where the amount of bandwidth allocated on each link should be less than the capacity of the link as,

$$\sum_{k=1}^{K} y_{rk}^{i} b_r^i \cdot \beta_k^{\hat{e}} \leq b_{\hat{e}},\ \forall \hat{e} \in U_k\ \forall r. \tag{8}$$

The availability constraint specifies that the traffic-weighted availability of the slice should be greater than or equal to the availability requirement of the SFC. When base bandwidth is requested for an elastic network slice, and there are enough resources on the physical infrastructure to provision for the base request, the availability constraint is denoted by the following equation:

$$A_0^i \geq A_{req}. \tag{9}$$

When the peak bandwidth is requested, there could be two possible outcomes. In the first case, we have sufficient resources available to satisfy the peak requirements. In the second case, we do not have sufficient resources available to satisfy the peak requirements. In the second case, we continue to provide base requirements for the slice. The overall availability is affected in the second case. The availability constraint is denoted by the following equation:

$$A_1^i \geq A_{req}. \tag{10}$$

Furthermore, the fraction of bandwidth provisioned on each path should be between 0 and 1. The case in which the fraction of bandwidth is 1 on each path is equal to the dedicated protection approach. This constraint is given by,

$$0 \leq y_k^i \leq 1,\ \forall k \in \{1.....K\}. \tag{11}$$

It is essential to map VNFs onto the designated slice nodes and subsequently map these slice nodes onto the physical compute nodes in the infrastructure. In this context we assume that each instance of a VNF is mapped to a distinct slice node. However, multiple slice nodes can be mapped to the same compute node, ensuring efficient utilization of the physical resources.

In this case, similar to [13], the placement of VNFs along the path does not affect the availability of the SFC. We select compute nodes randomly on the path to map the required slice nodes. The capacity constraint of the compute nodes is given by:

$$\sum_{m \in M} \sum_{k=1}^{K} r_m \cdot \alpha_k^{mi} \leq s_i,\ \forall c_i \in U_k, \tag{12}$$

where α_k^{mi} is 1 if function m (or slice node with function m) is mapped to compute node i on path k, 0 otherwise.

TABLE I
DESCRIPTION OF VARIABLES

Parameters	Description
C^i	cost of establishment for all paths between s and d.
b_0^i	base bandwidth requirement between s and d.
b_1^i	peak bandwidth requirement between s and d.
C_u	cost per unit bandwidth used on a link.
M	set of VNFs in the SFC.
C_m	cost of VNF m.
K	total number of paths in the slice.
$\beta_k^{\hat{e}}$	1 if physical link $\hat{e}$ is on path k, 0 otherwise.
$b_{\hat{e}}$	bandwidth capacity of physical link $\hat{e}$.
x_k	state, 1 if path k is available, 0 otherwise.
χ	set of all possible states.
$\pi_{\mathbf{x}}$	the fraction of time that the system is in state $\mathbf{x}$.
A_{req}	availability requirement of the slice.
Variable	**Description**
y_{0k}^i	the amount of base bandwidth on path k between s and d.
y_{1k}^i	the amount of peak bandwidth on path k between s and d.

V. NUMERICAL EVALUATION

A. Network Setting

We evaluate the performance of our approach over a 14-node USNET network topology. In each experiment, the availability of all physical edges in the topology is set to one of these two values: [0.999999, 0.9999999]. The distance of the edges is in the range of 350-1200 km, which is fixed for the network topology. We set the cost of allocating 1 Gb/s to each physical resource as 400 cost units, and we assume that the capacity of a physical link is 400 Gb/s. To host VNFs, 10 nodes are arbitrarily selected as compute nodes. The capacity of all compute nodes is kept constant at 400 units. The arrival times of the SFC requests follow a Poisson process with 2 requests/hour and an SFC holding time follows an exponential distribution with an average holding time of 1/2 hours. Each of the requests is an elastic request and transitions from base rate to peak rate according to a Poison process with a rate of 5 requests/hour. The amount of time that the request remains at the peak rate before transitioning back to the base rate follows an exponential distribution with an average holding time of 1/5 hours. For each data point in our experiments, we generate a set of 1000 trials consisting of individual SFC requests between randomly selected source-destination pairs in the topology and have used K-shortest paths for routing with the value of K being 2. For each SFC request, we select its bandwidth requirement for the base to be 20 Gb/s and for peak, it is considered to be 40 Gb/s. We randomly select the number of VNFs in the SFC request to be between 2 and 6. To understand how the bandwidth allocation is affected by different levels of peak rate blocking, we consider values of P_B ranging from 0 to 0.1. Note that, with the maximum of 10% peak blocking, we can still satisfy the availability requirement of the request by allocating additional base rate resources on the secondary path. We consider that the VNFs require a number of computing resource units ranging from 1 to 3. We assume that the availability of each VNF is 0.99999 and we run our experiments for six different levels of SFC availability requirement, ranging from 0.9 to 0.999999.

