Aalto University
School of Business

A Dynamic Allocation Mechanism for Network Slicing as-a-Service

Master's Thesis
Markus Myllymäki
Aalto University School of Business
Economics
Spring 2019

Contents

Glossary of Terms

Abbreviation Used	Full term	Brief explainer
API	Application Programming Interface	Set of definitions, protocols and tools for building software on a platform
B2B	Business-to-Business	Products or services sold to another company directly
B2B2C	Business-to-Business-to-Customer	Products or services sold to another company that uses it to sell a product or service to an end user
B2C	Business-to-Customer	Products or services sold to its end user
CDNaaS	Content Distribution Network as-a-Service	On-demand access to distributed networks and data centers
C-RAN	Cloud Radio Access Network	RAN operated on a set of virtualized network resources
eMBB	enhanced Mobile Broadband	The transfer of user data and messaging via mobile network, e.g. video streaming
IL	Instantiation Level	Different quality iterations of the same network structure
IoT	Internet of Things	Interconnected network of sensors
mMTC	massive Machine-Type Communication	Related to IoT, ubiquitous connection of devices to the network
MNO	Mobile Network Operator	Operate mobile broadband businesses and own spectrum, e.g. Telia, DNA or Elisa
(V)NF	(Virtual) network function	NFV-driven functional building block within a network infrastructure, e.g. a network node
NFV	Network Function Virtualization	Allows for network functions to be software-driven and separate from physical hardware
NS	Network service	Composition of network functions, with at least 1

		VNF. Has a defined performance purpose in the network
QoS	Quality of Service	Quantified measurement of the performance of a network, based on requirements
RAN	Radio Access Network	Located between the core network and user devices, connects devices together
SDN	Software-defined networking	Decouples the data and control plane, allowing for programmable networks
SFC	Service Function Chaining	How ordered functions in a network are routed through different network services
SFP	Service Function Path	Predefined logical paths of how SFC is implemented
TOFaaS	Traffic-Offloading as-a-Service	On-demand management of traffic loads on networks
URLLC	Ultra-Reliable, Low Latency Communication	Mission critical or time critical services, e.g. autonomous vehicles
vMEC	Virtualized Multi-access Edge Computing	Computational resources on the cloud that allow offloading of computation by users

Abstract

In my thesis, I explore the design of a market mechanism to socially efficiently allocate resources for network slicing as-a-Service. Network slicing is a novel usage concept for the upcoming 5G network standard, allowing for isolated and customized virtual networks to operate upon a larger, physical 5G network. By providing network slices as-a-Service, where the users of the network slice do not own any of the underlying resources, a larger range of use cases can be catered to.

My market mechanism is a novel amalgamation of existing mechanism design solutions from economics, and the nascent computer science literature into the technical aspects of network slicing and underlying network virtualization concepts. The existing literature in computer science is focused on the operative aspects of network slicing, while economics literature is incompatible with the unique problems network slicing poses as a market. In this thesis, I bring these two strands of literature together to create a functional allocation mechanism for the network slice market.

I successfully create this market mechanism in my thesis, which is split into three phases. The first phase allows for bidder input into the network slices they bid for, overcoming a trade-off between market efficiency and tractability, making truthful valuation Bayes-Nash optimal. The second phase allocates resources to bidders based on a modified VCG mechanism that forms the multiple, non-identical resources of the market into packages that are based on bidder Quality of Service demands. Allocation is optimized to be socially efficient. The third phase re-allocates vacant resources of entitled network slices according to a Generalized Second-Price auction, while allowing for the return of resources to these entitled network slices without service interruption. As a whole, the mechanism is designed to optimize the allocation of resources as much as possible to those users that create the greatest value out of them, and successfully does so.

1. Introduction

The 5G spectrum promises to change the way mobile networks are used: not only does it provide the performance gains that are being currently touted for new mobile broadband connections, but it also promises a broad level of improvements that are not limited simply to mobile phones. Rather, the larger evolution of 5G is likelier to be built upon the flexibility of the 5G standard, not only in its performance, but in the underlying technologies and concepts that can be realised under the spectrum. With the possibility of mass customization of networks through technological changes that we will talk about in our second section, two completely isolated networks can function using the same hardware but would have vastly different network services tied to them and could be used by different clientele entirely. This is called network slicing and is key to the way in which 5G will be operated in the future. Such mass customization means that 5G spectrum holders could generate more value out of their spectrum licenses through higher or more valuable traffic, without cannibalizing their own mobile broadband businesses. However, how would these network slices be created in a market with vastly different needs and private valuations by prospective network slices? This is what I will seek to answer in this thesis.

Gartner expects that network slicing will only appear in commercial usage after 2022 (Keene, Fabre, & Takiishi, 2019), but there is no clear understanding on how this should happen: how should network slices be created and sold, how should the resources used to operate these networks be allocated? The current literature on the subject is still new, with the main focus in computer science literature on the operation of an individual network slice, sidestepping how they are formed. In my thesis, I focus on forming network slices in an as-a-Service manner, with the users of network slices not owning the underlying resources, but rather using them only as a service. This allows for a broader range of usage cases, since the shared economies of scale will be far cheaper than a built-for-purpose network slice. Also, designing the network slices as services allows for competition between bidders, forming a price discovery mechanism. Thus, I try to provide a practical model to the theoretical work in academia so that network slicing as-a-Service can be developed further.

The purpose of this thesis is to design an allocation and a re-allocation mechanism that, both combined, allocate the resources needed for network slices to the bidders in a socially optimal manner. I will seek to show that these network slices can be designed in an as-a-Service manner, with an allocation phase that is both envy-free and Bayes-Nash incentive compatible for all participants and is understandable enough to be practically applicable. I will also provide a re-allocation mechanism that allows for statically allocated resources to be dynamically re-allocated when they are unused, thus creating a more efficient and welfare-maximizing system. This is a unique addition to the current literature, which currently only focuses on static allocation mechanisms. In all, I'm seeking to show that an as-a-Service manner of designing network slicing markets would be both viable and socially efficient, thus the optimal way to operate the market.

The latest generation of mobile communication, the 5G standard, will possibly change the way mobile networks are utilized by businesses and consumers. The core aspect that makes 5G unique to other networking solutions or concepts is its flexibility to the demand that network users will have. The main design bodies have set up the following targets, compared to the previous generation (The Next Generation Mobile Networks Alliance, 2015):

- 1,000 X in mobile data volume per geographical area reaching a target of 0.75 Tb/s
- 1,000 X in number of connected devices reaching a density ≥ 1M terminals/km²
- 100 X in user data rate reaching a peak terminal data rate ≥1 Gb/s for cloud applications inside offices.
- 1/10 X in energy consumption compared to 2010 while traffic is increasing dramatically at the same time.
- 1/5 X in end-to-end latency reaching delays ≤ 5 ms.
- 1/5 X in network management Operational Expenditure (OPEX).
- 1/1,000 X in service deployment time reaching a complete deployment in ≤ 90 minutes.
- Guaranteed user data rate ≥ 50 Mb/s.
- Capable of IoT terminals ≥ 1 trillion.
- Service reliability ≥ 99.999% for specific mission critical services.
- Mobility support at speed ≥ 500 km/h for ground transportation.
- Accuracy of outdoor terminal location ≤ 1 m.

Although the actual realized specifications that can be achieved on the new standard are unclear, it is sure that 5G will improve mobile communication in varied, specific ways. However, not all use cases will need to achieve all these milestones: guaranteeing haptic touch, being able to feel things remotely as if located there, in remote surgeries requires critical reliability and short latency, but low data volume is transferred. Meanwhile, video streaming has vastly different demands with a need of high data transfers rates. Thus, there is a desire to be very efficient with what performance your network needs, causing a desire for high differentiation in network specifications. There are more examples in figure 1, where one can see how varied the demands can be within a single field or industry. Serving these different needs is possible through new technological solutions that enable the concept of network slicing. We will go through the technological aspects in section 2, but the value and challenge of network slicing can already be explained: building up tens of isolated individual networks within a larger network, each with a unique usage and structure, has tremendous value in enabling more varied technological solutions. This in turn would improve the social utility gain from 5G spectrum.

However, this sets up a challenge on how the network resources should be allocated and how the license holders for 5G spectrum should price these assets. Each use case would have its own demand and valuation for the resources that make the capabilities possible, but we have no way of knowing *ex-ante*, what exactly each resource is worth to them, and therefore pricing the service accordingly. Since the packaging of resources is often heavily reliant on the business case of the bidder, the true value of the resources is based on information that is likely unavailable for the market orchestrator.

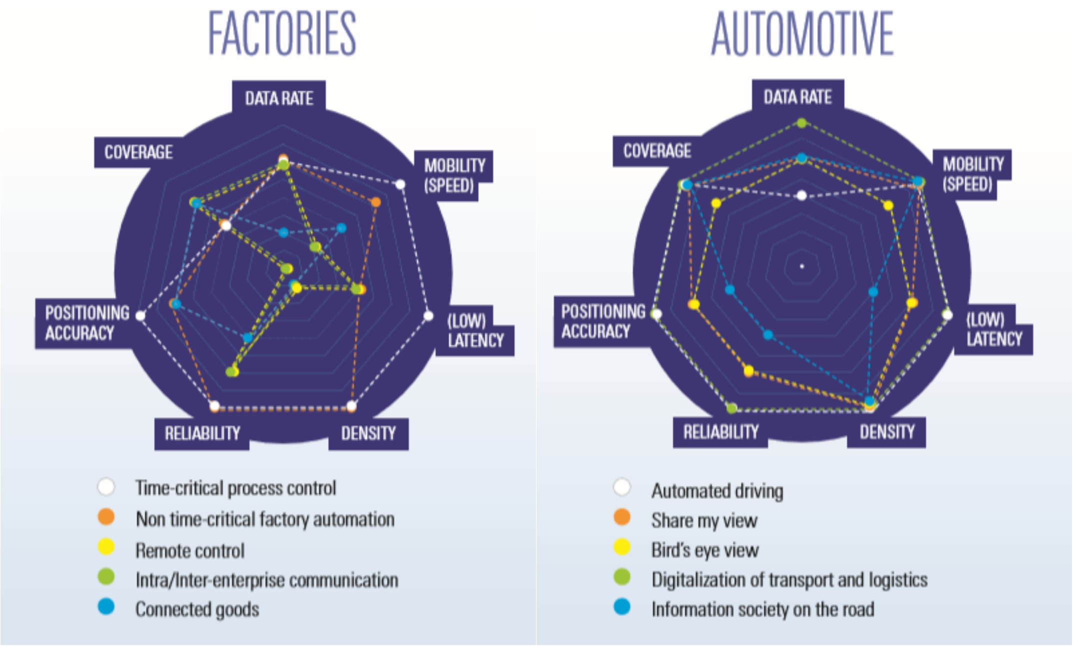


Figure 1: Requirement structures of different use cases in 2 industries. Source: 5G Empowering Vertical Industries, 5GPPP

It will be hard to define what the specific use cases for 5G will be, as well as being beyond the framework of economics. Rather, the aim should be on creating a marketplace that allows the 5G spectrums to be used by various novel use cases, where we don't know what their demands will be beforehand. The market mechanism would thus need to incentivize use cases to share their demands, as well as allow the market to form the ideal combination of network resources to create the most valuable network slices, before dividing up the resources and collecting payments efficiently.

In this paper, I will look at creating this network slice auction, built with a practical focus to be usable as a real-world market. This market would function through three stages: the first, a creation phase that collects the available resources and the bids that potential network slices offer for different capabilities. Second, an allocation phase that decides on who gets what and for what price. Third, a re-allocation phase that decides over time where excess resources, either newly added or unused by other network slices, are allocated. I will go through all these in section 4, with the comparative work on existing literature on the topic mainly handled in this section.

This thesis is structured as follows: section 2 covers the technology that enables network slicing, section 3 will make a short literature review on the current solutions in the pre-5G world for competition and innovation trade-offs that I'm trying to resolve in this thesis. Section 4 will go through the three phases of the market mechanism explained earlier, with further literature review on the specific solutions my model is trying to utilize. Finally, section 5 will cover a discussion section on the implementation challenges and opportunities my mechanism might face. I finish off with a brief conclusion in section 6 on what I've found: it is possible to create network slices as-a-Service, and thus make the trade-offs to competition that currently exist with network operators less relevant to the future. This is done through an allocation phase based on a VCG mechanism and a re-allocation phase based on a Generalized Second-Price (GSP) auction mechanism.

2. Technological components of Network Slicing

> *"5G is an end-to-end ecosystem to enable a fully mobile and connected society. It empowers value creation towards customers and partners, through existing and emerging use cases, delivered with consistent experience, and enabled by sustainable business models."*(The Next Generation Mobile Networks Alliance, 2015)

Before we can talk about the network slicing market, we need to understand the basics of the technology that enables it, and what exactly would be the product in our mechanism. 5G can be roughly defined as a collection of new technological advances that have been amalgamated into one standardized format with basic requirements, an effort by regulatory bodies, standardization organizations and industrial organizations (Morgado, Huq, Mumtaz, & Rodriguez, 2018). New software and hardware solutions allow for a wide array of functionalities to the 5G broadband, creating a need for a very flexible network architecture to satisfy all possible use cases. Rather than 5G being revolutionary in itself, it is the accompanying solutions that can change how mobile networks are utilised.

5G is expected to have three broad design and architectural principles, covering the basic purposes that the technology can offer to users. Some use cases can require ability in more than one of these areas (see figure 1), which are usually called extensive Mobile Broadband (eMBB), massive Machine Type Communication (mMTC) and Ultra-reliable Low Latency Communication (URLLC).

eMBB – This covers the standard purpose of mobile networks on the current generation, focused on providing high throughput of information between two devices or the network and a device, as well as high coverage and mobility support.

mMTC – This covers the desire for cheap and ubiquitous connection of devices to the network, such as different Internet of Things or Industry 4.0 solutions. The focus here is on high density of devices, with long battery life and high position accuracy or reliability in some cases.

URLLC – This covers the need for device connection that is both ultra-reliable (0.001% package drop rates) and has very low latency (<5ms connections). The focus here is on providing mission critical or time critical services the ability to connect as fast and securely as possible.

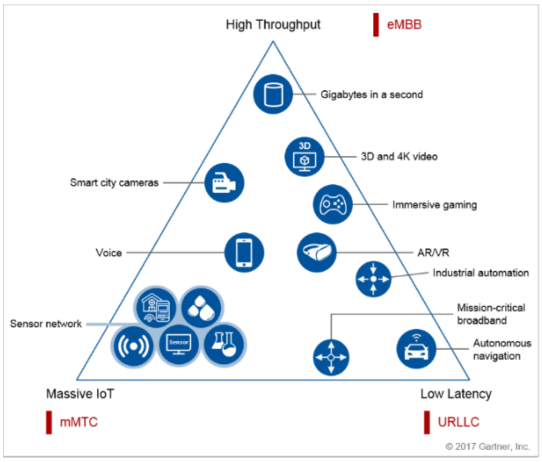


Figure 2: Examples of the three broad areas and their use cases. Source: Gartner

In the next five sections, I will briefly look through different technologies that will enable the changing use cases involved with 5G. We obviously cannot go through every single tangential technological change involved in the 5G architecture, so I will rather concentrate on the ones with the highest impact, before talking about what sort of solutions these would allow.

However, before we go into the technological aspects of network slicing, it would be valuable to preface it with the core economic changes that the technology will have, to prime you on the value of these concepts. The core driver of change is the virtualization of networks, since this changes how networks can be sold. This is made possible by the upcoming concepts. In the pre-5G world, networks were tied to physical resources, and thus the design of a network was pre-determined and fixed, since changes would require the addition of a physical mast or remote radiohead somewhere, for example. This would make the design of networks less responsive to demands, while also requiring those that might desire a network to build one themselves, since resources were more tied to a specific network topology. This also meant that the use cases were likely limited to mobile operators and companies that were large enough to make owning their own network viable.

Virtualization allows for mobile networks to be less tied to physical resources, and therefore more customizable in its design and users. Since the owner of a mobile network could share a part of its resources to another user without affecting the structure of the network itself, called multi-tenancy, it has more options on maximizing its own revenue. A multi-tenant network is the logical consequence of trying to get the most out of your network, and it follows seeking to provide the network as a service to lower the cost barrier for smaller but possibly more lucrative bidders (Zhou, Li, Chen, & Zhang, 2016). There is no change of mindset in the world, rather a change in technological possibilities. Moreover, customization through a virtual network allows for the exact definition of the network slice to no longer be pre-defined. This means that anything could actually be provided as a service under this market, since the combination of resources is not restricted. This further opens up the market to a broader ranger of usage cases (Taleb, Ksentini, & Jäntti, 2016). To summarize, the technological changes incentivize the market to open up from being mainly between competing network and spectrum owners to those same owners offering their resources to other users that would use their network without owning or operating it.

