Towards Assessing Effects of Isolation on Determinism in Multi-Application Scenarios

Stanislav Lange, Marija Gajić,
Thomas Zinner
first.last@ntnu.no
NTNU
Norway

Jane Frances Pajo, Håkon
Lønsethagen, Min Xie
first.last@telenor.com
Telenor Research
Norway

Ricard Vilalta
ricard.vilalta@cttc.es
CTTC
Spain

ABSTRACT

In addition to the already present heterogeneity in the network service and application ecosystem, new challenges come in the form of emerging applications which require deterministic delay performance characteristics that go beyond traditional SLAs. While hard network slicing with strict isolation is a promising candidate for addressing such requirements, it might be overly restrictive and hard to scale, hence leading to an inefficient use of available resources.

In contrast, soft slicing can help balancing isolation and determinism by allowing a configurable degree of resource sharing between slices. In this work, we follow a simulation-driven approach to study and quantify the trade-offs between variable isolation and the degree of delay-related determinism in a single-hop environment with multiple applications. The results from this scenario will serve as a foundation towards assessing the end-to-end problem.

CCS CONCEPTS

• **Networks** → **Network simulations**; **Programmable networks**; **Network performance analysis**.

KEYWORDS

Slice Isolation, Soft Slicing, Determinism, Multi-Application.

ACM Reference Format:
Stanislav Lange, Marija Gajić, Thomas Zinner, Jane Frances Pajo, Håkon Lønsethagen, Min Xie, and Ricard Vilalta. 2022. Towards Assessing Effects of Isolation on Determinism in Multi-Application Scenarios. In *FIRA '22: ACM SIGCOMM 2022 Workshop on Future of Internet Routing & Addressing (FIRA '22), August 22, 2022, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3527974.3545724

1 INTRODUCTION

Today's networks need to support a multitude of heterogeneous applications whose requirements not only cover a wide range of Quality of Service (QoS) factors and ranges, but also differ w.r.t. the degree of determinism that is associated with these requirements.

With the emergence of new vertical applications such as autonomous vehicles and remote surgery, those requirements - specifically in terms of delay - are becoming increasingly more stringent. In addition, the requirements can include different degrees of determinism ranging from guarantees for delay and jitter as in Time Sensitive Networking (TSN), to more traditional Service Level Agreement (SLA) promises regarding quantiles like five-nines or averages of the delay performance.

Hard network slicing with strict isolation can help addressing this broad range of requirements by reserving and allocating a fixed amount of available resources to specific applications. This results in a high degree of determinism since isolated applications can neither access each other's resources nor interfere with each other. However, hard slicing limits the possibility to take advantage of inter-slice multiplexing gains and consequently achieve higher resource and energy efficiency.

In contrast to hard network slicing, soft slicing allows for a variable degree of isolation which enables resource sharing between slices. However, this could also potentially reduce determinism due to interference, particularly in very dynamic environments and highly delay-sensitive applications.

To quantify these trade-offs, we investigate the interplay between the isolation level and determinism in a multi-application setting. We employ a single-hop network topology as a starting point towards end-to-end analyses and routing schemes that are aware of the aforementioned trade-offs. We evaluate determinism in terms of the following key metrics: mean, five-nines, and maximum end-to-end delay, as well as jitter. The effects of isolation on determinism are quantified via simulations using the OMNeT++ / INET framework and the Hierarchical Token Bucket (HTB) for emulating soft network slicing behavior. The HTB configurations allow for varying the amount of shared resources and therefore the degree of isolation. It is important to note that while HTB might not necessarily be used for network slicing in production environments, it is a useful tool to evaluate the implications of soft network slicing on the traffic characteristics of the services.

The rest of the paper is structured as follows. Section 2 provides an overview of related work, while Section 3 outlines the problem context alongside key components, metrics, and the simulation environment that we use for our evaluations. In Section 4, we identify the main impact factors in the relationship between isolation-related parameters and KPIs, and quantify the resulting trade-offs. Finally, we conclude the work in Section 5 and discuss directions for future work.