B. Experiments and Discussions

Equation (4) is used to calculate the total cost associated with each accepted slice request. As depicted in Fig. 2, the cost is compared across various availability requirements (0.999, 0.9999, and 0.99999) in relation to a dedicated 1+1 protec-

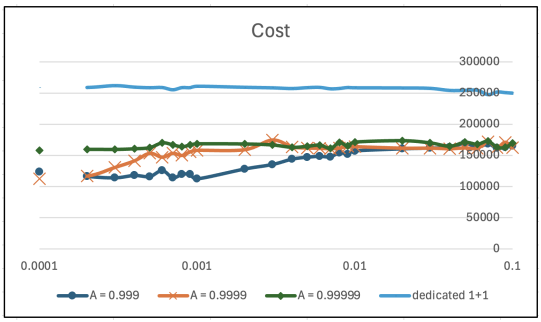

Fig. 2. Cost versus P_B.

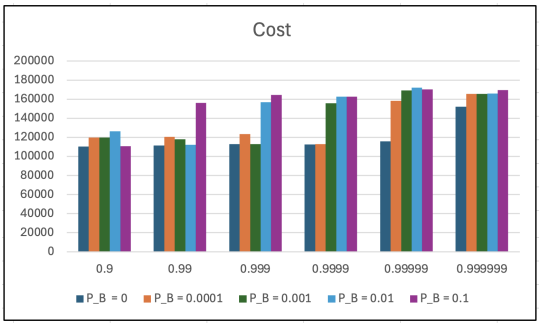

Fig. 3. Cost versus Availability Requirement for different P_B values.

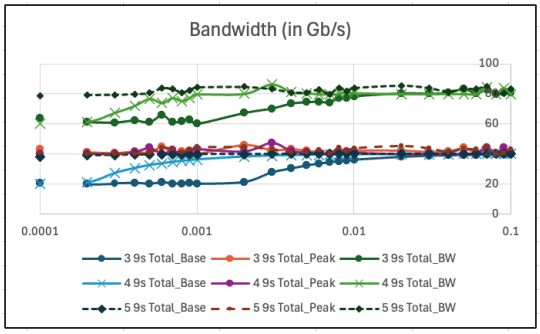

Fig. 4. Bandwidth versus P_B.

tion scheme. The K-link disjoint path selection algorithm is utilized to construct the dedicated 1+1 protection. The results indicate a gradual increase in total cost for the 0.999 availability requirement as the P_B for peak bandwidth increases from 0 to 0.1. Conversely, for higher availability requirements (0.9999 and 0.99999), the total cost remains relatively consistent across varying P_B from 0.001 to 0.1. This trend can be attributed to the increased resource provisioning necessary to meet stringent availability targets. The dedicated 1+1 protection scheme consistently exhibits a constant cost of about 26,100 cost units, regardless of the specified availability requirement. This highlights the fixed overhead associated with implementing such a protection mechanism. When plotting the establishment cost against different availability requirements for various blocking probabilities (0.9 to 0.999999), a clear correlation emerges. As the desired availability increases (shown in Fig. 3), the corresponding resource provisioning and associated costs also rise. This demonstrates the inherent trade-off between service assurance and cost optimization in network slicing.

