How to Accommodate Network Slicing & Network Neutrality?

Yassine Hadjadj-Aoul, Patrick Maillé, Bruno Tuffin

How to Accommodate Network Slicing & Network Neutrality? A View from the ERMINE Team

Y. Hadjadj-Aoul*, P. Maillé†, B. Tuffin‡

June 30, 2022

Abstract

Network slicing has emerged, with 5G mobile networks, as a response to the increasing networks' complexity and the inherent scaling limitations of traditional methods. That technological building block allows, in fact, a very agile orchestration of services, bringing dynamicity, differentiation, and most importantly the fulfillment of various service constraints.

Slicing is therefore seen as a key characteristic of 5G-and-beyond networks, however it seems in contradiction with neutrality principles promoted worldwide. We detail the two contradictory but considered compulsory notions, and discuss how they can be accommodated.

1 5G & Network Slicing

5G stands for the latest generation of wireless networks, with new capabilities. That generation is indeed expected to increase throughput by a factor up to 100 with respect to 4G, allowing to connect all available devices, from phones to cars and any type of object all over the world. Beyond that, 5G enables the support of a broad range of services, even the most demanding ones, including real-time interactivity, such as autonomous driving. As of February 2021, 30% of countries had already implemented a 5G network, and it is believed that in 2025, at least 3.6 billion 5G connections will be active. While the first phases of deployment are currently limited to access networks, 5G is designed to go far beyond that, with the ability to provide on-demand guaranteed end-to-end services. This feature is made possible with the adoption of a new concept in 5G and beyond networks, named *network slicing*.

Network slicing consists in creating multiple dedicated logical and virtualized networks over several domains, cutting the infrastructure into "slices" managed

*University of Rennes, Inria, CNRS, IRISA
†IMT Atlantique/IRISA,
‡Inria, University of Rennes, CNRS, IRISA

independently. That approach provides flexible and scalable resource provisioning for applications and services in order to align resources with needs for quality in terms of throughput, latency, reliability and other metrics, by appropriately dimensioning the slices. Slicing allows to simultaneously and efficiently manage heterogeneous traffic and offer tailored solutions to customers and industries, some being demanding such as telemedicine, online gaming, or augmented reality.

There are several reasons for the adoption of network slicing in 5G mobile networks. One is obviously the support of services regardless of their constraints, as described above. In fact, this is an opportunity for Network Operators (NOs) to get more value out of the transport of services, by being part of the content distribution value chain. Indeed, NOs currently have very limited control over their own infrastructure rental revenues, which are currently being determined by national regulators. Network slicing is also seen as one of the most important building blocks for network automation, as it brings more agility in the management of services. It also helps to deal with the growing complexity of networks. For all these reasons, we believe that network slicing will continue to be part of the next cellular network standards.

2 Network Neutrality

On the other hand, telecommunication networks have been under the scrutiny of regulators concerning neutrality issues. Network neutrality is the principle that traffic should be treated equally, without discrimination, restriction or interference, independently of the sender, receiver, type, content, device, service or application [4]. Due to the historic nature of a free and open Internet, the move to a commercial network ignited fears and threats that some services would not be available to all and/or that some would receive a degraded treatment. According to the rule currently in place in the EU, no differentiation is allowed, with exceptions i) in case of a legal action; ii) to ensure the security and integrity of the network if confronted to attacks; iii) in case of temporary congestion if no commercial consideration is taken into account.

The principle has been highly debated and laws have been passed in most countries worldwide.

3 Contradictions Between the Two Notions and Traditional Propositions

There seems to be a contradiction between the notions of neutrality, now a pillar of the Internet and telecommunication networks, and of slicing, a pillar of 5G networks: how to conciliate equal treatment among flows, with reserved resources for *some* traffic flows?

Regulators such as ARCEP in France or BEREC in Europe are proposing to define some services, called *specialized services*, which could be excluded from neutrality constraints. This would concern services with strict quality of

service requirements, services usually not supported by the Internet network. Typical examples are video on demand, online gaming, autonomous vehicles or telemedicine [1]. Network slicing allows allocating slices to each of these services, hence guaranteeing a sufficient quality, but potentially at the expense of other "regular" services.

This opens some breaches to the implementation of neutrality principles. Indeed, the definition of a specialized service is vague: what prevents a new service and even an existing one from claiming the "specialized service" label? It would also be tempting for an NO to consider its own services (VoIP for example) as specialized ones to obtain a competitive advantage. What are then the objective and clear criteria to be part of that category and to define the quality associated to each slice? Who chooses the relevance of pertaining to a slice, and the service levels associated with that slice? Who will ask for a service to be of this particular class? The Service or Content Provider (SCP), the NO? If it is the NO, isn't there a risk to favor commercial partners? What if an NO does not warrant a request from an SCP? How to proceed without economic considerations, which could constitute incentives to favor some services over others, even between slices and specialized services, something excluded by neutrality principles? Those open questions need some clear answers.