FIRA '22, August 22, 2022, Amsterdam, Netherlands

ACM ISBN 978-1-4503-9328-7/22/08.
https://doi.org/10.1145/3527974.3545724

2 RELATED WORK

Network slicing is a multi-faceted topic in the networking domain and has several degrees of freedom when it comes to its scope and concrete implementation [3]. In this section, we focus on isolation-related characteristics of network slicing and first present an overview of aspects, types, and scopes of isolation. We then cover works that deal with evaluating the performance of slice isolation mechanisms. Finally, we discuss slicing and slice isolation approaches that are related to hierarchical QoS (hQoS) and therefore have configuration parameters similar to our HTB-based emulated slices.

Different degrees of network slice isolation are an important parameter for 5G and beyond network technologies. In this context, isolation can include aspects of performance, security, and dependability [10]. Furthermore, a wide spectrum of isolation types is actively discussed and standardized [11]. These isolation types range from service, process, and Virtual Network Function (VNF) isolation to virtual resource and physical isolation. The notion of isolation can also extend over different scopes like RAN slicing [18], core and transport network slicing [20], as well as end-to-end slicing [12, 16].

In [12], an overview of isolation capabilities and approaches for their realization is presented. The authors suggest that the level and strength of isolation comes with a set of properties and should be chosen based on the needs of a particular use case. Furthermore, an architecture showcasing the capability to provide traffic isolation while at the same time satisfying different service requirements like bandwidth and delay within transport networks is presented in [7]. Additional evaluations of isolation performance in terms of traffic rate and end-to-end latency have been conducted in [9] using iperf-based traffic sources. At a higher level of abstraction, [15] and [22] use simulations where packet interarrivals are generated via Poisson processes. In contrast, we employ a packet-level simulator that covers both application-layer logic as well as protocol behavior.

Similarly to the discussions related to isolation, meeting the QoS targets of individual slices also poses a challenge that can be addressed by different means and at different scales. For instance, multi-queue QoS mechanisms can be employed at individual network elements [5], whereas network virtualization hypervisors can control the sharing of the physical infrastructure between several independent tenants [4]. Additionally, acceleration mechanisms can speed up packet processors along a packet's way through the network [17], and scheduling mechanisms can manage compute resources of VNF instances [13]. In our previous work [14], we provided initial evidence that applying mechanisms such as prioritization or slicing with hard isolation in a QoS-aware manner can improve resource efficiency while meeting stringent delay requirements of heterogeneous applications. However, the investigated mechanisms did not offer the option to fine-tune the degree of isolation and the simulations have been conducted at a high level of abstraction, ignoring application- and protocol-level behavior. We address these shortcomings by using the packet-level simulator OMNeT++ in conjunction with the INET framework as well as an HTB-based slice isolation scheme.

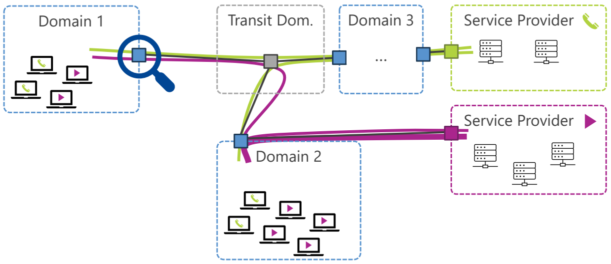


Figure 1: Inter-domain traffic aggregation. Heterogeneous traffic can be aggregated in different ways at each station, resulting in trade-offs w.r.t. performance, costs, and scalability. The focus of this paper is on the behavior within a single element as a starting point for E2E analyses.

The HTB enables two-level bitrate control of all nodes that are arranged in a tree-based hierarchy. Each node in the HTB tree has its assigned guaranteed and maximum bitrate. Other mechanisms from the family of hierarchical QoS have been used previously to derive slicing and slice isolation mechanisms [8, 19]. The approach for HTB-based slice emulation was used in [5], where authors provided evidence that multiplexing gains from intra-slice sharing can improve resource efficiency in multi-application contexts with strictly isolated slices. In contrast to the per-flow bitrate guarantees that were considered in [5], in this paper we perform only per-slice QoS treatment, but allow different degrees of inter-slice borrowing.