Figures 4, 5, and 6 illustrate the bandwidth consumption

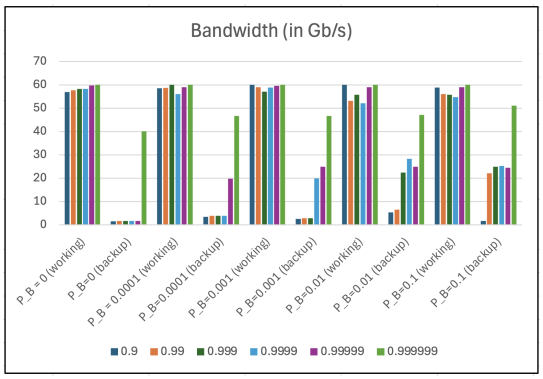

Fig. 5. Bandwidth versus Availability Requirement for different P_B values.

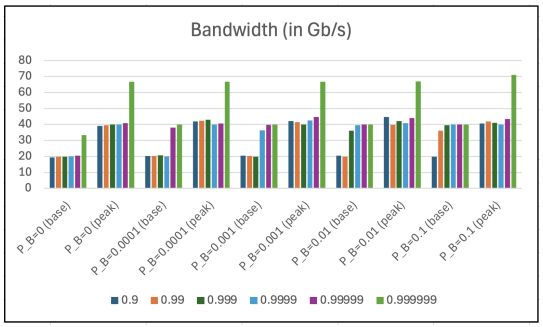

Fig. 6. Bandwidth versus Availability Requirement for different P_B values.

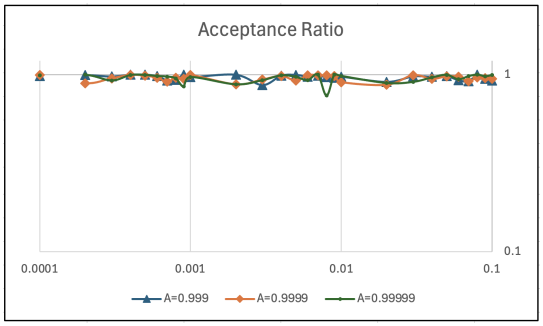

Fig. 7. Acceptance Ratio versus P_B.

patterns for base and peak traffic under various availability requirements (0.999, 0.9999, and 0.99999 in Fig. 4 and 0.9 to 0.999999 in Fig. 5 and Fig. 6). To visualize the total bandwidth utilized on both primary and backup paths for P_B ranging from 0 to 0.1 along the x-axis plotted against total bandwidth units utilized on the y-axis. For availability requirements of 0.9999 and 0.99999, the bandwidth utilization remains relatively constant across different P_B ranging from 0.001 to 0.1. At the 0.999 availability level, a noticeable increase in bandwidth utilization is observed for P_B between 0.001 and 0.01. This indicates a direct correlation between lower availability requirements and increased resource allocation. As the desired availability level increases, there is a corresponding rise in bandwidth allocation to the backup path. For the highest availability requirement (0.999999), the bandwidth utilization remains consistent even at higher P_B values, demonstrating the system's ability to effectively manage resources. Overall, the analysis reveals a trade-off between bandwidth utilization and

availability requirements. Higher availability levels necessitate increased resource allocation, including on backup paths, to ensure service continuity.

In Fig. 7, we compare the acceptance ratio of new slice requests for the availability requirements of 0.999, 0.9999, and 0.99999 against the P_B for the range of 0 to 0.1. The acceptance ratio plotted is calculated as the ratio between the number of requests accepted for the network to the total number of requests. The acceptance ratio consistently exceeds 90%. Variations observed in the graph are attributed to the stochastic nature of the 1000 requests, where specific links may experience overutilization, resulting in network bottlenecks and subsequent request rejections.

We also conducted experiments when the availability of the underlying physical components was 0.999999. The cost, bandwidth, and acceptance ratio followed a similar trend to the graphs shown in Fig. 2 to Fig. 7, where the physical component availability is 0.9999999. The cost increases as a proportion to the usage of the bandwidth, thus irrespective of the number of nodes in the network the trend should depend on the number of requests in the system.

VI. Conclusion

We focused our work on optimizing network slice composition for SFCs with varying bandwidth requirements. Our primary objective is to minimize the total cost of slice establishment while ensuring adherence to specified availability constraints. We adopted a flexible approach to slice topology design, allowing multiple paths (potentially non-disjoint) to enhance resilience. We calculated the necessary resource allocation for each path within the slice on both computing nodes (for VNF deployment) and physical links (for bandwidth). The effectiveness of our proposed approach was evaluated based on the total cost incurred by the network operator. By optimizing slice composition and resource allocation, our approach demonstrated the potential to significantly reduce the cost of deploying SFCs while maintaining desired availability levels. This contributes to the development of more efficient and cost-effective network-slicing solutions. Future work will focus on incorporating an admission control policy for new requests. This policy will aim to balance accepting incoming requests with maintaining scalability during peak demand, ensuring that the P_B remains within a desired threshold.