2.1. Software Defined Networks (SDN)

Akyildiz et al. (2015) give us a basic explanation what and how SDN differentiates from the current network technology. The main ideas are “(i) to separate the data plane (which carries user traffic) from the control plane (which plans the flow of traffic); and (ii) to introduce network control functionalities based on a virtualized network. These are achieved by (i) removing control decisions from the hardware, e.g. switches; (ii) enabling the hardware to be programmable through an open and standardized interface; (iii) using a network controller to define the behaviour and operation of the network forwarding infrastructure.” (Akyildiz, Lin, & Wang, 2015) This allows for a third party to control network resources, thus letting them control their own physical and virtual resources individually, giving them greater control of how their information is handled (Alexiadis & Shortall, 2017).

By separating the control and data planes of a network, SDN allows for these two to be operated independently from each other, rather than being tied to the same hardware. We can see this in figure 3: each forwarding device in a traditional network has embedded controls and may have middleboxes too. In a software-defined network, these controls are decoupled, with a single software control operating this for all forwarding devices. Thus, we are able to create programmable networks, simplifying network management and allowing high network programmability (Nunes, Mendonca, Nguyen, Obraczka, & Turletti, 2014). This also allows for operational structure for network slicing, which we will talk about later.

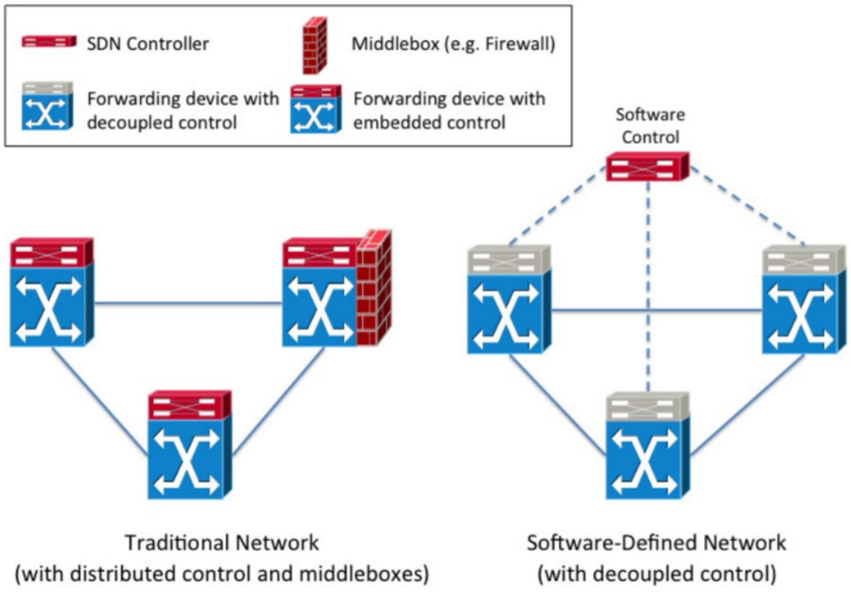

SDN Controller
Middlebox (e.g. Firewall)
Forwarding device with decoupled control
Forwarding device with embedded control
Software Control
Traditional Network
(with distributed control and middleboxes)
Software-Defined Network
(with decoupled control)

Figure 3: An explainer on SDN architecture logic. Source: Nunes et al. 2014

2.2. Network Functionality Virtualization (NFV)

Building on software-defined networks, Network Functionality Virtualization helps to abstract network functionalities and to implement these on software (Akyildiz et al., 2015). The NFV framework was originally developed and formalized by (ETSI GS NFV, 2013), creating a basic division of physical and vertical layers that has been used by later works. Mijumbi et al. (2015) gives us a comparative of NFV to the current way of network service provisioning:

a. *Decoupling software from hardware*: As the network element is no longer a composition of integrated hardware and software entities, the evolution of both are independent of each other. This allows separate development timelines and maintenance for software and hardware.
b. *Flexible network function deployment*: The detachment of software from hardware helps reassign and share the infrastructure resources, thus together, hardware and software, can perform different functions at various times. This helps network operators deploy new network services faster over the same physical platform. Therefore, components can be instantiated at any NFV-enabled device in the network and their connections can be set up in a flexible way
c. *Dynamic scaling:* The decoupling of the functionality of the network function into instantiable software components provides greater flexibility to scale the actual virtual network function performance in a more dynamic way and with finer granularity, for instance, according to the actual traffic for which the network operator needs to provision capacity. (Mijumbi et al., 2015)

So, rather than being reliant on specific and proprietary hardware to run, network functions can be migrated to the cloud, allowing for self-service in their operation (Morgado et al., 2018). The key effect of this technology on the utilisation of the 5G network would be the flexibility and control that a vertical, an operator in a specific industry or trade outside of the Mobile Network Operator’s own industry, can exercise on its own network usage. Rather than being tied to the Mobile Network Operator’s (MNO) proprietary hardware and software, it is able to import its own control mechanism and applications over to a new network. Moreover, NFV allows for general purpose servers, storage units and switches to be used on the physical level, which serve the new network based on the software that drives it (Han, Gopalakrishnan, Ji, & Lee, 2015). This liberalizes their ability to choose how they utilize a network, as well as move between different MNOs.

For a network as diverse in its applicability as 5G, the combination of SDN's control independence and NFV's ability to drive networks via software allows for a greater freedom to define the network in the way most useful for the individual use case. Combined, there is a greater ability to run network functions through cloud computing, rather than being reliant on physical hardware structures as networks currently are. Moreover, virtualization of the network allows for an easy and cost-efficient network isolation, since the isolation itself happens on a software level, rather than a hardware level.

2.3. Virtualized Multi-access Edge Computing (vMEC)

Although edge computing is not a new concept to 5G, it still plays a large part in explaining the flexibility possible on the platform. Mobile cloud computing as a general concept allows for a mobile device to offload its computation to the cloud, thus saving battery life, enabling more sophisticated applications to be run, and giving greater storage possibilities (Barbarossa, Sardellitti, & Di Lorenzo, 2014). However, utilizing the cloud can have high latency and high load on the radio access network due to the potentially distant location the cloud servers, both in terms of geographical and network topological distance. Edge computing seeks to bring this computation power closer, being distributed closer to the connected devices, thus straining the network less and providing a lower latency (Mach & Becvar, 2017).

In the context of creating network slices, edge computing resources are useful to many applications, and vital for a few, meaning that they will be a highly valued but scarce resource. Moreover, their geographic location will be highly strategic and value-defining, since they will only be useful to network slices if they bring latency down through proximity to the intended users. Thus, location requirements will be a critical part of slice design, and the pricing of these resources will have high fluctuation from one edge resource to another. We will thus look into the location of the resources within our market mechanism.

Edge resources allow for a network to improve its service in two broad areas: either by using computation resources to ease the load on a device, or from using storage capacity in strategic locations for content delivery. For the first, (Taleb et al., 2017) provides examples on how computation can be offloaded, where a mobile device transfers a computation-intensive task to an edge cloud to compute, such as the encoding of data in a video conference call. This also allows for faster decision-making by devices that are not well-equipped to deal with the computation efforts, such as small cell networks (Intel, 2013) and by mMTC devices to provide memory replications to the central cloud (Abdelwahab, Hamdaoui, Guizani, & Znati, 2016). By allowing for computation to be externalized from the user device itself, networks can have more creativity in terms of how it operates its services. Since computation isn't only tied to device that seeks it, more services can be provided to users than before.

The other example use case (Taleb et al., 2017) focuses on is the use of edge storage resources to create caches of content closer to the users that need it. The basic example of this is video streaming, where content is stored closer to the users, so that the network doesn't need to seek this data from a central cloud storage. Taleb et al. say this helps limit the "jittering" of a network through a faster and more secure latency. By not forcing devices to seek content from the central cloud, MEC helps reduce the load on the core network, reducing backhaul requirements for the network by 35% (Intel, 2013). Another example of content caching can be found in multiplayer gaming, with the stored data allowing for lower latency between the players. In these caching examples, location again is key, and thus will be a critical part of the re-allocation we'll talk about in section 4.4.

2.4. Network Slicing

Now, we'll look into how these three technologies combine to allow for network slicing to be done, as well as explaining the mechanics and the value of network slices. The core value of a network slice is allowing tenants to reap the benefits of better performance and lower cost from sharing resources, while retaining control of the data and the operations of your own network (Caballero, Banchs, Veciana, Costa-Pérez, & Azcorra, 2018). (Nakao et al., 2017) define a slice as "an isolated set of programmable resources to implement network functions and application services through software programs", which allows for separate sets of network functions to be run congruent to each other on the same network spectrum. To create a network slice, user demands are developed into Quality of Service (QoS)[1]

[1] Quality of service is the end-to-end performance of the network that is affected by both the network and non-network performance of the components that create the service. The definition of QoS is done on quantifiable and measurable terms, e.g. bit error rate, latency, repair time or performance provision time (ITU-T, 2008).

requirements based on software defined networking. SDN generates a logical network that optimizes how the virtualized radio resources are managed, drawing from edge nodes and centralized core network resources, depending on the needs of the network (Zhang et al., 2017). Thus, we can create a network that is tailored directly to the needs of its user, giving it access to only the resources it needs, while keeping their network isolated from the activities of other network users.

In figure 4, you can see the proposed NFV architectural framework that network slices would function on (ETSI GS NFV, 2013). Here, we have physical hardware resources that are chosen for the slice, which are then virtualized to create virtual machines (VMs) that can be used by any network slice within the greater network. The VMs are then used in different amounts to create a virtualized network function (VNF), which the slice itself uses to generate the services required from it, based on the stipulated QoS. Simplified, these VNFs are blocks of resources that serve to generate a less abstract service, such as acting as a switch or router, which are usually called network services (NSs). Each network slice is likely to need a different combination of VNFs to generate the one or more network services that the network should provide, and thus most slices are going to be utilizing a different combination of the shared hardware resources. We'll look further into these combinations in section 4.2.2.

Each network would also have a management and orchestration (MANO) plane to it, handling how the resources and VNFs are utilized. Part of the isolation concept of network slicing is that each slice has its MANO plane, orchestrating how the resource it is provided get used within its own network. This includes the virtual infrastructure manager (VIM), that manages how the resources are used, the VNF manager, that controls how the NSs are generated, and the NFV orchestrator, that orchestrates the virtualized section of the slice. The NFVO is then in contact one level upwards in the network slice, to the operations support system, which contains the slice controller. This section instantiates the slice itself, and creates it as part of the greater network it is in (Chatras, Tsang Kwong, & Bihannic, 2017). The slice orchestrators of different slices within the network would then be in contact with the network manager of the whole network itself, in order to seek the resources that its network services would need. Sometimes called multi-tenancy, this system of multiple network slices within the network would require the complex orchestration of all resources to maximize efficiency and the QoS available to all slices (Samdanis, Costa-Perez, & Sciancalepore, 2016).

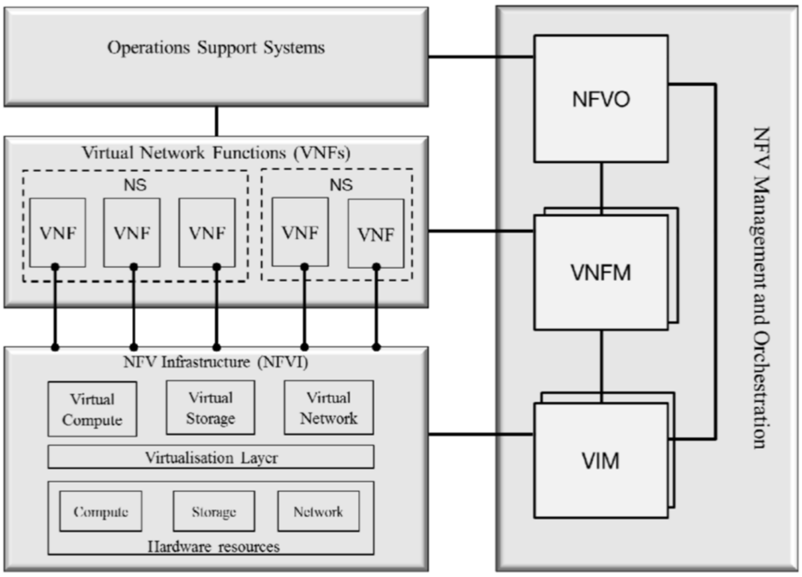


Figure 4: NFV architectural framework. Source: ETSI, 2013

2.5. Cloud Radio Access Networks (C-RAN)

The next thing we consider on the application of these technologies and solutions is the virtualization and moving to the cloud of the radio access networks (RAN), the data transfer network that connects devices to the core network. As (Ksentini & Nikaein, 2017) describe it, using SDN principles allows us to create a fully programmable network slice with a programmable RAN. This is done through the same principles talked about earlier: separating the control and data plane of the information going through the radio access network. What this means in practice is separating the physical remote radio heads (RRHs) that receive and transmit signals from the baseband units (BBUs) that process and direct the traffic. This would allow lower latency between RAN units, as the BBUs would now be pooled together, thus closer to the units they'd need to communicate with (Bhamare, Erbad, Jain, Zolanvari, & Samaka, 2018). Ksentini and Nikaein suggest using a Slice ID to direct the traffic through the pooled RAN resources to the right slice, from the user device to the core network, which would be encoded into the device, e.g. a SIM card (Zhou et al., 2016).

A critical thing for our mechanism is that it is yet unclear to what degree C-RAN will be implemented, i.e. how much of the transfer network will be virtualized or kept on their physical level, and which specific functionalities will be concerned. (Marsch et al., 2016) suggest that functions that are time-asynchronous (functions that don't need to occur at the same time) are likelier to be implemented virtually, while time-synchronous functions would still be physical. However, this is hard to decide beforehand, simply based on the needs of potential network slices being varied. This changes the input of the resources into our auction that potential network slices will compete for, especially as the market is continued from its initial allocation. Virtualized resources allow for dynamic scaling, allowing generic devices to be connected to the hardware pool, meaning that these resources can be added swiftly to respond to increased demand. Meanwhile, non-virtualized resources are more costly and slower to add into the network. Thus, the orchestrator of the market must be able to differentiate between the two, no matter how the radio access network is designed for the network slices.

3. Literature Review

I will look into existing literature on network slicing -specific topics throughout the following sections, but I will do a brief literature review into how competition was regulated and operated before network slicing, since it is a separate consideration from the rest of the thesis. Pre-5G and early 5G literature studying the role of innovation and competition of mobile network infrastructure and telecommunications focused heavily on the investment incentives and access pricing[2] imposed on incumbents. At its core, there was an implicit trade-off between the incentive for legacy incumbents to invest into the next generation technology and the ability of new entrants to compete in the market, creating a balance of social welfare between old and new technologies (Bourreau, Cambini, & Doğan, 2012). There was a big problem of enforcing competition on an investment-heavy industry due to *ex-ante* threat of regulation: the incumbent feared enforced access pricing that opened competition in only its profitable areas (Kotakorpi, 2006), and also causing unwillingness to invest into unprofitable areas (Bourreau, Cambini, & Hoernig, 2012) and a fear of investment spillover benefiting competing entrants (Foros, 2004). There was also a lot of focus on the best way to enable market entry (Brito, Pereira, & Vareda, 2010; Brito & Tselekounis, 2016; Lestage & Flacher, 2014) and the value of higher competition into the telecommunications market (Houngbonon & Jeanjean, 2016; Jeanjean & Houngbonon, 2017; Vogelsang, 2017).

[2] Access pricing is the concept of enforced access at a fixed price to the network of the incumbent for an entrant business, creating competition to the incumbent on their own network without new investment. This has been one of the standard competition mechanism pre-5G. To read more on this subject, see (Economides & White, 1995; Pindyck, 2005).

The concerns that come along with 5G have moved on from this core trade-off: the value addition for MNOs that happens through network slicing comes from attracting new users to their mobile network. Assuming that the incumbent MNOs don't have an interest in entering emergin vertical markets, narrow markets that meet specific needs for a targeted customer base, the MNOs aren't inhibited to make investments into their network. Since the verticals utilizing network slices are likely to have little to no effect on the mobile broadband business of MNOs, eMBB explained in section 2, there's little harm from these slices on their own profits. What Zhou et al. (2016) calls network slicing as-a-service (NSaaS) is going to be a key driver of how networks will used in the versatility of 5G. Rather than just creating network slices for their own business portfolio, MNOs would seek to provide different business verticals a network platform on which they can run their own businesses.

There is a big question regarding how the 5G solution will develop the mobile network business: whether into a "business-as-usual" model of mobile broadband -driven network slice utilization called the "evolutionary" image, or whether it'll be a more "revolutionary" image, with a broader demand for and competition of network slices (Cave, 2018). Under the evolutionary vision, network slicing will be used to mainly drive cost-effectiveness to offset rising network deployment cost[3] on internal markets, while the revolutionary vision would lead this being used into external markets with currently untapped demand (Lemstra, 2018). The core value of a NSaaS paradigm will be a shift of focus from Business-to-customer (B2C) to B2B and B2B2C utilizations of licensed network, allowing MNOs to expand their revenue stream and broaden the firms able to benefit from dedicated network services (Zhou et al., 2016).