3 METHODOLOGY

In this section, we first provide the wider problem context of QoS-aware inter-domain connectivity with multiple applications and discuss available mechanisms for differentiated traffic treatment and resource allocation. From that, we derive a minimal scenario that allows us to study the interplay between isolation and delay performance in a single-hop setting that is the focus of this work. The scenario description covers the chosen HTB-based resource allocation scheme, the measures for quantifying the degree of resource sharing and isolation, as well as the overall simulation setup and applications.

3.1 QoS-Aware Inter-Domain Connectivity

Figure 1 displays an exemplary multi-domain, multi-application setup that features spatially distributed application clients that interact with application servers located at service providers' premises. Depending on the chosen E2E path, packets can traverse black-box transit domains as well as different administrative domains with potential (re-)aggregation of traffic flows at domain boundaries. Hence, when choosing an E2E path, operators not only have to consider the path itself, but also different mechanisms for aggregating flows that can have a significant impact on the QoS performance. Aggregation options include treating all traffic in a best effort manner, using prioritization, and network slicing with hard or soft isolation. Each of these options represents a trade-off regarding isolation, resource efficiency, deterministic QoS performance, costs, as well as (control plane) complexity and scalability. Since this work revolves around the impact of isolation on determinism, we mainly focus on slicing with soft isolation as it allows controlling the degree of isolation whereas the other listed options have either implicit or binary forms of isolation.

3.1.1 Per-Hop Isolation and Delay. A key challenge for operators in the outlined scenario is to reap multiplexing gains at each domain traversal, but without sacrificing the required QoS characteristics of affected application flows. However, before tackling the problem in an E2E fashion, a deep understanding of the effects and trade-offs between different degrees of isolation on a per-hop basis is required. Hence, we propose the single-hop setup displayed in the top part of Figure 2 which allows for a quantitative analysis of such trade-offs.

The setup consists of a single link that interconnects clients who use Voice over IP (VoIP) and Video on Demand (VID) applications with the corresponding application servers. In the following subsections, we elaborate how HTB is used to mimic slicing with soft isolation and establish a relationship between HTB parameters and measures for quantifying the amount of shared resources and therefore the degree of isolation. Furthermore, we provide details about the applications as well as the simulation environment and parameters.

3.2 Hierarchical Token Bucket (HTB)

The central data structure for configuring hQoS in HTB is the HTB tree, an example of which is displayed in the bottom part of Figure 2. Starting at the root node which represents the entire link, each node in the tree has two parameters that control the Guaranteed Bitrate (GBR) and Maximum Bitrate (MBR) of the matching traffic, respectively. By arranging the nodes in this tree and setting appropriate GBR and MBR values, the degree of resource sharing (“borrowing”) and therefore isolation can be adjusted.

On arrival, packets are classified based on their header information and stored in the queues of the corresponding leaf nodes. Afterwards, packets are extracted from the queue according to a Deficient Round Robin (DRR) queuing discipline and sent out. In this work, we use the HTBQueue module for OMNeT++ / INET from [6] which follows the Linux implementation of HTB.

It is worth noting that the HTB offers additional parameters such as burst sizes and DRR quantum granularity for fine-tuning its behavior. However, we limit our studies to the primary impact factors regarding the capacity split and leave other parameters at their default values. Finally, although HTB allows deeper tree structures that could be used to provide even per-flow QoS, we argue that per-flow QoS would not be a scalable solution for E2E scenarios and hence focus on per-traffic-type resource allocation.

3.2.1 Resource Sharing Measures. Our main interest in this work lies in the relationship between the degree of isolation and various network KPIs. Hence, we introduce two measures that allow quantifying different aspects of isolation or, equivalently, the degree of resource sharing. Figure 3 illustrates how the HTB parameters GBR and MBR relate to the amount of shared capacity and degree of contention in case of two competing applications, VoIP and VID, on a single link with a given capacity.

Using the total available capacity c and the GBR / MBR parameters, we define two measures $S = 1 - \sum_i \frac{GBR_i}{c}$ and $S' = \sum_i \frac{MBR_i}{c} - 1$ which are also highlighted in the figure. While S depends on GBRs and denotes the amount of potentially contested link capacity, S' considers the MBRs and reflects the actual overlap between configured HTB capacities. Note that an overlap between the MBR of one traffic type with the GBR of the other is also possible.