Acknowledgment

This work was supported in part by the National Science Foundation under Grant No. CNS-2008856.

References

[1] R. Gour, G. Ishigaki, J. Kong, and J. P. Jue, "Availability-guaranteed slice composition for service function chains in 5G transport networks," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 13, no. 3, pp. 14–24, 2021.

[2] B. Shariati, L. Velasco, J.-J. Pedreno-Manresa, A. Dochhan, R. Casellas, A. Muqaddas, O. Gonzalez de Dios, L. Luque Canto, B. Lent, J. E. Lopez de Vergara, S. Lopez-Buedo, F. Moreno, P. Pavon, M. Ruiz, S. K. Patri, A. Giorgetti, F. Cugini, A. Sgambelluri, R. Nejabati, D. Simeonidou, R.-P. Braun, A. Autenrieth, J.-P. Elbers, J. K. Fischer, and R. Freund, "Demonstration of latency-aware 5G network slicing on optical metro networks," *Journal of Optical Communications and Networking*, vol. 14, no. 1, pp. A81–A90, 2022.

[3] Y. Wang, L. Kong, M. Zhu, J. Gu, Y. Cai, and J. Zhang, "Availability-aware and delay-sensitive RAN slicing mapping based on deep reinforcement learning in elastic optical networks," *IEEE Transactions on Network and Service Management*, pp. 1–1, 2024.

[4] Y. Wang, M. Zhu, X. Cai, J. Gu, X. Liu, and J. Zhang, "DRL-assisted fine-grained function placement and routing of 5G RAN slice with reuse scheme in elastic optical networks," in *ICC 2023 - IEEE International Conference on Communications*, 2023, pp. 1958–1963.

[5] H. Cao, S. Garg, G. Kaddoum, M. M. Hassan, and S. A. AlQahtani, "Intelligent virtual resource allocation of QoS-guaranteed slices in b5G-enabled VANETs for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 704–19 713, 2022.

[6] Q. Tian, S. Li, F. Wang, R. Gao, H. Yao, F. Tian, L. Yang, Q. Zhang, and X. Xin, "Elastic adaptive network slicing scheme based on multi-priority cooperative prediction in Fi-Wi access network," *Journal of Lightwave Technology*, vol. 41, no. 2, pp. 396–403, 2023.

[7] Y. Wang, M. Zhu, J. Gu, X. Liu, W. Tong, B. Hua, M. Lei, Y. Cai, and J. Zhang, "Security-aware 5G RAN slice mapping with tiered isolation in physical-layer secured metro-aggregation elastic optical networks using heuristic-assisted DRL," *Journal of Optical Communications and Networking*, vol. 15, no. 12, pp. 969–984, 2023.

[8] Z. Wu and J. P. Jue, "A reinforcement learning-based routing strategy for elastic network slices," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 5505–5510.

[9] S. Saxena and K. M. Sivalingam, "Slice admission control using overbooking for enhancing provider revenue in 5G networks," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, 2022, pp. 1–7.

[10] Z. Min, Z. Liu, Z. Kang, S. Zhou, S. Gokhale, S. Shekhar, C. Mahmoudi, and A. Gokhale, "A novel 5G digital twin approach for traffic prediction and elastic network slice management," in *2024 16th International Conference on COMmunication Systems NETworkS (COMSNETS)*, 2024, pp. 497–505.

[11] S. Saxena and K. M. Sivalingam, "DRL-based slice admission using overbooking in 5G networks," *IEEE Open Journal of the Communications Society*, vol. 4, pp. 29–45, 2023.

[12] J. Y. Yen, "Finding the K shortest loopless paths in a network," 2007.

[13] J. Kong, I. Kim, X. Wang, Q. Zhang, H. C. Cankaya, W. Xie, T. Ikeuchi, and J. P. Jue, "Guaranteed-availability network function virtualization with network protection and VNF replication," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec 2017, pp. 1–6.