There has been a lot of work put into the optimal allocation of virtualized resources in a network slice (Leivadeas, Falkner, Lambadaris, & Kesidis, 2017; Pianese, Gallo, Conte, & Perino, 2016; Tadayon & Aissa, 2018; Tang et al., 2018), where the core idea has been on finding the most efficient routing and combination of resources. This work goes beyond my thesis, since this intra-actor decision-making on which user gets which resources would be beyond the orchestrator of the auction, but it is good to know what the desired inputs would approximately be to estimate resource demand. Moreover, these papers do not answer the more fundamental question I am posing in this thesis: which bidders or network users should

[3] Since the transition from 4G (LTE) to 5G would mean that the spectrum used will move from the 2,1GHz band to the 3,6GHz band (the NR-78 band), there would need to be a densification of the cell tower network. The core trade-off of spectrum is higher throughput for a narrower coverage per tower, when increasing in frequency. For more on network densification, see (Bhushan et al., 2014; Oughton & Frias, 2018).

be receiving these resources in the first place? Intra-network efficiency is a later phase question that requires an understanding of how the network slices are formed. Thus, this paper attempts to resolve the trade-offs that regulators have faced in the past in foster competition in a natural monopoly like network spectrum through a mechanism that allows for competition, without cannibalizing the business of license holders.

4. Market Setup

4.1. Market Structure

In this section, we will look at how an NSaaS market mechanism would be operated, in order to gauge the plausibility and value of creating network slices. Our market would be executed in three stages: a creation phase that allows for bidders and sellers to understand the market and for the orchestrator to build up the market based on demands. Next, the allocation creates network slices on minimum instantiation levels of their desired networks, which are individually developed based on the bidder's demands. Finally, the re-allocation phase dynamically re-allocates resources that are not currently in use by other entitled network slices, once the network slices are fully in operation. In this market, there are inputs in the form of network, storage and computation resources, the sellers of these resources and the firms competing for resources with their own demands. Crucially, the market also needs an orchestrator, who forms prospective network slices to bidders and allocates resources to them by maximizing valuations, charging the prices agreed on from these buyers.

There is some consensus in papers on the topic of network slice or radio resource allocation that the orchestrator of the market should be neutral and independent towards both the sellers and buyers (Jiang, Condoluci, Mahmoodi, & Guijarro, 2017; Ordonez-Lucena et al., 2018; Tadayon & Aissa, 2018). With a neutral orchestrator, competitors on both sides can be sure that their offered resources or their valuation is not being twisted or skewed by the orchestrator, thus creating risk that would need to be compensated for. For an MNO to run the market for itself, this would both limit the ability for the bidder to compare between MNOs on other network resources without rival markets, as well as causing bidders uncertainty of their bid being treated neutrally. Critically, the problem in non-neutrality is private information that the bidders would not want to share with the MNO, fearing dishonesty in the mechanism and that their bid would signify their preferences or type (Myerson, 1983).

In what we seek to resolve in this paper, a market mechanism that allocates and prices the resources to bidding network slices, there is no definite answers in literature to the two questions that are outstanding on it: how do we define the inputs that the bidders want, and how do we allocate resources efficiently? On the former question, there's a challenge of creating accurate representations of the resources, while making it understandable to the bidder. Firstly, should the service description a bidder requests be human-readable or function- and network component-focused. Secondly, should the resources be coarsely- or finely-grained, in terms of the detail given (Foukas, Patounas, Elmokashfi, & Marina, 2017). A created allocation model suggested creating network chunks of each resources, which bidders can bid for separately (Jiang et al., 2017). Another solution provided is to create network services out of the virtual resources, thus having units of performance that are understandable enough for the bidder, while not reducing the efficiency of the resource combination too much (Ordonez-Lucena et al., 2018). The trade-off involved is thus the accuracy and efficiency of the network slices that can be created, and how understandable the market is for the bidders.

On the latter question, how we intend to allocate the resources, it is critical to understand how the resources are put up for sale: since bidders aren't seeking to gain resources, but rather to execute desired services, they will only be satisfied by a specified combination of resources. Thus, models where bidders bid for individual resources separately (Jiang et al., 2017) face the problem of conditional value: edge computing resources only have value if they're also guaranteed a RAN solution to transfer data across. An opposite solution, pre-packaging all resources into blocks to be bid on would remove conditionality as a problem, but would require pre-planning and knowledge of bidders' needs. Pre-packaging would require knowledge that the orchestrator cannot hold, since this is private information that the bidders themselves have. Designing packages based on lobbying suggestions would lead to mismatched packages skewed in favour of wealthier bidders, limiting efficient allocation. This would likely lock development out of the packages, or at least make this development driven by those within the network already. Thus, we have a trade-off between the resources being conditional to each other to form a bundle, and the efficiency of the resources being allocated. We will seek to solve this in the creation phase -section.

Next, we'll want to look into the resources on offer on the market, what would they be, how would they be differentiated from each other, as well as who would be providing them. The neutral orchestrator we envision would allow free entry of resources into the market, thus limiting the ability for any anti-competitive behaviour from incumbents. Before the pricing and allocation takes place, each resource provider would provide a set of resources and their quantities on offer to the orchestrator, as well as a minimum price with which they are willing to provide it. Pre-allocating the resources to the market is necessary to allow the resources to be connected to the network itself, making the creation phase of the network slices efficient. Moreover, each resource should have an exact geographical location provided with it, with the value of the resource being sometimes tied strongly to the where it's located. Finally, each resource must have an API tied to it, so that it can be easily integrated to the bidder (Taleb et al., 2016).

IoT service providers are likely to be pushing the market on edge computing, so creating the computation resources that are heavily location and proximity dependent (Hsieh, Chen, & Benslimane, 2018). Since this is likely to be a market driven by both price and connection security, the edge computing providers are likelier to be firms that provide such services outside telecommunications, rather than requiring industry-specific knowledge. On the same vein, core computing and storage functions, cloud functions that don't have high proximity demands, are most likely going to be provided by firms with high economies of scale. Moreover, this section of input is likely to be highly commodified, purely by the simplicity and ubiquity of the service provided. A second edge resource to be used are content distribution networks (CDNs), which provide storage of content on the edge of the network to provide lower latency and more reliability to the user, as well as making transfer of data on the backbone network less necessary (Cave, 2018). With 77 percent of network traffic being forecast to cross CDNs, these edge resources will be valuable (Cisco, 2017).

The telecommunications firms are in the greatest likelihood going to be providing two main input into the market, the spectrum the slice uses and the C-RAN and RAN resources the slice needs, along with controlling the creation of the network slices themselves for the bidder. Assuming that the upcoming 3,6GHz and 700MHz band auctions are going to follow the same pattern as the concluded rounds of auctions in e.g. Italy, Finland and the UK, most licensed spectrum is going to be picked up by telecommunications operators that have legacy network resources and spectrum from 4G and 3G operations. Thus, these legacy network owners are also likeliest to hold a cost advantage in bundling these services together, along with being unlikely to freely provide outside access to their own licensed spectrum. An alternative spectrum that can be used is unlicensed spectrum like TV white spaces[4] or licensed shared access (LSA) spectrum[5]. LSA spectrum requires the MNO to agree to sharing its spectrum with a third party, and they would likely prefer to do this under their own network slice, where they would retain greater control of the resource and would not compete against their other slices. While white spaces can be utilized in spectrums that MNOs cannot control or deny access to, their temporary nature means that QoS cannot be guaranteed on them, rendering them problematic for network slicing purposes. Thus, spectrum is likely to drive the usage of RAN resources, both from value of the routing power spectrum gives, and the bundling MNOs practice.

Next, let's delve into the creation of the network slice itself: what it entails and what value and control this gives the MNO. To do this, the service function chaining (SFC) architecture of the network slice must be created, which is the instantiation of an ordered set of functions that steers the traffic from end to end of the network (Halpern & Pignataro, 2015). This SFC architecture would be based on the demand description of the network slice, along with all the network services implemented within it. Under SDN principles, the logical and network topology path are separate within the network slice. The SFC will likely be pathed in a dynamic routing, which allows the pathing to happen without a predetermined topology, thus allowing for a more responsive network to VNF congestion, link failure and any changes to the network services provided (Choyi, Abdel-hamid, Shah, Ferdi, & Brusilovsky, 2016; Medhat et al., 2017). This sort of a dynamic planning would require a more complex knowledge of the entire network, not just the services that are being chained together, and thus is best served by the MNO itself. For the market to work as intended, with limited lock-in cost for the network slices on the MNO providing them the network connection, it is key for open access API to be developed that allow the movement of tenants between different MNOs with minimal cost (Lemstra, 2018).

A critical thing for the potential network slices is the need for all auctioned inputs to be understandable in both their quantity and quality; if they win, what exactly are they getting? Here, the so-called "Kelly's mechanism" can be very useful, which aims for fairness with the weighted proportional fairness criterion (Afèche, 2006, p. 110): allocation of

[4] The usage of TV frequencies that are not used in certain areas by unlicensed low-power devices, such as IoT devices. (Morgado et al., 2018)

[5] Exclusive access granted by a spectrum owner to secondary users on bandwidths that are underutilized, designed under a QoS agreement. (Marsden & Ihle, 2018)

resources should match to the valuation of each bidder (Kelly, Maulloo, & Tan, 1998). With different combinations of resources, this might not be always possible to gauge, although the principle stands in more limited and comparative circumstances. A fair allocation would also ensure an envy-free system: no actor in the market would rather have the package that another actor received for the amount they paid for. The market must also be simple enough that this envy-freeness is comprehensible to all buyers in the market.

4.2. Creation Phase

In the creation phase, the orchestrator seeks to bring in the bids from the potential slices, what network services they are seeking, and the offerings from the resource providers, what they are willing to provide for the market. The market is created during the creation phase, allowing both sides to indicate what they are seeking, without having to share this information with their competitors for the resources and the sellers with the resources. Thus, a loss in the market doesn't cause a loss of information to others, which would increase the hazard rate of the bidder, changing their behaviour in the market (Tadayon & Aissa, 2018). This highlights another reason for a neutral orchestrator: that the information of bids or offerings does not affect the behaviour of the other side of the market, thus not necessitating any *ex-ante* strategic behaviour.

The information on the network slice demand is asymmetrical in the market: the buyers have an understanding what they seek from the market, and the sellers can only anticipate this with what they seek to offer. Therefore, the decision-making on what sorts of network services are formulated from the available resources should come from the buyers, since they'll only share this information through a purchasing decision. However, they have a worse understanding on what exactly they need on a technical level, i.e. what resources in what network services are needed to create the network slice that they desire. The generation of a clear network service catalogue by the orchestrator for the bidders is therefore crucial: providing both an understanding of what each network service provides, and what they need to be combined with to function. Since these network services would likely combine resources from multiple sellers, the design of the would be pre-set before the bids are placed, but the creation of the network services themselves would happen after the resource allocation phase. This allows for the geographical and temporal customization of the network service for the slice, since the information of the desired combinations are not available prior to the bids.

In the next three subsections, I will look at how the creation phase will happen in effect, and how the different actors will behave and interact under my mechanism. To help understand the interactions between the actors and the order of action during the phase, see figure 5. First, sellers provide resources to create the marketplace of the mechanism, after which the orchestrator creates and allocates resources to network slices. Once the network slices are operational, the seller and bidder interact through the resources being made available to the bidder by the seller. Since the bidder behaviour in the market define the actions of the sellers, I will first go through the second phase mentioned, before going chronologically to look at the sellers' interactions with the orchestrator and bidders respectively.

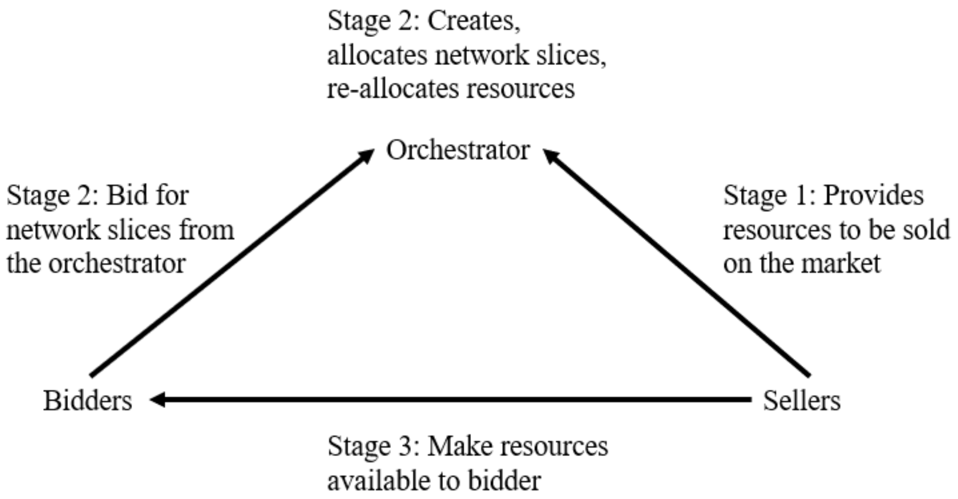

Stage 2: Creates, allocates network slices, re-allocates resources
Orchestrator
Stage 2: Bid for network slices from the orchestrator
Stage 1: Provides resources to be sold on the market
Bidders
Sellers
Stage 3: Make resources available to bidder

Figure 5: Interactions and action order between the market's three actors

4.2.1. Bidder Strategy

Our first focus will be on the stage 2 from figure 5, the interaction between the bidder and orchestrator. Specifically, we will look at how the bidder interacts with the network slices that the orchestrator creates for them. The bidder seeks the creation of a network slice, and we will follow the basic creation phase outlined by Ordonez-Lucena et al. (Ordonez-Lucena et al., 2018). The bidder has a set of performance demands for a network slice (e.g. wants their network slice to perform certain end user services at a certain quality), which it provides to the orchestrator. These demands are not formatted based on the resources it would need, only on the outcomes they would want to gain from the slice. The orchestrator uses these demands to form the ideal network structure, which would generate a resource-linked quality of service (QoS)[6] for bidder i, Q_i. For the purposes of this thesis, we will assume that the network slice generated by the orchestrator is the best-effort ideal solution for the bidder needs, and the bidder takes this quality of service as a metric for its own services. Thus, they will base their decisions on valuations from here on based on this quality of service.

To understand the creation of quality of service better, we can think of it as perfectly complementary isoquant lines of two network services, y_1 and y_2, which can be thought of as the switches and routers a network needs to function (see figure 6). In reality, a network slice will likely have more than two network services, but this allows us to conceptualize the QoS of a network slice with these isoquants. The problem illustrated would exist in a multi-dimensional space as well. Critically, adding more computational resources to a network slice that does not have enough bandwidth to operate has no value, and vice versa. Although the isoquant form might not always be perfectly complementary, as presented, this core demand for enough resources to be operative is key.

A network slice needs a minimal level of each resource or network service to be functional, i.e. to reach the lowest isoquant, or the minimal instantiation level (IL_0) in our case. The concept of IL_0 is key throughout my thesis: the orchestrator ties a bidder's Q_i to their minimal instantiation level, the required combination of resources to guarantee the bidder the service they demand. The bidder has valuations for each instantiation level, conditional on these resource demands, but the orchestrator is seeking to only meet the minimum demands in the allocation phase, thus only caring about the minimal instantiation level. If this minimal instantiation level of all resources or network services is not reached, the network slice is not operational, and thus has no value to the bidder. So, while the bidder mainly cares about the QoS service that is guaranteed to their network slice, Q_i, the orchestrator and the seller cares about the minimal instantiation level, IL_0, required to satisfy these demands.

[6] Quality of service is the end-to-end performance of the network that is affected by both the network and non-network performance of the components that create the service. The definition of QoS is done on quantifiable and measurable terms, e.g. bit error rate, latency, repair time or performance provision time (ITU-T, 2008).

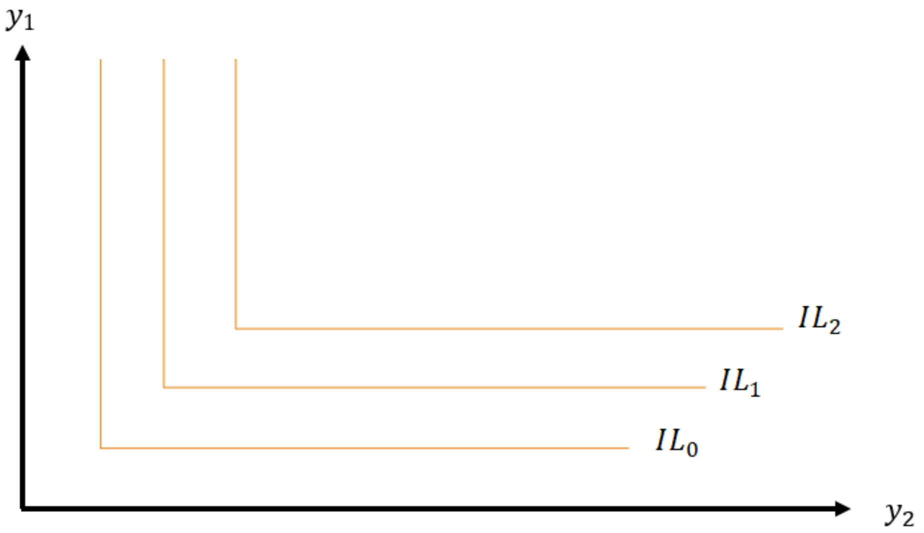


Figure 6: Isoquants of a simple network slice with two network services, y_1 and y_2, creating a level of QoS at different instantiation levels

The demand for the network services provided through our mechanism doesn't appear out of nowhere, but rather comes from either a growth in customer willingness-to-pay or through an expansion or efficiency gain in the B2B-offerings a company has. As mentioned in the introduction, network slicing causes a paradigm shift in spectrum usage from B2C to B2B and B2B2C paradigms, meaning the end users are not directly connected to a Mobile Network Operator (MNO) (Zhou et al., 2016).