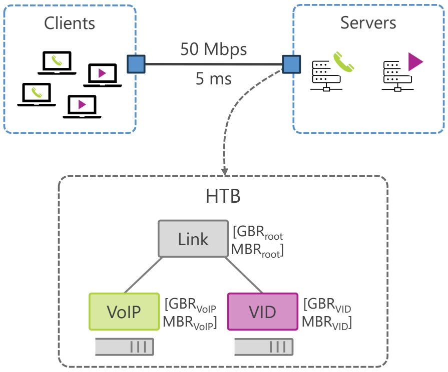


Figure 2: Top: simplified simulation setup for investigating the relationship between isolation and delay performance in a single-hop scenario. Bottom: HTB tree and parameters that control resource sharing.

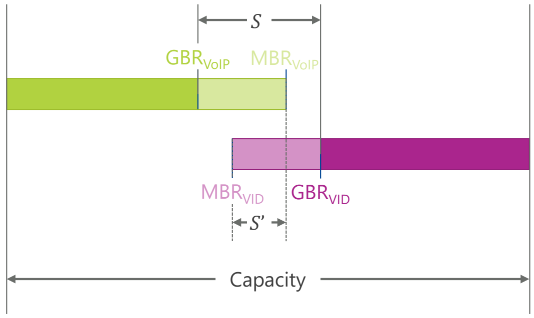


Figure 3: Bandwidth sharing relationships when two traffic types are configured, with particular focus on the relationship between GBR and MBR HTB settings and isolation / contention indicators S and S'.

3.3 Applications and Simulation Parameters

As indicated in Figure 2, we mix traffic that is generated by clients of two applications. These include a TCP-based adaptive video streaming application (VID) as well as a UDP-based VoIP application. For each application client, we can collect packet-level statistics such as E2E delays that are used to compute means, quantiles, and the variation of packet delays. Additionally, per-session aggregated metrics like the Mean Opinion Score (MOS) are obtained via well-established Quality of Experience (QoE) models, namely the ITU-T P.1203 model [2] in case of VID and the ITU-T G.107 e-model [1] in case of VoIP.

In order to emulate emerging and more demanding applications such as the ones involving haptic feedback [23], we modify the VoIP application to increase its bandwidth usage and delay sensitivity by

Table 1: Simulation parameters.

Parameter	Value(s)
Link capacity	50 Mbps
Link delay	5 ms
Simulation time	400 s
Video duration	Uniform(280, 320) s
Number of clients per application	$\{12, 16, 20, 24\}$
(p, q)	$\{0, 1, 2, \ldots, 10\}^2$
GBR for VoIP traffic	$(5.5 - 0.5 \cdot q)$ Mbps
MBR for VoIP traffic	$(5.5 + 0.5 \cdot p)$ Mbps
GBR for video traffic	$(44.5 - 0.5 \cdot q)$ Mbps
MBR for video traffic	$(44.5 + 0.5 \cdot p)$ Mbps
Number of repetitions per configuration	5
Total number of simulation runs	2,420

a factor of 10 each. To this end, we decrease the inter-packet delay to reach a bandwidth consumption of 300 kbps when a client is active and feed the QoE model inflated delay values. Using the default parameters for the duration of talk spurts and silence periods[1], VoIP clients are active roughly 65 % of the time and therefore consume an average of 195 kbps.

While each VoIP client generates UDP packets at a constant bitrate when active, adaptive video streaming clients are more dynamic and make decisions about the quality level of downloaded video segments based on their buffer fill level. Hence, their bandwidth is less predictable and can range from few hundred kbps to multiple Mbps, corresponding to resolutions between 480p and 1080p. As a reference, the average bitrate at 720p equals 1,600 kbps in our setup.

An overview of all simulation parameters is provided in Table 1. Given a fixed link capacity and baseline delay, we vary the load that is offered to the system by gradually increasing the number of clients per application. To investigate the effects of isolation on the delay performance and general system characteristics like the overall link capacity usage, we first determine a split between the two application types based on their expected bandwidth consumption. Given the previously mentioned mean bitrates, $\frac{195}{195+1600} \simeq 11\,\%$ of the link capacity (5.5 Mbps) is used as a starting point for VoIP traffic and the remaining 44.5 Mbps for VID. From these reference points, both GBR and MBR values are varied to cover all combinations that are listed in the table, representing a wide range of isolation and resource sharing conditions. We perform five repetitions per parameter combination to obtain statistically significant results.