The definition of specialized services is therefore surrounded by many grey areas still to be sorted out, otherwise complaints are expected to pop up.

4 ERMINE's Accommodation Proposition

Given the many remaining questions, our aim in this section is to propose a solution which could accommodate network slicing and network neutrality from Inria's ERMINE team perspective. The idea is to define a procedure as objective and automated as possible, limiting the holes and antagonist principles between slicing and neutrality.

What we propose is made of the following steps.

1. First, analyze the different kinds of traffic, their requirements and constraints, and classify them into different "types", or classes. Classification can be either based on packet-level information, through deep packet inspection, or performed using statistical approaches [5], the latter case being possible even with encrypted traffic [3].

2. Define slices, one per traffic type, in terms of a given level of quality of experience (QoE) [6]. QoE "explains" the users' perceived quality for a given service, in general related to quality of service parameters (packet loss, delays, jitter , ...) but in an extremely complex way. Quantifying the QoE has been the topic of an extensive research activity during the past couple of decades [2].

3. If congestion and as a consequence quality of service degradation occurs, differentiate service to offer a sufficient or satisfying QoE for each slice,

while treating all flows equally within each slice (which may imply interrupting service for a whole slice–or several, or uniformly selected flows–if resources are insufficient).

Proceeding this way, flows with similar quality requirements and perceptions will be treated neutrally, and differentiation will be made possible at the benefits of end users. Users and service providers will be served in an optimal way according to the conditions, and NOs will be allowed to manage services.

Of course, this requires monitoring from regulators to verify that the behavior of actors corresponds to what is expected, here in terms of sufficient QoE. Specific procedures have to be designed toward that goal.

The proposed method allows to respect a new vision of neutrality, oriented toward classes of service, but aiming simultaneously at offering the best possible experience to users, at providing a sufficient quality to SCPs, and at leaving flexibility to NOs. It seems to us a reasonable trade-off for accommodating slicing and neutrality. This obviously requires some knowledge about service requirements and the ability to measure their quality of experience, whether they are encrypted or not. In an over-provisioned system, the operator will have the ability to meet the targeted quality criteria.

In the case of an under-provisioned system, the proposed slicing technique presents some challenges. First, what is the share of resources to be dedicated per class of service? Once we have partitioned the resources, how do we guarantee the quality of the supported services? What incentive is there for a network operator to set up several solutions (i.e., traffic classification, QoE measurement, services' monitoring, etc.)? The answers to these questions are not necessarily unique, but several directions and accommodations exist to achieve these goals while respecting the principles of neutrality.

On the issue of resource sharing, there is the simple possibility of fixed sharing, prior to service deployment and which might be up to the preference of the operators and their interest in favouring one slice or another. But a better solution is, instead, to have a dynamic sharing of resources that maximizes the satisfaction of the users belonging to the different slices.

To guarantee the quality of the flows in a particular slice, there is a clear need to monitor the traffic of this slice in order to provide a fair intra-slice resource sharing (i.e., offering the same QoE). In the case where resources are insufficient to reach the minimum QoE, we can include an admission controller: it could act by randomly selecting some flows which will blocked on this slice to the QoE objectives for those served. The selection would be according to a uniform law to ensure a form of neutrality/fair treatment.

Finally, the proposed solution allows the NOs to offer differentiation in their network, an option they do not have today, even if depriving them from differential control within a slice.

References

[1] IETF. Network Slicing Use Cases: Network Customization and Differentiated Services, June 2017. [Online; accessed 26. Jun. 2022].

[2] M. J. Khokhar, T. Ehlinger, and C. Barakat. From network traffic measurements to qoe for internet video. In *2019 IFIP Networking Conference (IFIP Networking)*, pages 1–9. IEEE, 2019.

[3] C. Liu, L. He, G. Xiong, Z. Cao, and Z. Li. Fs-net: A flow sequence network for encrypted traffic classification. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 1171–1179, 2019.

[4] P. Maillé and Tuffin B. *From Net Neutrality to ICT Neutrality.* Springer, 2022. In press.

[5] T. T.T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys & Tutorials*, 10(4):56–76, 2008.

[6] K.U. Rehman Laghari and K. Connelly. Toward total quality of experience: A qoe model in a communication ecosystem. *IEEE Communications Magazine*, 50(4):58–65, 2012.