The total utility, π_i, gained for each network slice, i, is driven by the total utility for existing customers (denoted by k), along with any new customers that would be drawn in by the technology (denoted by n). The utility is denoted by u and the cost by c. The cost of offering this technology is not additive, but rather a function of the total customers, K and a set of other factors including network effects and product synergy, θ.

$$\pi_i(Q_i) = \sum_{k=1}^{K} (u_{k,i}(Q_i) - c_{k,i}(K,\theta)) + \sum_{n=K+1}^{N} (u_{n,i}(Q_i) - c_{n,i}(K,\theta)) \quad (1)$$

This increase in total utility is driven by the change technology has on the service or product used, which customers value in a normally distributed manner. Moreover, customers will have a valuation for growth in service quality (e.g. throughput, latency, reliability) that is part of the network slice, in different ways, depending on the service. Each customer can have its own way of valuing the service, and we'll model the valuations as either step-wise, flat or constant functions. For the bidder, the aggregation of these demand curves generates their valuation for this service, what they stand to gain from bidding for the network slice.

The bidders report their valuations for their assigned Q_i to the orchestrator, denoted as v_i. As shown in equation (1), their total utility is formed from the perceived utility gain of both existing and new customers or businesses from gaining the network slice, compared to their current situation. This is simplified into the form of equation (2), with different utility levels with and without the guaranteed QoS of the network slice. Critically, the efficiency of the system should be based on the valuation that they hold for network slices, rather than just the QoS they are seeking. Even if bidders seek identical or near-identical network slices, their valuations for these can be vastly different.

$$v_i(Q_i) = \pi_i(Q_i) - \pi_i(0) \ (2)$$

As an example of how a bidder formulates its valuation based on its underlying demands as in equation (2), we will look at Dorsch et al., who analyse the economic benefit that comes from SDN communications on a network slice for smart grid communications (Dorsch, Kurtz, & Wietfeld, 2018). Smart grids are electricity grids that counter a more fluctuating energy production and consumption flows with distributive applications, connected with ICT solutions, in this case a network slice. The driving factor in using an SDN solution is its ability to guarantee hard service guarantees for the critical services in the distribution grid of the network and by enabling OPEX savings in the transmission grid, outperforming legacy networks within 4 years. Moreover, the competitivity of the technology comes necessarily from sharing resources with other end-users through having one of many network slices, rather than having a dedicated cell network.

An SDN-controlled network can provide a more secure and reliable operation of the smart grid than a legacy network, but the traffic itself is pre-planned and constant, and therefore the demand for high variation capability in the network is limited. Traffic is classified between *priority-random* (c_p)*, regular* (c_r)*,* and *scheduled* (c_s) (column 5 of table 1), with most of the services with a random or regular function being of small packages, needing a guarantee against package dropping. The scheduled traffic can be shifted in time flexibly, and thus the network is unlikely to need a spike in network services for any of the larger packages being sent. Moreover, the customers of the electricity companies are unlikely to have a larger valuation for a better smart grid infrastructure that wouldn't impact their personal quality of service, i.e. how they receive electricity to their homes.

The authors expect the aggregate data rate per connection to grow over time, creating a need for more resources reserved for their network services, and thus likely to generate a step-wise demand function for the network slice over years. Under our bidding mechanism, the smart grid operators would only bid for a Q_i that would cover the current network needs during the allocation phase. In the re-allocation phase, they would seek to meet their higher data traffic with additional resources, specifically the scheduled traffic, c_s, which is flexible in time but needs far greater resources (see column 2 in table 1). This paper thus gives an example of a model bidder for our mechanism design: they have a well-defined need for a network that varies in their capacity needs, which is improved by being operated as a 5G network slice. They form their valuation based on the better service guarantees and OPEX savings that the slice provides them.

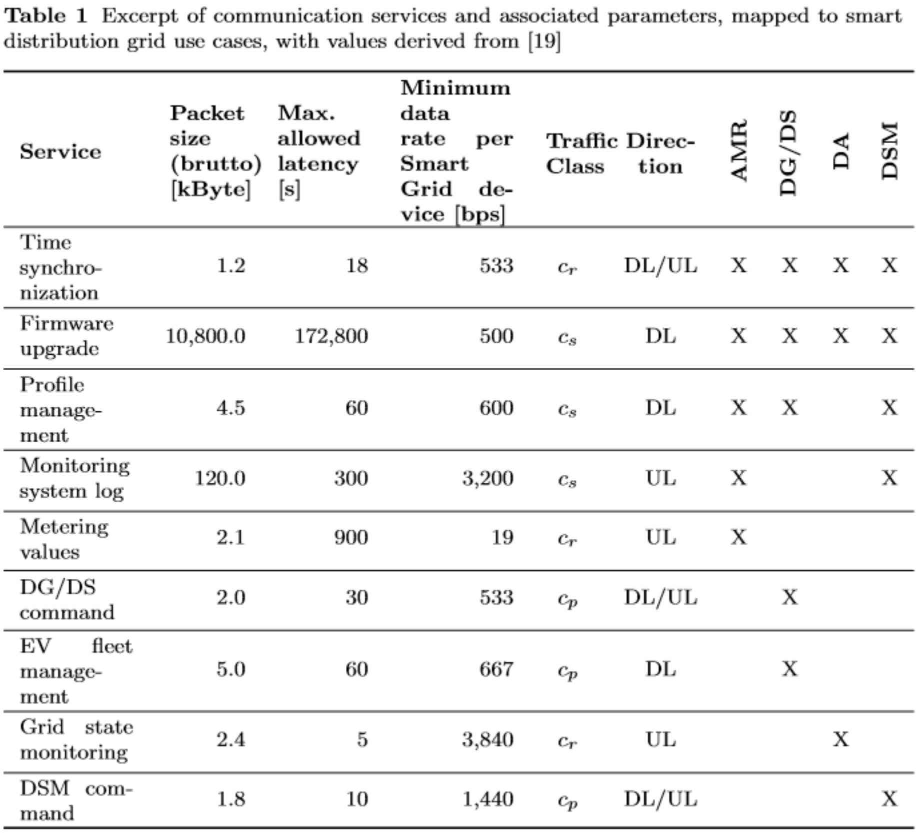


Table 1 Excerpt of communication services and associated parameters, mapped to smart distribution grid use cases, with values derived from [19]

Service	Packet size (brutto) [kByte]	Max. allowed latency [s]	Minimum data rate per Smart Grid device [bps]	Traffic Class	Direction	AMR	DG/DS	DA	DSM
Time synchronization	1.2	18	533	c_r	DL/UL	X	X	X	X
Firmware upgrade	10,800.0	172,800	500	c_s	DL	X	X	X	X
Profile management	4.5	60	600	c_s	DL	X	X		X
Monitoring system log	120.0	300	3,200	c_s	UL	X			X
Metering values	2.1	900	19	c_r	UL	X			
DG/DS command	2.0	30	533	c_p	DL/UL		X		
EV fleet management	5.0	60	667	c_p	DL		X		
Grid state monitoring	2.4	5	3,840	c_r	UL			X	
DSM command	1.8	10	1,440	c_p	DL/UL				X

Table 1: Smart grid communication needs. Source: Dorsch et al. (2018)

So, let's look at the strategy that the bidder should implement in our market, which seeks to maximize the profit that comes out of their network. To do this, I base my work on (Maskin & Riley, 1989). First, let's model the revenue so that the network slice only finds value from the resources allocated if they generate the minimal QoS demanded, i.e. reaches the lowest isoquant line given in figure 6. Critically, the seller cannot force a bidder to purchase a network slice, hence there needs to be a selling procedure in place. There, each action is a pair of QoS and price $\{Q_i, m_i\}$, and bidder i's strategy maps their type v_i to actions, as $s_i(v_i) = \{Q_i, m_i\}$.

We'll denote the strategy the bidder i chooses as s_i, with the other bidders adopting the strategy rules:

$$s_{-i}^*(\cdot) \equiv [s_i^*(\cdot), \dots s_{i-1}^*(\cdot), s_{i+1}^*(\cdot), \dots s_n^*(\cdot)] \text{ (3)}$$

As a non-cooperative game, bidder i's profit is affected by both its own optimal strategy, s_i^*, and the optimal strategy of all other bidders, s_{-i}^* . The strategy of each bidder is a series of pairs, with a network slice, Q_i, tied to a payment m_i. Since we want to simplify our calculations, we will formulate this pair in a deterministic form: we assume deterministic strategies, and thus rule out the randomization of actions. The strategy is a best-response to other players' strategies, thus we won't take subpar strategies into account when pondering the bidder's type. We can therefore note this pair as:

$$\left[\hat{Q}_i(s_i^*, s_{-i}^*), \widehat{m}_i(s_i^*, s_{-i}^*)\right] \text{ (4)}$$

We must also note a few key assumptions that must be made for the Maskin and Riley's mechanism to be incentive compatible. Firstly, we must assume a non-decreasing price elasticity: $\frac{\partial}{\partial v}\left(-\frac{q}{d}\frac{\partial m}{\partial Q}\right) \leq 0$. This implies that absolute risk aversion with respect to consumption is non-decreasing in their valuation. Also, we will choose a parameterization for which the increases in demand price are non-decreasing as v_i increases, $m_{22}(Q_i, v_i) \leq 0$. The combination of these two assumptions creates a reasonable expectation for a market like ours: that both the demand elasticity and valuations of the bidders are non-decreasing, as long as their individual rationality, $U_i(0,0) \geq 0$, is respected.

The valuations firms hold for a potential network slice is based on the combined end user valuation change for their services with the network slice (equation 1), since this is what drives their profits. Thus, in a market mechanism where company bids would be proportional to the network slice would have on its product's efficiency or increased value to its customers, the allocation of the resources would be to those companies that would generate the most utility out of them. To understand why bidding truthfully is optimal, Maskin and Riley consider the expected surplus of a bidder that has their valuation v_i, but bids something else, x_i. Thus, the bidder i of type v, with a total of N bidders, expects a utility, U_i, of:

$$U_i(x_i, v_i) = E_{v_{-i}}\left[N\left(\hat{Q}_i\left(s_i^*(x_i), s_{-i}^*(v_{-i})\right), v_i\right) - \hat{m}_i\left(s_i^*(x_i), s_{-i}^*(v_{-i})\right)\right] \quad (5)$$

where $N(\hat{Q}_i, v_i) \equiv \int_0^R p(z, \sigma_r, v_i)dz$, meaning that we assume the cumulative distribution function of valuations for each resource to be strictly increasing and continuously differentiable. Simplified, we assume that for each separate resource, σ_r, the probability, p, of being allocated the network slice strictly increases with the valuation, given all other resources valuations are constant. Next, we can suppress the strategy functions to define

$$Q_i(x_i, v_{-i}) = \hat{Q}_i\left(s_i^*(x_i), s_{-i}^*(v_{-i})\right) \quad (6)$$

$$m_i(x_i) = E_{v_{-i}}\hat{m}_i\left(s_i^*(x_i), s_{-i}^*(v_{-i})\right) \quad (7)$$

Then, we can rewrite (5) as

$$U_i(x_i, v_i) = E_{v_{-i}}[N(Q_i(x_i, v_i), v_{-i}) - m_i(x_i)] \quad (8)$$

Since the surplus is built on the expected actions of other actors within the markets ($E_{v_{-i}}$), the bidder knows that they cannot unilaterally change the market behaviour. They thus know that given other bidders are choosing strategy based on their demand, they must choose the optimal strategy too, or face getting a suboptimal amount. It then follows that, for all i and v_i,

$$U_i(v_i, v_i) = \max_{x_i} U_i(x_i, v_i)\,. \quad (9)$$

To generalize what we've learned from Maskin and Riley, a mechanism only induces true valuations, v_i, from a bidder if this is their best response, given that other bidders are playing according to their truthful strategies. Further, for the mechanism to induce truthfulness from all bidders, providing a truthful valuation must be the Nash equilibrium for all bidders, which requires that all bidders that gain something out of the mechanism must hold a nonnegative expected utility (Myerson, 1981). This nonnegative utility is driven by the information that the bidder holds, and which the orchestrator needs to create an efficient allocation. This comes in the form of an information rent, the value of the private information each bidder holds, which Myerson shows to be the integral between a non-truthful and truthful valuation (Myerson, 1981, p. 17). This means that each bidder with an accurate valuation is rewarded by the mechanism, giving them cause to be truthful even when the orchestrator has no knowledge of their valuation type. I have thus shown that our market would be a Bayesian-Nash incentive compatible mechanism.

Since we know that the bids placed by firms for network slice resources is driven by the efficiency or additional customer willingness-to-pay, and that the bid itself maximizes firm profit by matching this increased demand with the amount of resources bid for, we can now know that the bids themselves will be efficient in the market. Due to the truthful bidding of all bidders, we can create a socially efficient allocation mechanism. In Maskin and Riley (1989), this is driven by a single price in the market, since the multi-unit product being divided is competed on by all. In our case, however, a solitary price is not possible to enforce effective allocation, given that the unique combination of resources required means that the price charged of each bidder can be unique as well. We will look into this at section 4.3.1.

4.2.2. Creation of the Network Slice

Now, we will look at stage 1 from the figure 5, how the sellers provide resources to the orchestrator. Given that the market consists of multiple different multi-unit resources, the sales process is more complicated than in Maskin and Riley. Therefore, we need to figure out a model for how the cost of each network slice is structured with respect to the object of the mechanism being combinations of resources, so that we can quantify what is being bought and sold. The easiest way we can characterize the costs on the input-providing firms is to model all these inputs as a group of vectors, with pairs of input and output vectors for each product on offer. A combination of input vectors, resources in our case, creates an output vector, or a network service in our case, that can be combined with other output vectors to form a full network slice, Q. Critically for the orchestrator, it seeks to minimize the inputs it has to put into generating this Q. Thus, the multiproduct minimum cost function for the network slice is:

$$C(y, w) = \min_{\sigma_r}\{w \cdot \sigma_r : (\sigma_r, y) \in Q\} = w \cdot \sigma_r{}^*(y, w), \; (10)$$

where $\sigma_r{}^*(y, w)$ gives the minimum input vector, σ_r, to achieve the desired output vector y, given the factor prices, w (Panzar, 1989). Panzar creates a regularity condition for this function, as the technology set Q, which Panzar defines as T, is a “nonempty closed subset of R × Y”, where R and Y are the compact sets σ_r and y respectively are drawn from. It includes two additional properties: i) $(0, y) \in Q$ if $y = 0$ and ii) if $(\sigma_r, y) \in Q$, $\left(\sigma_{r_1}, y_1\right) \in R{\times}Y, \sigma_{r_1} \geq \sigma_r$ and $y_1 \leq y$, then $(\sigma_{r_1}, y_1) \in Q$. Property i) shows that positive outputs require positive inputs, and ii) proves that the production process is at least weakly monotonic, i.e. an increase in input must at least weakly increase the output (Panzar, 1989).

This framework is useful for our market: each bidder seeks a specific sort of combination of output vectors into a multiproduct package, its network slice. It wants to achieve its desired minimum requirements, and the seller seeks to meet these requirements with the minimal network, storage and computation resources to achieve the Quality of Service, with the orchestrator creating these combinations as a proxy.

Meanwhile, the seller is generating a collection of output vectors, or network resources in our case, that it seeks to sell on the market for network slices. To do this, they have their own cost function in providing these resources to the market. Let's say that the firm is producing 2 of the 3 output vectors on this market, y_1 and y_3:

$$C(y_1, y_2, y_3) = F^{13} + c_1 y_1 + 0 + c_3 y_3 \ (11)$$

The firm has both fixed costs and production costs, which we'll now assume to be linear for the sake of simplicity, although this would likely be a more complex cost structure for a real-world firm. The core thing here for our mechanism is the differentiation between fixed costs and running costs: although these would drive the prices the sellers offer on the market, this will need to be designed into a total cost at different quantity levels.

Combining with the isoquants we saw in figure 6, we can see the basic structure of the optimization problem the orchestrator faces in figure 7. Here, we see the isocost function given in equation (10) seeking to minimize the usage of resources for the seller to produce the requires QoS level that the bidder requests, the isoquant. This optimization, with more numerous inputs and simultaneous requests, is what the orchestrator is trying resolve.