4 EVALUATION

In this work, we are mainly interested in evaluating how and to which extent different levels of isolation that are controlled by the GBR and MBR parameters impact different delay, QoE, and system-level KPIs. To this end, we first identify key factors that affect the respective KPIs by means of a correlation- and ANOVA-based analysis of main effects. Then, we study the actual KPI values in-depth in order to quantify their sensitivity to our isolation and resource sharing measures.

4.1 Analysis of Main Effects

Table 2 displays the results of a correlation- and ANOVA-based analysis of isolation-related HTB parameters which directly impact the amount of remaining and shared capacity between traffic types. The effect of each of the two parameters on a total of six KPIs is studied. These include the mean, five-nines, and maximum delay of the VoIP traffic, the variance of its delay as a measure of jitter, as well as the QoE-related average MOS value among VoIP clients and finally, the mean link capacity usage as an indicator of resource efficiency. In addition to calculating Spearman's correlation coefficient r_s between each pair of metric and parameter, we additionally perform an Analysis of Variance (ANOVA) and report the resulting values of the F statistic alongside the corresponding p-values.

While both GBR and MBR values have a statistically significant impact on all delay and QoE-related metrics, the correlations and F values suggest that the guaranteed bitrate is the main influence factor on these metrics. This is in line with the fact that a higher GBR corresponds to a higher degree of isolation and therefore minimizing the interference that can occur. In case of the mean link capacity usage, the MBR value is identified to have the larger impact. Again, this matches the expectation that a higher degree of resource sharing leads to a more efficient use of available resources, i.e., temporarily idle resources of one application can be utilized by another.

4.2 Impact of Isolation on Determinism

Based on the results of the main effects study, we perform an in-depth analysis of the relationship between the resource sharing measures S and S' and the individual KPIs at different load levels.

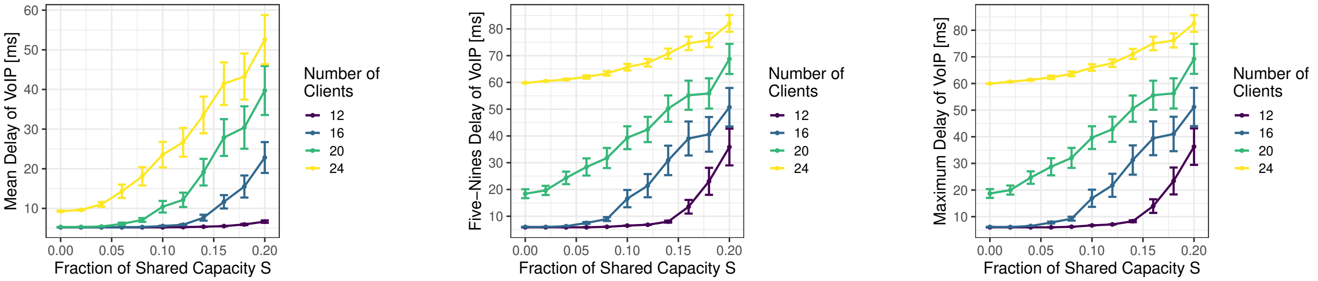
The first row of Figure 4 displays the results for the three delay metrics, i.e., mean, five-nines, and maximum delay experienced by the VoIP clients. Given the fraction of shared link capacity S on the x-axis, the y-axis shows the corresponding metric, while differently colored curves denote different load levels in terms of the number of clients per application. Furthermore, error bars display 95 % confidence intervals. As suggested by the results of the previous analysis, a positive correlation between S and the delay metrics can be observed. Additionally, the strictness of the delay requirements dictates the sensitivity, e.g., the five-nines delay deteriorates significantly earlier than the mean delay. For instance in case of 12 clients, the mean delay of VoIP clients remains almost unchanged even up to $S = 0.2$ whereas a sharp increase starts at $S = 0.16$ in case of the five-nines delay. We can observe in Figure 4d that for the same number of clients, the variance is the delay metric that deteriorates earliest - around the $S = 0.12, 0.14$ mark - and is therefore most sensitive to interference. While the KPIs discussed so far are related to pure QoS, Figure 4e also provides a QoE perspective. Since the default QoE model is based on measured delays per talk spurt, the