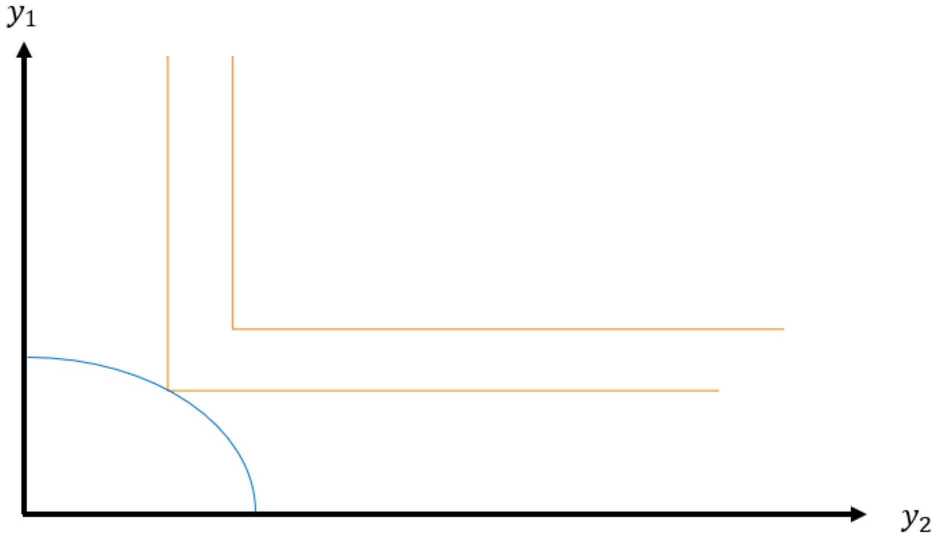


Figure 7: Orchestrators isocost function seeking to minimize resources while meeting the minimum isoquant of the bidder

With this information, the orchestrator must then generate the network slice for the winning bids, according to the service requirements that the bid is tied to. The most practical way for this is the slice creation being operated by the MNO whose spectrum and C-RAN resources were won by the bidder, along with the 3rd party resources that form the other network services. This allows for the topology of the network slice to be integrated with the other network slices operating under the network resources of that MNO, thus guaranteeing the isolation of the slices. Thus, both the buyers and sellers provide their valuations for what they seek to buy or wish to sell through the same medium, bundles of output vectors in the form of network services combined into a network slice. Sellers list exactly what resources they are willing to put up for sale, and this forms the market that we need to allocate to the bidders that have their private valuations for the services these resources can provide.

4.2.3. Seller Strategy

Finally, we will look at stage 3 of figure 5, looking at how sellers make resources available to bidders, and their limitations. The question of how exactly sellers will be compensated by the mechanism is outside the core questions of allocation I want to explore in this thesis, and thus lies mainly outside its scope. We can assume that the sellers have some compensation scheme agreed upon with the orchestrator by the start of the network slice creation, thus this interaction has no effect on the orchestrator's decisions. Still, it is important for us to define the role that the seller can and cannot play within the mechanism for the market to stay functional.

For the resource seller, the market mechanism allows for a more efficient allocation of excess resources from their own businesses. The opportunity cost for the resources is usage in their own business, such as the MNO using their network resources, such as their spectrum and C-RAN capacity, on their own eMBB business. A faster connection and higher throughput have value to their customers, but there is a strong diminishing return for these resources (Oughton & Frias, 2018). With SDN- and NFV-enabled resource-sharing, resource owners can allocate their resources to users that have a higher utility, and thus valuation for the resource. The seller will seek a minimum price for their offered resources that at least is higher than the utility it provides in their own business. The seller will demand a price for its marginal resource that is higher or equal to the marginal utility of the resource for its own businesses, linked either to the cost efficiency gain or willingness-to-pay of their customer. Crucially, this minimum price must be fixed over time, rather than fluctuating based on the current best price for the firm. Such a market would make the resource allocation unstable on the lowest instantiation level, such as with the minimum price demand above the agreed IL_0 cost of a bidder. As mentioned, this thesis will not focus on the transactions between the orchestrator and sellers, but there's an inherent question of who holds risk in the market between these two that could be subject for future research.

If the seller could withdraw its committed resources from the market, the IL_0 of all bidders cannot be guaranteed, meaning that the valuations of the bidders would have to take this risk into account, potentially making the market mechanism untenable. Since there is no stability with the allocation of resources, the incentive rationality constraint of the bidder is broken, if they value a null allocation to the sub-optimal allocation, x.

$$U_i(v_i, v_i) > U_i(0,0) > U_i(x, v_i) \ (12)$$

A decrease in resource availability pushes the price of the marginal network service up, given no change to the demand for these network services. Once a suboptimal allocation breaks the incentive compatibility of the marginal network slice through the budget limit being hit, their valuation of all other resources drops to zero as well. A network slice without the ability to perform the services sought from it has no value to its owner and would cause a withdrawal from all the other resources. Thus, dynamic minimum price alteration could make the entire mechanism unstable for other sellers and the value generation to the bidders' customers uncertain, making the entire market mechanism untenable. A settled minimum price that can only be altered at set times of renewed allocation phases removes the uncertainty of dynamic minimum price setting, without locking sellers into the mechanism at a fixed price that would change *ex-ante* behaviour.

4.3. Allocation Phase

4.3.1. Mechanism Design

In this section, we will go through the allocation phase, where the orchestrator takes the network slice demands that the bidders have and creates prospective network slices for them to bid on, finally creating the network slices that maximize the network utility with the resources made available by the sellers. We will briefly go through how the design part of the market mechanism would work, before we look at what the *ex-post* optimal allocation phase would look like. This allows us to look at what we are optimizing the network towards, and how we can achieve something close to it in an *ex-ante* setting.

The problem we are designing for is specifically a multi-unit market with nonidentical resources being sold. Each bidder is seeking a combination of nonidentical resources from the orchestrator to create a network slice, based on private information. Thus, an Ausubel multi-unit auction is incompatible with our problem, specifically due to the necessity of nonidentical resources being combined (L. Ausubel, 2004). Rather, my starting point is going to be the design of nonidentical object multi-unit mechanisms offered by (Krishna, 2010b). I will reword Krishna's example to fit our parameters. The model defines a finite set of resources on sale, $K = \{a, b, c, \dots\}$, with N buyers. Each bidder $i \in N$ assigns some valuation $v_i(S_i)$ for each subset $S \subseteq K$, which are packages of resources, a network slice in our case. We will suppose that $v_i(\emptyset) = 0$ and if $S \subset T$, S is a subset of T, then $v_i(S_i) \leq v_i(T_i)$. This gives us that participation in itself has no value to the bidder and that all resource valuations are nonnegative. The orchestrator then creates an allocation of all the resources on offer $(S^1, S^2, \dots S^P)$ into P packages, with the requirements that (a) every resource must be allocated to a bidder, and (b) no resource is allocated to more than one bidder. Note that S_i is similar to Q_i, in that package S_i gives the bidder a level of QoS just like Q_i does. However, I will show why pre-packaging like S_i is a suboptimal way of allocation in the next section. With all bidders reporting valuations for the packages, the orchestrator allocates the packages. An allocation rule is efficient if, for every bidder valuation $v \in V$, the allocation $\mathbb{S}(v)$ maximizes social welfare over all allocations, so

$$\mathbb{S}(v) \in \arg \max_{\langle S_1, S_2, \dots S_N \rangle} \sum_{i \in N} v_i(S_i) \quad (13)$$

The payment a bidder pays for a network slice is based on the externality on the other $N - 1$ bidders by their presence in the bidding mechanism. Effectively, their payment is the difference in social welfare when they bid v_i, rather than 0. I will look further into this mechanism through this section of the thesis.

In our specific case of mechanism design, Krishna's proposed VCG mechanism for efficient allocation and network slice price, as well as the vector formulation of the individual packages, is the right starting point that I will work on from. However, I believe that the menu-based design of these subsets or packages that Krishna proposes is incompatible with our specific market design, and thus must be worked around. I will first go through the incompatibility, before giving an alternative mechanism to package creation that is built on existing computer science literature. I will show that this alternative maintains the truthful direct mechanism that the VCG mechanism would require, while solving the issue of tractability inherent when creating network slices. Thus, my mechanism would be in line with the current literature on nonidentical object mechanism design.

4.3.1.1. Issue with Menu-Based Allocation

Myerson lists two levels of uncertainty regarding the valuation of a bidder by the orchestrator and other bidders: preference and quality uncertainty. These would be the uncertainty of knowing the bidder's personal preferences to the object offered and the uncertainty of the bidder having special information on the intrinsic quality of the object respectively (Myerson, 1981). In our case, preference uncertainty relates to the orchestrator not knowing what kind of a network slice the bidder values, while quality uncertainty relates to the underlying value of the network slice itself. The orchestrator wants to coax this information out of the bidder, while the bidder wants this information to stay private from the other bidders and will only give the orchestrator this information if they gain from this. This second point relates to information rent: the bidders that win in the mechanism want to be rewarded for their information by surplus from the market. Thus, they are only willing to share private information, shed light on the two uncertainties, if they gain from it.

This is the opposite that any revenue-maximizing orchestrator would want to do: the orchestrator would want to create network slices to all bidders at prices that would leave them no surplus. However, this would require information that the orchestrator knows exactly what the bidders value and by how much, and a menu structure would not coax this information. By a menu structure, I refer to the orchestrator giving the bidders a list of options of network slice structure, a combination of distinct resources in a network topology, at a fixed price. The bidder would then browse through the menu and pick the option that maximizes their utility, without ability to negotiate on the content or price of the packet.

The next problem the orchestrator faces is specifically regarding the infinite combinatorial options that it has for the network slices. It has limited information at the start of the market mechanism on what combinations the bidders would want, and what combinations create value in a network slice. Thus, the preference uncertainty of the orchestrator might mean that the created network slice options would fit the bidders poorly, causing the created options to be less valuable to the bidders, and thus to the orchestrator. If the orchestrator were to ask the bidders for this information, the bidders with the most accurate private information might fear losing this information to other bidders, and thus give information with little value to the orchestrator.

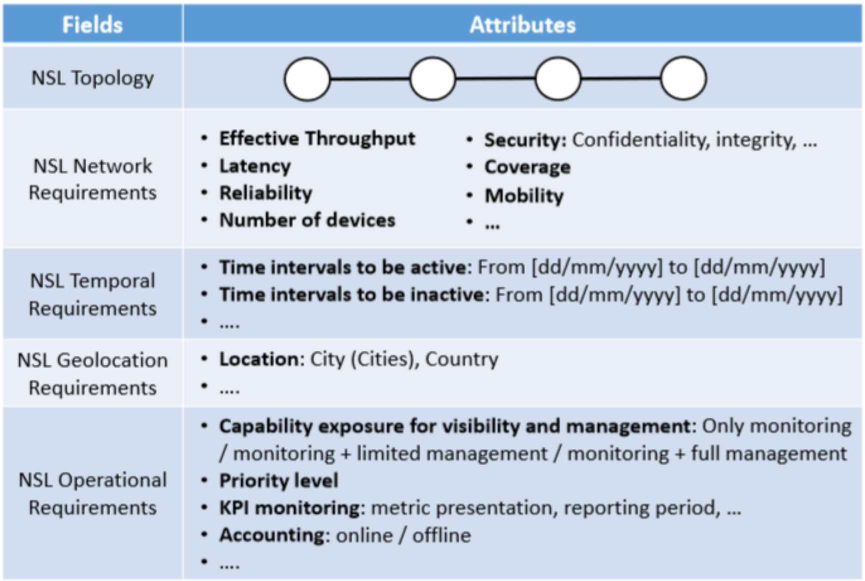
In the end, a menu design would face a trade-off between efficiency and tractability of the options. Given that the orchestrator could create near infinite options of network slices, with marginal changes in resources from one to the next, each bidder could theoretically find the optimal network slice from this list. However, the bidder would firstly have a difficulty in finding the optimal network slice from this list. Not only the task of finding the right option, but also since the bidder can only poorly translate their network demands into a bundle of resources. Thus, the bidder will make sub-optimal decisions without the orchestrator's help. I will look into how the orchestrator supports this in the next section. But more importantly, the orchestrator would still be uncertain on what the price of this optimal network slice would be. Even if the bidder were able to find the right network slice option, the orchestrator would need to coax a valuation from the bidder, without a mechanism to do so. The orchestrator might thus end up giving the right network slice at a much lower price than they could. The core problem here is that the bidder's price is dependent on their valuation, and thus they have an incentive to be strategic with their bid.

Moreover, the combination of these two uncertainties is important: if bidder i finds out that bidder j's valuation, v_j, is very low, they will assume that bidder j has some private information regarding the quality of the product (Myerson, 1981). If bidder j knows its private information to hold value, i.e. that it knows something about the intrinsic quality of the network slice others don't, it would seek to protect this from other bidders. Thus, even if we were to give the orchestrator full knowledge of the bidder valuations to create an optimal menu of network slices, the bidder might still wish to avoid signalling this private information to other bidders through its decision (Myerson, 1983). The bidder might seek to shelter its knowledge and make a sub-optimal choice from the menu to retain the quality uncertainty others have.

From these points, we can note that the orchestrator is in a weak position to utilize a menu-based market mechanism to distribute the resources to bidders. The allocation must be based on the private information that the bidders have, thus requiring the orchestrator to provide information rent to those bidders with critical private information. Given this information, the orchestrator can optimize the allocation to be efficient based on bidder valuations and learn the true value of different combinations to diminish the quality uncertainty that exists *ex-ante* in the system.

4.3.1.2. Ex-post Optimal Allocation

As mentioned earlier, the solution thus must have both the individual capability to influence the package allocated for you as the bidder but would not require conditional bids for multiple combinations. This can be achieved using a starting point of service ordering by (Ordonez-Lucena et al., 2018), where each bidder provides the orchestrator with what they'd be seeking from a network slice (see figure 8). Using this template, the bidder can show the orchestrator exactly what they value, allowing the orchestrator to create a prospective combination of network services to create the ideal network slice for the bidder. This would then be the item they would be bidding for, along with any alternative combinations created by the orchestrator as next-best options for the bidder. Examples of this would be creating packages with different edge computing resources in urban areas the orchestrator believes will be highly contested. Such a pre-planning allows for each bidder to bid for something their own optimal product, without valuations affecting what they end up paying and not requiring any conditionality in bids. Also, alternatives can be used to minimize the possibility of inefficient allocation caused by one highly desired resource skewing the allocation results.



Fields	Attributes
NSL Topology	○—○—○—○
NSL Network Requirements	• **Effective Throughput** • **Latency** • **Reliability** • **Number of devices** • **Security:** Confidentiality, integrity, … • **Coverage** • **Mobility** • …
NSL Temporal Requirements	• **Time intervals to be active:** From [dd/mm/yyyy] to [dd/mm/yyyy] • **Time intervals to be inactive:** From [dd/mm/yyyy] to [dd/mm/yyyy] • ….
NSL Geolocation Requirements	• **Location:** City (Cities), Country • ….
NSL Operational Requirements	• **Capability exposure for visibility and management:** Only monitoring / monitoring + limited management / monitoring + full management • **Priority level** • **KPI monitoring:** metric presentation, reporting period, … • **Accounting:** online / offline • ….

Figure 8: A service template structure. The bidder provides as exact definitions of all requirements and their importance to the orchestrator. Source: Ordonez-Lucena et al.

Once all bidders have their network slices and their alternatives designed for them, the next stage would be the valuations for these network slice templates. Each bidder is free to value each alternative they have of the templates as they wish, with the safe knowledge that their valuation is private and that their valuation is separate to the end price they'd pay for a package. As shown in section 4.2.1 and equations (3-9), the mechanism is Bayesian-Nash incentive-compatible, and thus the valuation will be truthful, given truthful bids by all other bidders. The orchestrator collects all the bids of the individual network slice templates offered and starts forming combinations of network slices. The orchestrator seeks to create the most valuable combination possible from all the bids, i.e. an allocation of the available resources in such a way that it maximizes the total valuation of the network slices within the network. This way the mechanism not only guarantees that the bidders winning a network slice have a set of resources that is personalized to their needs, but that the end effect is an allocation with each winning bid getting the optimal package for them. Moreover, the orchestrator can maximize the value of the resources to the sellers by allocating to the bidders with the highest valuations.

To generate the QoS we talked about in section 4.2, the orchestrator combines a basket of inputs from the resources available to it from the sellers on the market. Since network slices would value identical resources located in different places differently, we must separate each resource instance into its own input. We will simplify the multiproduct cost function presented in section 4.2.2 by focusing only on the inputs or individual resources, σ_r. The orchestrator seeks to minimize the outputs that are created by the inputs (equation 14), the main restriction and thus concern is the limited inputs or resources it has.

$$\sigma_r\colon Q_i \leq F\left(\sum_{r=1}^{n} \sigma_{i,r}\right)(14)$$

$$\sum_{r=1}^{n} \sigma_{i,r} = I \cdot \begin{bmatrix} \sigma_{i,1} \\ \cdots \\ \sigma_{i,n} \end{bmatrix}(14a)$$

So, the orchestrator combines resources together to create the desired QoS, meaning the network slices can be abstracted into a function of a basket of resources, $F(\sum_{r=1}^{n} i, \sigma_r)$. The ratio of each resource that the bidder gets can therefore be defined as $\frac{\sigma_{i,n}}{r_n} = [0,1]$. Critically, the bidder does not know or care what the inputs within its network slice are, it only cares about the QoS it can gain, as well as its cost. The orchestrator collects all the QoS valuations of the prospective network slice owners, the bidders, to set up the combinations it must create the network from. Thus, the vector Q_i has different meanings for the bidder and orchestrator: the bidder views it as a guarantee of service from the network slice at a specific quality, while the orchestrator views it both as a service to the bidder and a bundle of specified resources that the bidder would gain if they win. Note that the definition of a bidder's Q_i does not mean that they automatically get this. Rather, this is what they are bidding for in the allocation phase, and only gain the network slice and its guaranteed resources, if they win in the auction.

With all the valuations collected, the orchestrator has an optimization problem (equations 15-15d) in its hands: it seeks to maximize the utility gained from the network, while keeping within the resource limits, with R denoting the total resources. $\sigma_{r,i}$ meanwhile denotes the amount of resource, r, that is assigned to bidder i's network slice. Thus, the limits for the orchestrator are that the combined allocated resources do not go above the total resources on offer, and that all the individual resources, σ_r, should be offered at least in the amount that is in accordance to the bidder's Q_i.

$$\max_{\sigma_r} \sum_{i=1}^{I} v_i(Q_i)\ (15)$$

$$s.t. \sum_{i=1}^{I} Q_i \leq I \cdot R\ (15a)$$

$$s.t\ F(\sum_{r=1}^{R} \sigma_{r,i}) \geq Q_i\ (15b)$$

$$R = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix} (15c)$$

$$r_n = \sum_{i=1}^{I} \sigma_{n,i}\ \ (15d)$$

The orchestrator resolves the optimization problem by maximizing the valuation that is created on the market, stacking up network slices using the total resources on offer. Bidders pay for their effect on the rest of the market according to the Clarke pivot rule: their payment is the difference to the total market valuation from them being and not being included, i.e. the effect of bidder i on all other bidders (Clarke, 1971).

$$m_i = \max \sum_{i=1}^{I} v_i(Q_i) - \max \sum_{i=1}^{I_{-i}} v_{i-1}(Q_{i-1}) \quad (16)$$

Essentially, each bidder that gains a network slice is taking a network from a rival bidder $i - 1$, and thus by maximizing the total valuation between all bidders and all bidders save for bidder i, we are able to value their load on the network. Their payment is precisely the total loss of utility that other bidders see from their participation. The pivot rule uniquely allows both a feasibly truthful VCG mechanism and to create prices for bidders that are seeking vastly different packages from our network. Their price comes from the effect they have on the market, rather than from any valuation for a specific resource on offer.

For this allocation to be a truthful direct mechanism, we need to prove two propositions: (1) that bidder utility is best served by bidding truthfully and (2) the bidder values the package created for them the highest. Proving these, we show that the bidder is best served being truthful, and to bid for their assigned network slice.

Proposition 1: Bidder utility is maximized by bidding truthfully, $U_i(Q_i, m, v_i) \geq U_i(Q_i, m, x_i)$

I already showed in equations (3-9) how a bidder has no incentive to bid anything other than truthfully, given that this is an optimal strategy for all other bidders as well. We can note that the bidder's utility is improved by the probability of getting the network slice rising (see equation 5), with $c_i(Q_i, v_i)$ giving us the conditional probability of winning network slice Q_i with a valuation of v_i. Here, it is tied to probability p_i, which captures the effect of the expected valuation of other bidders on the valuation of bidder i of equation (8):

$$c_i(Q_i, v_i) = \int_{v_{-i}} p_i(v_{-i}, v_i(Q_i)) f_{-i}(v_{-i}) dv_{-i} \quad (17)$$

This means that the conditional probability of a bidder winning the network slice with their true valuation, v_i, is based on both this valuation and the function of all other bidders' valuations. Thus, the incentive-compatibility conditions of bidder i that must be satisfied, which (Myerson, 1981) set, are the same as how they are set up in equation (8). A difference in notation is that Myerson displays utility in integral form, while Maskin & Riley don't.

In sum, the bidder would have to be better off with a truthful valuation than anything else, given that other bidders choose actions according to their truthful valuations. This is the Bayes-Nash incentive compatibility proven in equations (3-9). For a more formal explanation of the incentive-compatibility conditions, see equation (3.5) from (Myerson, 1981). Given equation (4) and the assumption that bidder i chooses their action, based on truthful valuations by others, we can formulate how bidder i responds to these valuations according to Myerson's equations (4.5-4.6):

$$= \int_{v_{-i}} \left(Q_i(v_{-i}, x_i) + (v_i - x_i)\right) p_i\left(v_{-i}, x_i(Q_i)\right) - m_i(v_{-i}, x_i) f_{-i}(v_{-i}) dv_{-i}$$

$$= U_i(Q_i, m_i, x_i) + (v_i - x_i) c_i(Q_i, x_i) \text{ (18)}$$

Thus, connecting this to equation (8), and assuming that $v_i \geq x_i$, he concludes that utility is driven by the conditional probability of winning a network slice:

$$U_i(Q_i, m_i, v_i) \geq U_i(Q_i, m_i, x_i) + (v_i - x_i) c_i(Q_i, x_i) \text{ (19)}$$

Using (19) twice (once with the valuations switched), we get:

$$(v_i - x_i) c_i(Q_i, x_i) \leq U_i(Q_i, m_i, v_i) - U_i(Q_i, m_i, x_i) \leq (v_i - x_i) c_i(Q_i, v_i) \text{ (20)}$$

Then follows that, given $v_i \geq x_i$, $c_i(Q_i, v_i) \geq c_i(Q_i, x_i)$. In equation (20), we see the basic demand for incentive compatibility to hold in our modification of Vishna: that the QoS gained from package $Q_i(v_i, v_i)$ must always be at least equal to package $Q_i(x_i, v_i)$. Given the assumptions made ahead of equation (5) regarding the demand price never being non-decreasing with a rising valuation and a non-decreasing price elasticity, the inequality between v_i and x_i holds. It therefore follows that a higher valuation improves the conditional probability by improving the expected utility of your individual valuation. Thus, follows Myerson's conclusion that the utility gained must be higher, given a truthful valuation, and thus we have proven proposition (1).

Proposition 2: Bidder's utility is highest on its own assigned network slice: $U_i(v_i, Q_i) \geq U_i(v_i, Q_j)$

Built on the assumption that the orchestrator, as a neutral actor, creates the optimal network slice based on the non-quantitative demands the bidder gave to the orchestrator, we assume that the network slice created best fits these demands. A suboptimal network slice would therefore require strategically misreporting the demand, which is unlikely given that

(a) the bidder assigns no valuation to this demand, (b) this demand is not shared to other bidders, thus no hazard rate, and (c) this would harm their expected utility, given that they don't know the other composition of other network slices *ex-ante*, Q_{-i}. Thus, we can assume truthful reporting of demand and therefore an optimal Q_i.

Next, we must separate *ex-ante* and *ex-post* valuations, i.e. the expected and realised utility, of each bidder. Naturally, it is possible that *ex-post* the bidder would rather have Q_j than nothing:

$$\left(U_i(v_i, Q_i) \geq U_i\left(v_i, Q_j\right) \geq U_i(\emptyset)\right) (21)$$

However, this is only realised after Q_{-i} is revealed to the bidder. We'll look into this next section as well. *Ex-ante*, the bidder knows only that it's expected utility is best served by a network slice it's likeliest to win. Combining proposition (1), we see that the highest comparative valuation, thus conditional probability, is gained by bidding on the network slice that they designed themselves. This proves proposition (2).

On the usage of VCG mechanisms, these mechanisms critically need that the allocation is optimal, simply because the payments that the mechanism creates are intractable and irreversible (Bartal, Gonen, & Nisan, 2003). The benefit for our mechanism is that computational feasibility requires time for the orchestrator to create alternate allocations, which is possible in our allocation phase. Additionally, the value of creating personal network slices for the bidders to bid on allows for flexibility on what bidders bid for, without causing a problem in feasible optimization. The alternative way of creating alternatives, with bidders providing offers for separate resources is considered computationally infeasible, unless the amount of bidders and resources is small (Nisan & Ronen, 2007). This is supported by Lehmann et al., who show that the feasibility of a mechanism is reliant on reducing the possible types in the market (Lehmann, Oćallaghan, & Shohman, 2002). However, what Nisan and Ronen note is that the bidders consider the market based on the feasibly dominant actions available to them: actions that they have knowledge of. They will base their decision on this feasible best response, even if there are better strategies that they have no knowledge of (Nisan & Ronen, 2007). Thus, the orchestrator can trust that envy-freeness is guaranteed in a mechanism, where the bidder has no more optimal option available than the one created for them. The only reason for the bidder to lie, according to the authors, is to improve the algorithm to create a better result.

With the goal of maximizing the valuation of the network, the orchestrator is designed to create a market that rewards bidders for finding usages for the resources that provide the highest benefit to consumers, seen in equations (1) and (2). As a neutral principal in the market mechanism, the purpose of the orchestrator is to maximize the total utility of the market and all its participants through the metric of the total valuation. Neutrality also means that the orchestrator holds equal weight for the utility of both the sellers and the bidders and is willing to diminish the utility of one actor to improve the total utility. The possibility of favouritism by the orchestrator for some actor in the mechanism limits their discretion in the mechanism. This would drive the allocation process towards nonmanipulable factors like money (Laffont & Tirole, 1991). However, such a market would lack the ability to learn from previous allocation and re-allocation phases due to lower information rents for bidders, thus limit truthful bidding, and in the end would mean a non-optimal allocation during the allocation phase. We will look into this in the next subsection.

On the subject of what each bidder can bid on, it is the best option to keep the templates each bidder has been provided as private, so as to limit harmful strategic behaviour. The trade-off here is the ability for bidders to make more informed choices and possibly find an even better network slice option. However, this can be limited by accurate conversion of the template into the bidder's ideal network slice and by the creation of sufficient alternatives that limits the envy that bidders could have for other bidders' network slice templates. Moreover, the templates would become more accurate to the needs of the bidders due to social learning over time, as all actors learn more about the value of all the resources available. I will talk about this more in the re-allocation phase, section 4.4.

At its end, the orchestrator allocates all network services so that, given all network slices are operational at the same time, all could be serviced at the instantiation level (IL) they have bid for. Instantiation levels are the agreed upon Quality-of-service (QoS) levels that the network slices are willing to pay for, with exact guarantees on the minimum standards in latency, throughput, reliability and such that are created through the resources allocated to their network services. This allows them to show what the minimum level of service they're satisfied with is. A good example of this is given in section 4.2.1 through the smart grid example presented in (Dorsch et al., 2018). Thus, the service requests bidders provide give the orchestrator only a picture of what their demand is, allowing allocation to be efficient. The orchestrator simply needs what combination of excess resources network slices are willing to pay for, allocating these as a maximization problem.

4.3.1.3. Optimising Packages Ex-ante

As we noted in the previous section, the orchestrator might not achieve an optimal allocation of resources compared to the *ex-post* optimum, based on a lack of information. A bidder might have been better off *ex-post*, had they been given a non-optimal package by the orchestrator to bid for in the allocation phase. To make the comparative clear, let's take an example: company X seeks a network slice, which would achieve Q_X with any optimal combination of resources. An alternative composition would have all else the same, except computation resource B instead of A, which perhaps is in a different location. Thus, the latency and strain on the C-RAN resources would be slightly higher, causing the Quality of Service to drop to $Q_X - \varepsilon$. For simplicity, the bidder's valuation would also change to $v_X - \varepsilon$. The competition for resource A is very high, and other bidders have such high valuations for their packages that company X would lose with its pair of (Q_X, v_X), given that it is bidding for the package including resource A. Alternatively, there is less competition for resource B, and company X would win with its bid of $v_X - \varepsilon$. Thus, a naïve *ex-ante* allocation could create a situation in which the bidders have lost in the auction, although they shouldn't have in the view of *ex-post* efficient allocation.

A solution is provided by Nisan and Ronen in the form of a second-chance VCG mechanism: bidders are able to offer both their truthful valuation and an alternative bid that takes to account competition (Nisan & Ronen, 2007). In our version, this alternative would offer a slightly modified version of the optimal network slice, such as resource A being switched for resource B, as in the previous example. It would be possible to create more alternatives for the bidder, obviously tied to computational tractability of the mechanism itself, but we will limit our analysis on a two-option model. The orchestrator values maximum efficiency within the entire network, and thus might rather give a second chance to a bidder with a high valuation, but an optimal package that would be too highly competed.

This would be based on knowledge that the orchestrator has on bidder $i's$ situation, specifically the competition they are likely to face, with knowledge denoted as b_i. Each bidder has a type, v_i, under which they are given an optimal network slice, o_i, and a sub-optimal alternative network slice, l_i. Contrary to Nisan and Ronen's version, the orchestrator is the one that creates these two options based on the bidder's QoS requirements, and the bidder creates their valuations for them. Based on the knowledge the orchestrator has on the bidder situation, they create a slice formulation of $Q_i = \{o_i, l_i\}$ (Nisan & Ronen, 2007). Here, the orchestrator bases the alternative option on the knowledge they have on the other bidders, and what they are seeking, o_{-i}. Thus, the creation of optimal and alternative network slices is sequential: the orchestrator takes the optimal network slices it has created for all bidders to see what resources are likely to be too competitive for bidders, giving them an alternative slice that would help them avoid the competition. If the orchestrator were to base its knowledge on the valuations for the optimal network slice, bidders would gain information on the type of agent of other bidders, thus causing undesirable consequences (Nisan & Ronen, 2007).

To guarantee the truthful incentives of the bidders within our mechanism, it is key that the bidder themselves cannot create the alternative option for themselves. This would require them to hold knowledge on other bidders' valuations, $v_i(o_{-i})$, which could cause both strategic bids from them, as well as a fear of bidding truthfully for all other bidders. Moreover, this gives us another explanation why the orchestrator must be a neutral party: a seller would have high incentive to guide the most valuable network slices onto its own resources. Other sellers must trust the system to neutrally evaluate who uses their resources.

The change this does to our mechanism is quite simple: the orchestrator has two alternative valuations by the bidders for the two network slices to consider from when allocating resources to network slices. The maximization problem the orchestrator seeks to solve stays the same, maximizing based on the valuation of bidders, while restricted by the QoS guarantees and the total resources on offer. The valuation of each bidder is transformed into an array of valuations, $v_i = \{v_{o,i}, v_{l,i}\}$, with the orchestrator only able to choose a maximum of one of these. This is shown by equation (22e), which is added to (15-15d):

$$\max \sum_{i=1}^{I} v_i(Q_i) \; (22)$$

$$s.t. \sum_{i=1}^{I} Q_i \leq R \; (22a)$$

$$s.t \sum_{r=1,i=1}^{R} \sigma_{r,i} \geq Q_i \; (22b)$$

$$R = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix} (22c)$$

$$r_n = \sum_{i=1}^{I} \sigma_{n,i} \ (22d)$$

$$v_{o,i}, v_{l,i} \in \{0,1\} \ (22e)$$

4.3.2. Allocation of Resources

Given that the bidders have provided their network service allocation bids and sellers have provided their minimum pricing rates to the orchestrator, the allocation of the resources can commence. We will go through the orchestrator's decision-making in allocating the network services to the right bidders, looking at how it can be made efficient. We will also look at the creation phase of the network slices, leaning on the technical explanations in the 3rd section.

The key questions the allocation phase answers are: who gets what and what do they pay? It is critical to remind ourselves about the necessity of conditionality in all the bids that potential network slices make: they only find value from a complete network slice, with all the network services its operation requires. If, for example, a mobile video gaming network slice is lacking the network service that allows computation offloading, it cannot provide the latency rates that are needed for the service to function as intended (Mach & Becvar, 2017). Thus, the rest of the resources assigned are irrelevant for the network slice, simply because the sought-after service cannot be created. In a market where they face the risk of being stuck with a partial network slice, bidding would have to be strategic and consider the likelihood of bid failure in the value of the bid. Rather, we would want the bidders to freely bid for all the network services they seek, knowing that they will only pay if all their bids succeed.

To this effect, combining conditionality with the Clarke pivot rule for allocation of the network services and the payments that bidders must provide for their network slice does this in a way that guarantees truthful bidding. Using a price mechanism like the pivot rule, a mechanism in which the valuation creates the allocation but not the pricing of resources, has been proven to be effective in pushing bidders to be truthful with private information on the value of the bid object (L. M. Ausubel & Milgrom, 2005). The dominant strategy of being truthful is subject to assumptions regarding the bidder knowing their private value, not caring about the identity of other bidders and what they pay, and that budget constraints must never be binding (Milgrom, 2000). The auction design isn't guaranteed to provide the highest possible payment to the sellers, but can help allocate the network services to the bidders with the highest valuation (Myerson, 1981).