---

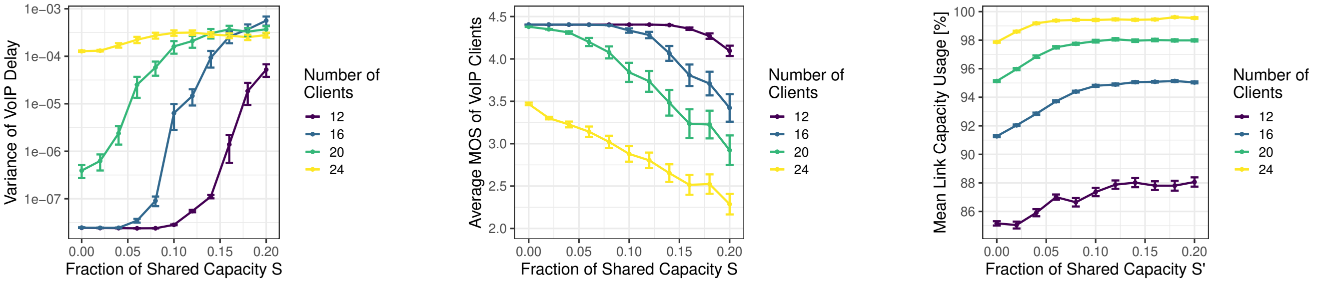
[1] https://doc.omnetpp.org/inet/api-current/neddoc/inet.applications.voip.SimpleVoipSender.html

Table 2: Correlation- and ANOVA-based analysis of isolation-related parameters and their effect on delay performance and link capacity usage. While r_s denotes Spearman's coefficient of correlation, F and p represent the value of the ANOVA test statistic and the corresponding p-value, respectively.

Parameter / **Metric**	GBR_{VoIP}			MBR_{VoIP}		
	r_s	F	p	r_s	F	p
Mean delay of VoIP traffic	-0.48	367.84	**< 0.001**	0.08	12.86	**< 0.001**
Five-nines delay of VoIP traffic	-0.43	316.20	**< 0.001**	-0.07	22.71	**< 0.001**
Maximum delay of VoIP traffic	-0.43	320.79	**< 0.001**	-0.07	22.38	**< 0.001**
Variance of delay of VoIP traffic	-0.43	257.54	**< 0.001**	-0.01	36.87	**< 0.001**
MOS of VoIP clients	0.42	350.94	**< 0.001**	-0.04	11.34	**< 0.001**
Mean link capacity usage	-0.00	0.08	0.78	0.20	73.01	**< 0.001**



(a) Mean delay of VoIP. **(b) Five-nines delay of VoIP.** **(c) Maximum delay of VoIP.**



(d) Variance of VoIP delay. **(e) Average MOS of VoIP clients.** **(f) Mean link capacity usage.**

Figure 4: Impact of isolation-related measures and number of clients on delay and QoE performance of delay-sensitive traffic and link capacity usage. Error bars denote 95 % confidence intervals.

previously observed delay increases also lead to a degradation of MOS values.

Finally, the mean link utilization has been identified to be mostly affected by the maximum bitrate and hence its relation to the S' indicator is shown in 4f. The results illustrate that depending on the baseline load level, an increase of S' from 0 to 0.2 leads to a steady increase of the mean utilization. Differences between the two extremes range from 2 % utilization at the highest load level to roughly 4 % with 16 clients per application. Such medium-load scenarios offer the most potential for multiplexing gains since high loads already have a high utilization to begin with and clients tend to have enough resources without sharing at low loads.

In summary, our simulations not only allow determining the qualitative effect of isolation-related parameters on QoS, QoE, and resource usage characteristics, but also enable us to quantitatively assess the relationships. Using such an approach in conjunction with application-specific requirements can help finding system configurations that hit the sweet spot between resource utilization and application performance.