A critical technological concept we need to briefly delve into is punctured scheduling, which allows for data traffic of different priorities and latency and reliability requirements to operate on the same network (Esswie & Pedersen, 2018; Pedersen, Pocovi, Steiner, & Khosravirad, 2018). In its simplicity, punctured scheduling is the way in which more critical data transfers are given priority by the virtualized base-station unit for the antenna it operates. The current solutions are built on algorithms that decide on the order of downlink traffic based on maximizing the transfer rate for eMBB users and minimizing the latency of URLLC users. The authors use puncturing to mean more time-critical data slotting into the scheduled transmission of another user that is deemed to be less time-sensitive (see figure 9). Thus, we are able to not only combine different users into the same network as network slices, but also to have different types of usage cases in the same network without an issue of Quality of Service requirements being failed.

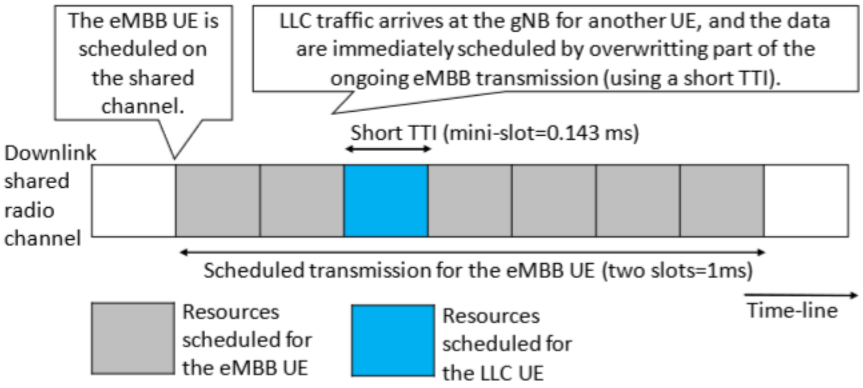

The eMBB UE is scheduled on the shared channel.
LLC traffic arrives at the gNB for another UE, and the data are immediately scheduled by overwritting part of the ongoing eMBB transmission (using a short TTI).
Short TTI (mini-slot=0.143 ms)
Downlink shared radio channel
Scheduled transmission for the eMBB UE (two slots=1ms)
Time-line
Resources scheduled for the eMBB UE
Resources scheduled for the LLC UE

Figure 9: The basic principle of downlink punctured scheduling. A more time-critical user (in blue) punctures the scheduled transmission of an eMBB (e.g. streaming) user (in grey). Source: Pedersen et al.

4.4. Re-allocation Phase

Once the network slices have been created with the appropriate resources and all network services have been integrated together to create functional network slices, each network slice would now be guaranteed their minimal instantiation level, IL_0 whenever they want to use it. However, the network slices are unlikely to all operate on this level simultaneously. Thus, there are resources that are not being used by any network slice and could be formed into a network service that some other network slice would value. This is something critically missing in current literature on network slicing: they formulate resource allocation as a static and one-off action, with network slices guaranteed a certain range of performance that they can freely use (Jiang et al., 2017; Ordonez-Lucena et al., 2018; Tadayon & Aissa, 2018). Another alternative in the literature is an allocation mechanism focused on maximizing network utilization and reducing QoS violations, but without mechanisms on allocating resources to the most efficient users (Sciancalepore, Cirillo, & Costa-Perez, 2017). Combined, that is specifically what I will seek to add: how to optimize the functionality of the network services in a dynamic setting to re-allocate resources to those with the highest valuation at each time. The reality is that most network slices need their resources at a different time, and at a different intensity. Efficiency comes from differentiating the valuation of different service requests, lacking in current literature. Therefore, the orchestrator must know where an unused marginal resource should be allocated within the market, based on who needs it the most. By re-allocating unused resources, we can achieve a higher usage rate for the resources on offer in the mechanism, creating more efficiency. An efficient re-allocation would generate more value from all network slices and higher welfare for consumers of the networks, without harming the isolation or functionality of all slices.

4.4.1. Mechanism Design

4.4.1.1 Basic Design Structure

To clarify the structure of the re-allocation phase, the phase will be formed into a sequence of bidding rounds, where the network services are formed for each round by the orchestrator and are re-allocated to the winning bidders. This period could be 30 minutes, with windows ending 15 minutes before for bidders to form their next bids, although this suggestion is not fixed. Critically, the usage of network slices can be modelled as partially predictable and partially random: there is a certain level of usage that the network slices can expect from their end users, while a part of the resource usage is driven by random decisions of end users. This of course is reliant on the business model of the network slice: the resource demand of a smart grid slice is more predictable than a video stream slice, since its operations are more pre-planned.

We'll separate between entitled network slices, slices that have a claim to resources based on the allocation phase, and re-allocation bidding network slices, slices that are seeking to use vacant network services above their own allocated network services. Entitled network slices are entitled to their minimal instantiation level of service without taking part in the re-allocation round, and is prioritized above all re-allocation bidding network slices in each round: if they seek to use their network slice at any time, the necessary resources will be vacated from re-allocation bidding slices to guarantee their quality service until the limit of their IL_0. The re-allocation phase simply seeks to re-allocate vacant network services to all of the bidders in the second group during each round of the phase, effectively sharing the crumbs that are not assigned between them.

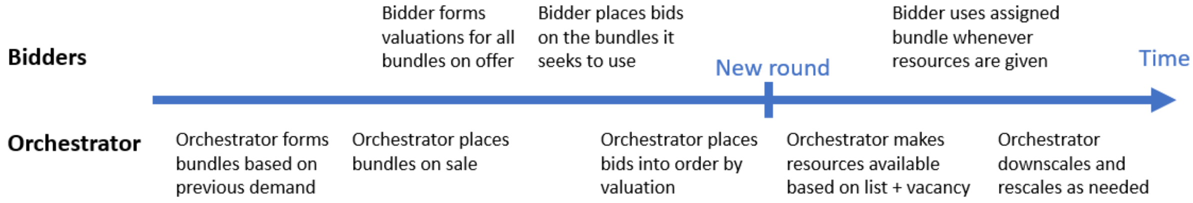

Bidders
Bidder forms valuations for all bundles on offer
Bidder places bids on the bundles it seeks to use
New round
Bidder uses assigned bundle whenever resources are given
Time
Orchestrator
Orchestrator forms bundles based on previous demand
Orchestrator places bundles on sale
Orchestrator places bids into order by valuation
Orchestrator makes resources available based on list + vacancy
Orchestrator downscales and rescales as needed

Figure 10: Timeline of a re-allocation phase round

For the re-allocation phase, resources need to be packaged in a way that allows for competition between network slices in the same auctions, while mitigating the harm of conditionality. This is why a mechanism handling each resource individually is sub-optimal: bidders would have trouble combining these individual resources together to create the service they need. On the other end of the scale, a similar allocation strategy as used in the allocation phase, pre-packaging resources into a higher instantiation level of the complete network slice (see figure 6), would be infeasible, due to the uncertainty that all resources would be available. Since all resources cannot be guaranteed to bidders, due to entitled network slices from the allocation phase having a right to these resources at will, the packaging loses its value. Rather, a happy middle is found from the orchestrator creating bundles that would then be re-allocated and removed according to valuations.

Critically, the orchestrator cannot know what the usage rate for resources on its networks will be, though it can estimate the demand through previous usage rates. Therefore, any pre-packaging by the orchestrator of network services together would fail on the problem of conditionality: the failure rate, the likelihood that a bundle is not available throughout the re-allocation period, is not the same for all network services for the bidder. The pre-packaging practiced in the allocation phase would be incompatible with a re-allocation mechanism, since bidders would not be competing for the same priority levels with their individual packages.

Accordingly, we would create two categories of bundles: network resources bundled either with computation resources or storage resources. All additional operation by a network slice above its own minimum instantiation level needs additional network resources to accommodate this, and therefore each bundle must contain them. Moreover, by splitting up computation and storage resources, we can separate the two broad usage cases for these re-allocated resources: the creation of over-the-top solutions that require computation power (Taleb et al., 2017) and smart caching solutions that require storage resources (Taleb et al., 2016).

To utilise the example presented in section 4.2.1. about a network slice for a smart grid solution (Dorsch et al., 2018), recall that they had a set of time-critical but small data packages that needed to be transferred, as well as much larger, time-flexible packages. While its minimum network would serve these smaller packages, it would likely compete for the computation and network resource bundle to transfer these larger packages. This bidder would likely only bid for these bundles, and would seek to bid for them during downtimes of the entire network, such as during the night. It is important to note, however, that network slices are not limited in how they can use these bundles: rather than providing them with specific network services, we are offering them additional capacity to operate their own network services. There can be some variation within the bundles, such as different amounts of spectrum and C-RAN resources to create variation in provided bandwidth, throughput, delay and reliability. This would allow bidders seeking resources for time-critical or reliability-critical services to value resources differently, without causing a fragmentation of the auctions themselves.

Bidders are not limited from bidding for any of the bundles, but realistically they are only going to compete for bundles within their own network: for example, a computation bundle with both network and computation resources that are from the same sellers they are currently receiving resources from. Accordingly, the bundles would be formed underneath these market structures: the orchestrator would seek to bundle resources together as much as possible to allow bidders to compete for bundles available in their own network. Although a network slice could use a re-allocated bundle outside its current network topology, they would face issues in connectivity, such as the application programming interface (API) that the applications operating the resources use. By comparison, a bundle already under the network slice's network would simply be re-allocated by the SFC operator within milliseconds. The trade-off here is the possibility of demand mismatch: concurrent overdemand and oversupply of the same type of bundle due to the difficulty to re-allocate resources. However, the re-re-allocation of resources wouldn't face difficulty, making the solution technologically feasible.

Each bundle can be produced for a finite amount of times, but each instantiation of the bundle would be produced on different underlying physical resources: during downtimes for network slices there would likely be a lot of resources available to create these bundles, compared to a time of high demand. In other words, the critical part of competition here would be priority: who gets a re-allocated bundle and whose bundle would be re-re-allocated to an entitled network slice when they demand it? Here, we can borrow from internet advertising auctions and generalized second-price (GSP) auctions (Edelman, Ostrovsky, & Schwarz, 2005). For each competed bundle, bidders provide bids, denoted as b_i, for priority slots that are assigned in decreasing order of bids. Each bidder is allowed to offer a single valuation for a bundle. There are C slots available, with N bidders competing for them: $C = \{s_1, s_2, \ldots s_C\}$. A key design is to have more bidders than available slots, $C < N$. Each bidder that wins a slot pays the bid directly below them, i.e. the winner of slot s_n pays b^{s+1}.

The higher up on the priority list, the less likely the bundle will be re-re-allocated from you, and therefore more valuable the position is. Contrary to the concerns in sponsored ad usage that bidders might not value slots as a strictly increasing function due to clickthrough rates being affected by the user, we do not have this concern. Since the only effect on valuation between s_n and s_{n+1} is the expected probability of using it *ex-ante*, we can assume incentive compatibility, as proven in equations (17-20). Each time an entitled network slice uses its network again, the bundles their slice uses is taken from the bidding network slices that are lowest on the re-allocation priority list (see figure 11). For example, if the resources were formed into 100 equal bundles, we'd see that the 101st ranked bid in figure

11 would never receive any bundle. Moreover, if entitled network slices were using 40% of the resources, the 61st ranked bid would be the first losing bid.

The value of such an allocation mechanism is that it does not require any specific usage of the network to work: priority levels allow for a division of the vacant resource, no matter how much of it is available at any moment, with variation only in the number of winners in re-allocation. The payment of each bidder is dependent on how long they were able to utilize the bundle: if it was re-re-allocated away from them halfway through the round, for example, they would only pay for half of their bid. Moreover, a mechanism like this allows for a grace period of the payment being contingent on continuous availability beyond a minute, for example. This prevents questions of incentive compatibility, where a bidder is paying for resources that are constantly given and pulled away from them, causing payment but barely any usability. So, a bidder that only receives a bundle seven times for a 10-second period each would not pay for this. The ideal grace period would require further investigation of the technological limitations and is thus outside the scope of this thesis.

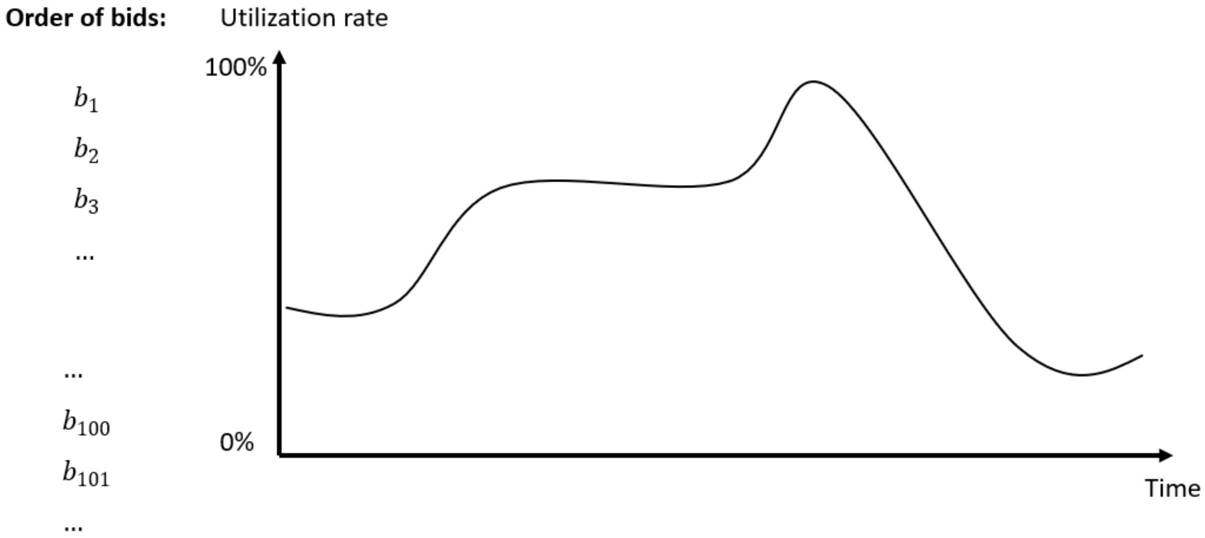


Figure 11: An example of GSP ordering of bids – all bids above the utilization rate would gain an allocated bundle at each point in time

In order to run the dynamic re-allocation, the mechanism needs a system that can dynamically redirect the service function path (SFP) of a user within a virtual network, so that the orchestrator of the network can re-allocate resources to a network slice seeking claim to its minimal instantiation level from other users in real-time. SFPs are the routes that data packages take through a virtualized network service in the network topology (Medhat et al., 2017). This is the core function of a service function chaining -solution (SFC), creating a management model that makes the routing decisions for a network (Halpern & Pignataro, 2015).

There are several SFC solutions that are possible to use, all differing in how they implement the chaining, that are evaluated in (Medhat et al., 2017). The authors look at how the control and data planes are managed by the software, and how good the flexibility and scalability of the software is. Flexibility is defined as "the efficiency of the traffic steering scheme implemented in the SFC solution" and scalability is based on "the number of rules needed to apply traffic steering for one chain". In our case, both are important, but having an efficient traffic steering scheme is paramount to us. For our usage, (Csoma et al., 2014) provides us with the best possible flexibility with a dynamic pathing solution in its ESCAPE tool, and thus would be what our re-allocation would be based on. I will not go into detail on how the software paths traffic, as this would go outside the scope of this paper, but for us the critical part is the software's ability to dynamically re-allocated resources within its own network in milliseconds, based on given policies. These policies would be based on maximizing the utility of the mechanism. Combined with the punctured scheduling explained in the allocation phase, there is a technological capacity to re-scheduling and re-path data in a way that allows for re-allocation to function.