5 CONCLUSION

Network operators need to cope with heterogeneity on many levels. These include services and applications with diverse and increasingly stringent performance requirements, end-user devices and preferences, as well as network devices and their capabilities. It is particularly challenging to meet delay performance requirements on a shared medium due to potential interference and different degrees of required and expected determinism, i.e., TSN-like guarantees, SLA-like promises, or plain best effort. Furthermore, operators would still like to benefit from multiplexing gains rather than performing hard slicing and fully isolating each and every flow or service, which is neither cost- nor resource-efficient.

In this work, we have explored how soft slicing as emulated via HTB can strike a balance between isolation and determinism, focusing on a specific application mix and a single-hop scenario. The proposed simulation-based methodology allows quantifying different delay performance characteristics that represent varying degrees of determinism and how they deteriorate when isolation is decreased.

Our evaluations illustrate that QoS-related KPIs can be characterized by their sensitivity to the degree of isolation. Hence, depending on the requirements of a specific application, there can be room for optimization in multi-application environments in order to neither under- nor overshare resources while maintaining a high level of user satisfaction.

Directions for future work include extending the repertoire of considered application profiles and traffic mixes as well as validating our findings in a physical testbed. Additionally, insights from the single-hop scenario can be used to widen the scope towards more complex network topologies where packets traverse multiple devices along an E2E path.

Another further step is the introduction of the concepts proposed in this article to the ETSI TeraFlowSDN controller [21]. The inter-domain slice management can benefit from the introduction of the proposed slice model with SLA.

ACKNOWLEDGMENTS

This work has been performed in the framework of EC H2020 TeraFlow (101015857).

REFERENCES

[1] ITU-T Study Group 12. 2015. *ITU-T Recommendation G.107: The E-model: a computational model for use in transmission planning.* Technical Report. International Telecommunication Union.

[2] ITU-T Study Group 12. 2017. *ITU-T Recommendation P.1203: Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport.* Technical Report. International Telecommunication Union.

[3] Ibrahim Afolabi, Tarik Taleb, Konstantinos Samdanis, Adlen Ksentini, and Hannu Flinck. 2018. Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. *IEEE Communications Surveys & Tutorials* 20, 3 (2018), 2429–2453.

[4] Andreas Blenk, Arsany Basta, Martin Reisslein, and Wolfgang Kellerer. 2015. Survey on network virtualization hypervisors for software defined networking. *IEEE Communications Surveys & Tutorials* 18, 1 (2015), 655–685.

[5] Marcin Bosk, Marija Gajić, Susanna Schwarzmann, Stanislav Lange, Riccardo Trivisonno, Clarissa Marquezan, and Thomas Zinner. 2021. Using 5G QoS Mechanisms to Achieve QoE-Aware Resource Allocation. In *International Conference on Network and Service Management (CNSM).* IEEE.

[6] Marcin Bosk, Marija Gajić, Susanna Schwarzmann, Stanislav Lange, and Thomas Zinner. 2021. HTBQueue: A Hierarchical Token Bucket Implementation for the OMNeT++/INET Framework. In *OMNeT++ Community Summit.*

[7] Chia-Yu Chang, Manuel A Jiménez, Molka Gharbaoui, Javier Sacido, Fabio Ubaldi, Chrysa Papagianni, Aitor Zabala, Luca Valcarenghi, Davide Scano, Konstantin Tomakh, et al. 2021. Slice Isolation for 5G Transport Networks. In *International Conference on Network Softwarization (NetSoft).* IEEE, 366–368.

[8] Ferenc Fejes, Szilveszter Nádas, Gergő Gombos, and Sándor Laki. 2022. DeepQoS: Core-Stateless Hierarchical QoS in Programmable Switches. *IEEE Transactions on Network and Service Management* (2022).

[9] Alessio Giorgetti, Davide Scano, and Luca Valcarenghi. 2021. Guaranteeing Slice Performance Isolation With SDN. *Communications Letters* 25, 11 (2021), 3537–3541.

[10] Andres J Gonzalez, Jose Ordonez-Lucena, Bjarne E Helvik, Gianfranco Nencioni, Min Xie, Diego R Lopez, and Pål Grønsund. 2020. The Isolation Concept in the 5G Network Slicing. In *European Conference on Networks and Communications (EuCNC).* IEEE, 12–16.

[11] Aihua Guo, Luis M. Contreras, Sergio Belotti, Reza Rokui, Yunbin Xu, Yang Zhao, and Xufeng Liu. 2022. *Framework and Data Model for OTN Network Slicing.* Internet-Draft draft-ietf-ccamp-yang-otn-slicing-01. Internet Engineering Task Force. https://datatracker.ietf.org/doc/html/draft-ietf-ccamp-yang-otn-slicing-01 Work in Progress.

[12] Zbigniew Kotulski, Tomasz Nowak, Mariusz Sepczuk, Marcin Tunia, Rafal Artych, Krzysztof Bocianiak, Tomasz Osko, and Jean-Philippe Wary. 2017. On end-to-end approach for slice isolation in 5G networks. Fundamental challenges. In *Federated conference on computer science and information systems (FedCSIS).* IEEE, 783–792.

[13] Sameer G Kulkarni, Wei Zhang, Jinho Hwang, Shriram Rajagopalan, KK Ramakrishnan, Timothy Wood, Mayutan Arumaithurai, and Xiaoming Fu. 2020. NFVnice: Dynamic backpressure and scheduling for NFV service chains. *IEEE/ACM Transactions on Networking* 28, 2 (2020), 639–652.

[14] Stanislav Lange, Jane Pajo Frances, Thomas Zinner, Hakon Lonsethagen, and Min Xie. 2022. QoS-Aware Inter-Domain Connectivity: Control Plane Design and Operational Considerations. In *International workshop on Fully-Flexible Internet Architectures and Protocols for the Next-Generation Tactile Internet (FlexNGIA).*

[15] Jieyu Li, Jiang Liu, Tao Huang, and Yunjie Liu. 2020. DRA-IG: the balance of performance isolation and resource utilization efficiency in network slicing. In *International Conference on Communications (ICC).* IEEE, 1–6.

[16] Xu Li, Rui Ni, Jun Chen, Yibo Lyu, Zhichao Rong, and Rui Du. 2020. End-to-end network slicing in radio access network, transport network and core network domains. *IEEE Access* 8 (2020), 29525–29537.

[17] Leonardo Linguaglossa, Stanislav Lange, Salvatore Pontarelli, Gábor Rétvári, Dario Rossi, Thomas Zinner, Roberto Bifulco, Michael Jarschel, and Giuseppe Bianchi. 2019. Survey of performance acceleration techniques for network function virtualization. *Proc. IEEE* 107, 4 (2019), 746–764.

[18] Dania Marabissi and Romano Fantacci. 2019. Highly flexible RAN slicing approach to manage isolation, priority, efficiency. *IEEE Access* 7 (2019), 97130–97142.

[19] Sebastien Martin, Paolo Medagliani, and Jeremie Leguay. 2021. Network Slicing for Deterministic Latency. In *2021 17th International Conference on Network and Service Management (CNSM).* IEEE, 572–577.

[20] Davide Scano, Luca Valcarenghi, Koteswararao Kondepu, Piero Castoldi, and Alessio Giorgetti. 2020. Network slicing in SDN networks. In *International Conference on Transparent Optical Networks (ICTON).* IEEE, 1–4.

[21] Ricard Vilalta, Raul Muñoz, Ramon Casellas, Ricardo Martínez, Victor López, Oscar González de Dios, Antonio Pastor, Georgios P Katsikas, Felix Klaedtke, Paolo Monti, et al. 2021. TeraFlow: Secured Autonomic Traffic Management for a Tera of SDN Flows. In *Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit).* IEEE, 377–382.

[22] Qiang Ye, Junling Li, Kaige Qu, Weihua Zhuang, Xuemin Sherman Shen, and Xu Li. 2018. End-to-end quality of service in 5G networks: Examining the effectiveness of a network slicing framework. *Vehicular Technology Magazine* 13, 2 (2018), 65–74.

[23] Qi Zhang, Jianhui Liu, and Guodong Zhao. 2018. Towards 5G Enabled Tactile Robotic Telesurgery. *arXiv preprint arXiv:1803.03586* (2018).