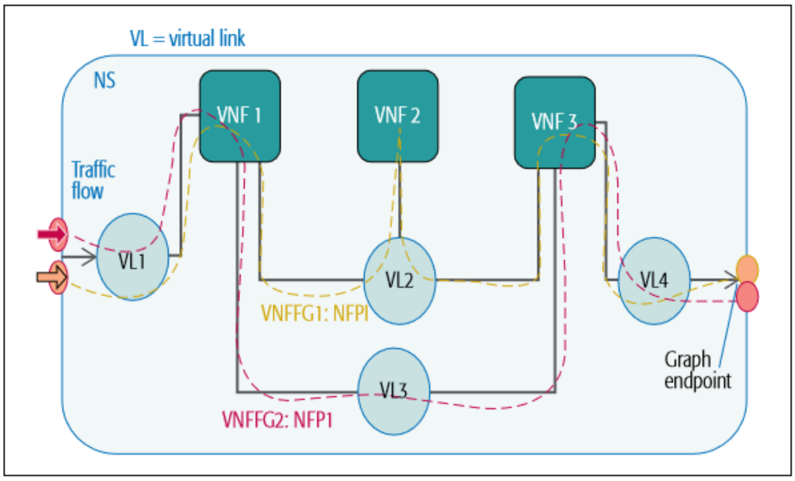
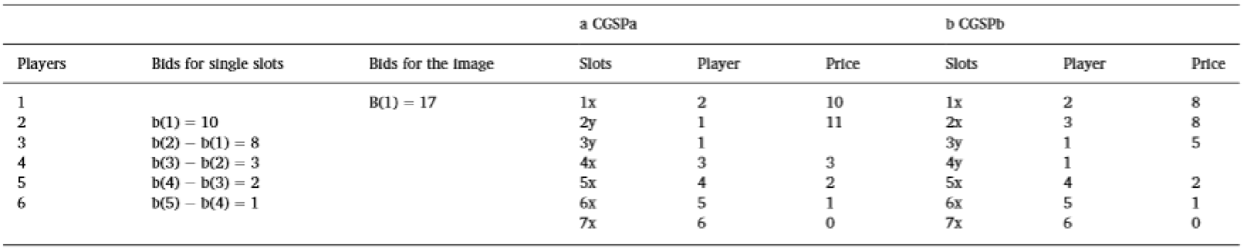


Figure 12: Example of two service function paths through a network service with Virtual Network Functions and Virtual Links. Source: Medhat et al.

4.4.1.2. Extension to Basic Mechanism

Conditionality is still a crucial problem within each bundle, even given the GSP mechanism, since bidders seek different levels of service. Video streaming services that serve a larger population like Netflix have demand for a vastly larger bundle, compared to a smaller service that offers video conference services for a few companies, for example. Thus, while the smaller company would only bid for one instantiation of the bundle, Netflix would want to bid for multiple instantiations. A solution offered by (Dimitri, 2018) creates a Combinatorial Generalized Second-Price (CGSP) auction, where bidders can either bid for one or multiple sequential slots. Sequential bids mean that the issue of only a part of the network services the bidder bids for being allocated to them is minimized, although it causes some inefficiency from the GSP auction requiring all bidders to make one bid. There is a possibility that a bidder seeking three units of the bundle loses, although their valuation for two units would have been high enough to win. However, the trade-off from a more feasible and understandable bidding system for the network slices is worth this slight inefficiency.

As in the current literature, the price for position s_n in the list is the bid b_{n+1}, as a general rule. However, with the CGSP auction structure, this is made more complicated with bids for a package of bundles amongst the single bids for priority: bidders are competing for the same item, but in different quantities. The general rule is still kept in place in the combinatorial GSP auction by Dimitri: the valuation of a package should be higher than the combined value of other bidders seeking the positions of the package (Dimitri, 2018). So, let's look at a market with three slots and five bidders, one seeking both positions and the four others seeking only one slot, with valuations for those single slots of 5, 7, 8 and 9, respectively. Here, the bidder seeking the package would need to have a valuation of 17 or higher to win the first slot. The bidder pays the higher option between either the combined single slot bids or the next highest equal package. If bidders have equal valuations, the orchestrator would randomly allocate the slots s_n and s_{n+1} between the package bidder and the highest single slot bidder. If the package-seeking bidder has a valuation of 16 or less, this would mean the package wouldn't get first slot. Dimitri offers two options in terms of allocation here: either the highest single slot bidder wins the first slot and the comparative is done between the package and the 2nd and 3rd highest single slot bidder (notated as CGSPa), or then both the 1st and 2nd highest single slot bidders get the first two slots (notated as CGSPb). A comparative example is given below:



			a CGSPa			b CGSPb		
Players	Bids for single slots	Bids for the image	Slots	Player	Price	Slots	Player	Price
1		B(1) = 17	1x	2	10	1x	2	8
2	b(1) = 10		2y	1	11	2x	3	8
3	b(2) − b(1) = 8		3y	1		3y	1	5
4	b(3) − b(2) = 3		4x	3	3	4y	1	
5	b(4) − b(3) = 2		5x	4	2	5x	4	2
6	b(5) − b(4) = 1		6x	5	1	6x	5	1
			7x	6	0	7x	6	0

Table 2: Comparing slot allocation with CGSPa and CGSPb. Source: Dimitri (2018)

What we can see from table 2 is a competition for seven slots with six bidders, one of which is seeking what the author calls an image, a package of two consecutive slots. With a valuation of 17, the package is less valuable than the combined bid of players 2 and 3. In CGSPa, player 2 gets the highest slot, and the comparison is again done between the package and the bids of players 3 and 4, where the package is valued higher to give player 1 slots 2 and 3. In CGSPb, this second comparison is not done, but rather players 2 and 3 gets slots 1 and 2, and the package is compared against the bids of players 4 and 5, where the package bidder wins to get slots 3 and 4. So, although the rules rely on similar principles, they can produce somewhat different results.

4.4.1.3. Pre-planning of the Market

The demand for higher performance can be driven by either push or pull effects, just like in the allocation phase. The first group would be one with a constant demand, where higher performance has no additional value, and thus they are not willing to pay for more. We'll call this group constant demand bidders. Demand can also be driven by a network slice perceiving value from a higher instantiation level, with the key being a predictable performance. They have their valuations for additional resources driven by firm-level demand for higher performance, which is tied to the time the resource is available. We'll call these bidders in the re-allocation phase fixed demand bidders. The third variant will have higher instantiation level requests being pushed by customer demand, such as a surge in network activity in a certain area. This demand is a more dynamic one, since the demand cannot be pre-planned, but can only be estimated. We'll call this group dynamic demand bidders. We'll look further into these bidders in section 5.2, discussing network composition.

A system in which a bidder can seek out resources whenever they wish would be computationally irretractable, simply because of the constant fluctuations that would cause constant recalculation of the re-allocations. A more realistic mechanism would be based on scheduled valuations for pre-designed bundles that would potentially become available, based on the network services that are left unused by other network slices on their minimal instantiation level. This is critical: the complete pre-packaging used in the allocation phase is untenable due to the uncertainty of what resources will become available. Alternatively, creating separate bidding for individual resources would create too large an exposure problem to the bidders from the conditionality of the bids.

For such a market of bundles, the network slices would have a clear understanding of what they would get, and thus there would be the ability to plan and schedule bids for these bundles. Even the dynamic demand bidders, such as streaming or video communication services, can fairly accurately estimate the future demand for their services in terms of when, where and how much their services will be used. There is a need for estimation of demand that means that over- and underbidding for resources happens, but this would diminish over time with the services becoming better at predicting usage by their customers. More importantly, this is a trade-off worth taking, with the computation by the orchestrator being less strained and now tractable, while the largest harm to the network slice would be the necessity of scaling down services slightly when they have underbid.

One of the core things the sellers in the mechanism seek is the ability to maximize the value of their provided resources. This means that the network should have as high a utilization rate as possible, along with those most willing to pay always being offered resources. The bidders meanwhile want a system where they can always be assigned resources when they have demand for it, but not to lose it if they value it heavily. Finally, the orchestrator is designed to make the market as efficient as possible, with bundles constantly being allocated to those that value them most, creating as much surplus as possible.

4.4.2. Downscaling and Re-re-allocation

A key part of the proposed mechanism is the guarantee that the Quality of Service agreed upon in the allocation phase is always available to the network slice until its minimal instantiation level, and this still needs to be guaranteed throughout the re-allocation phase. Thus, any entitled network slice must always be able to utilize the package that they won during the allocation phase, and those underlying resources need to be returned. This requires downscaling the resources allocated to at least one of the winners in the re-allocation phase, and the weighted proportional fairness criterion dictates that this should happen to the bidder with the lowest valuation (Afèche, 2006, p. 110). Thus, the priority list is used as the driver of re-re-allocation: the orchestrator simply takes the lowest valuation winning bidder for each bundle that is now returned to an entitled network slice and passes their bundles on.

Downscaling brings another benefit for packaging vacant resources into bundles for re-allocation: the impact of an entitled network slice reusing its resources can be contained into the bundles that their usage takes up, without limiting any other re-allocated bundles. The entitled network slice is provided with resources that guarantee the Quality of Service that they won in the allocation phase, meaning that the impact on re-allocated bundles is limited, if there is a limited usage of the entitled network slice. Critically, we limit the impact of a more conditional package, and how this would need to be downscaled. The trade-off naturally is the conditionality that some bidders have for the bundles that they bid for during re-allocation, but the harm is more limited when they still have their minimal instantiation level that is guaranteed to them. Bottlenecks will be formed due to re-re-allocation, but the network slice is still capable of operating.

On a technological level, downscaling of resources will operate in the same way that it is done in the opposite direction during re-allocation: the entitled network slice demands resources for its operations, thus triggering a need for downscaling. The management and orchestrator (MANO) plane (see figure 4 and section 2.4) of the resources that need to be re-re-allocated redesigns the pathing of the concerned network slices within its own networks, increasing the resources allocated to the entitled network slice and vice versa. Here, the value of virtualized resources comes to the forefront: this dynamic re-allocation happens in milliseconds, or as soon as the MANO plane can decide on its re-pathing decisions. The entitled network slice does not face any buffering in utilizing its package, and thus doesn’t have any disutility in participating in a market including re-allocation.

4.4.3. Bidder Valuation

A key thing for us to note is that not all bundles available would be ideal for each bidder, and thus their valuation would be based on the realized service, rather than simply their demand for service. A CDNaaS (content distribution network as-a-Service) bundle that is on the opposite side of the country than where the customers of all streaming network slices are would have little value to those network slices, even if they would be seeking extra resources. A CDNaaS is a combination of network and storage resources, that allow on-demand access to stored data in data centres, e.g. a Netflix user gaining access to a TV show that is stored at a data centre. Alternatively, a computation warehouse with excess computational resources would be of little value to a network slice that is seeking more storage resources, and thus their valuation for the available network service would be low.

We thus have to understand how exactly bidders would value an excess network service, since this understanding allows us to tailor the bundles on offer more exactly to the bidders and to develop the allocation process through social learning. Better understanding would allow the mechanism to allocate and re-allocate more efficiently, creating greater value from the same resources. Moreover, not all bidders can compete for all the bundles available: the underlying resources would need to be a part of their network topology to be usable by the network slice itself. As an example, being allocated more C-RAN resources is only valuable if they are underneath the same MNO that the bidder's network slice is on. This limits the competition within each bidding for network services, but is a necessary part of the way network slicing works, and thus cannot be avoided.

The next thing we will seek to do is answer how bidders evaluate the value of excess bundles added into the re-allocation market, and subsequently how much they are willing to pay for them. This is driven by internal calculations of what the entailed resources can do to improve performance and customer satisfaction with their service. As an example, given that a data centre has more excess storage capacity on offer for re-allocation than computation resources and these were packaged up together, this would be more valuable to network slices seeking for higher CDNaaS capacity than one seeking TOFaaS (traffic offloading as-a-Service) capacity, which allow selectively offloading traffic off a network to lighten its load (Taleb et al., 2016). Moreover, with the value in CDNaaS coming from the ability to cache popular content, such as a hit show on Netflix or HBO, the streaming services would likely seek to respond to local push demand for certain content with bigger caches of data (Taleb et al., 2016). Each network slice will have different demands from the excess resources they seek, and thus have different valuations for them. In effect, this means that the re-allocation market will likely have network slices competing on a wide range of bundles.

For a feasible system, the bids that the network slices offer in the re-allocation phase must be offered in advance of the excess resources becoming available. Such a combinatorial auction would have too many types of bidders to be computationally feasible, especially given the running-time that we seek to enforce in this situation (Lehmann et al., 2002). We thus make a trade-off of a slightly suboptimal solution for the re-allocation, in exchange for the ability for the mechanism to run in real-time. So, the bidder herself has to give a pre-planned value for any resource that becomes available for bidding in the re-allocation phase. Next, we will look at the factors that would drive this decision. Effectively, the valuation of a network slice for an available bundle on offer is based on how close to their ideal sought network service it is. The closer to the optimal network service the offered network service is for them, the closer their valuation is to their optimal valuation.

$$v_i(Q_i) = v_{optimal}[w_c(\theta_c) + w_s(\theta_s) + w_n(\theta_n)] \ (23)$$

$$v_{optimal} = \max_{Q_i} v_i(Q_i) \ (24)$$

The bidder has an optimal valuation that is based on their demand for a bundle that would exactly provide what they need, i.e. the most they are willing to pay for exactly what they want. In reality, the bundles available would likely not be perfect for any bidder, and thus all bids would be a part of this optimal value. The bidder has values w_c, w_s, w_n for the weight of computer, storage and network resources respectively within the bundle they are seeking. Each type of resource has its own set of coefficients, marked by θ, that are private information on how the bidder values performance, quantity and quality measurements of the bundle. Although bidders' *ex-post* utilities from a re-allocated bundle are linked to the likelihood of winning and how long they have the resources, this does not affect their *ex-ante* valuation of the resource, and thus their bid itself.

Now, let's look at some of the coefficients bidders would have for different parts of bundles. This isn't intended to be exhaustive or provide monetary values for bidder valuation, rather to give the reader an idea of what sort of things bidders value. First, a bidder has different valuations for excess resources based on the time they are available. A network slice that connects the sensors within a factory complex together, for example, would value excess computation and network resources more when the factory is at full production, rather than when it is on minimum capacity. They can plan a schedule for their connectivity needs, which would be tied to the value they place for excess resources that would allow this. Even a customer-demand driven network slice like a streaming or video service would have estimated needs, and thus valuations, for excess resources that could be scheduled for the re-allocation phase in advance. Additionally, since temporal requirements are not equally critical to all network slices, a re-allocation mechanism would incentivize the usage of cheaper resources during off-hours for those that are flexible time-wise.

Next, the location of the resources, and thus the ability of the bundle to only provide services in a fixed area, is valued by all bidders. This is especially true for network slices that seek CDNaaS services, as talked about earlier, who might have increasing demand in specific areas. Regionally operating network slices would also find a network service that is available on the wrong side of the country valueless. On the opposing side of location is the coverage that a bundle can improve their current services in. This is only relevant for network resources, such as the C-RAN resources available, which have a certain area that they can cover. Again, the locality of some network slices is going to limit their valuation to resources, likely meaning that C-RAN resources in a bundle in urban areas will likely be valued much higher than those in identical bundles in more rural areas.

Another set of coefficients that bidders will be concerned about is the performance of the network service they'd gain from the bundle, such as the effective throughput, latency and reliability of the underlying resources. Latency, for example, is partially defined by the physical distance between the user equipment and the edge computing resources, but also by the performance and load of the computation resource (3GPP, 2018) and the C-RAN resource connecting to it. The bidders may have either a constant valuation for better performance of their network slice, or they may have a certain cut-off point for the performance of a network service. Thus, it is likely that bidders will have vastly different understandings and valuations for the performance a bundle can provide.

4.4.4. Revenue Efficiency and Equilibrium

To understand the equilibrium that is formed within the priority list, it is useful to think of each bidder's situation in the form of up-Nash and down-Nash from (Lucier, Paes Leme, & Tardos, 2012). Up-Nash is when a bidder can't improve its utility by taking a slot above it on the list, while down-Nash means that they can't improve their utility by taking a slot beneath them on the list. Bidders that get a slot on the list must satisfy both up-Nash and down-Nash conditions, while bidders that didn't get a slot must satisfy the up-Nash conditions (Lucier et al., 2012). Moreover, a set of bidding functions is a Bayes-Nash equilibrium, if for all bidders, expected utility is highest by bidding truthfully, so b_i, not b'_i:

$$E[U_i(b_i(v_i), b_{-i}(v_{-i}))] \geq E\left[U_i\left(b'_i(v_i), b_{-i}(v_{-i})\right)\right] \quad (25)$$

We must also briefly look into the information conditions that apply in our mechanism. In the GSP literature, the main consideration is between a Bayesian partial information game, where bidders know their own valuations and only that other bidders are picked from independent and identically distributed valuations, and a full information model, where each auction is played repeatedly with the same group of bidders and the bids stabilize due to repeated play (Caragiannis & Kaklamanis, 2011). Clearly, our bidding game is somewhere in between these two settings, but closer to a full information game. Online advertisement auctions are based on constantly changing keywords, bidder preferences that may vary significantly over time, and advertisement quality scores that match ads that customers will be more receptive towards (Caragiannis & Kaklamanis, 2011). In our game, the bundles on offer will have little variation over time, bidder preferences are more stable and predictable over the long run, and the priorities would only be based on valuation. So, although the bidders cannot know the valuations of their competition, they will have a good estimate of the other bidders as a collective, especially over time. Thus, we will model the behaviour of the bidders based on a full information game.

This also brings us to the question of reserve price: should the orchestrator seek to improve the revenue through a minimum price in the mechanism? Here, a reserve price r means that any bidder with value $v_i < r$ would be dropped, after which the GSP mechanism would then be run. A reserve price would not serve any necessity to cover costs, since the underlying resources would still need to be constantly available to the entitled network slices, and the operating costs for the SFC software necessary for dynamic re-allocation is diminishingly small. The classic Myerson argument for a reservation price, that the orchestrator should rather keep the item on offer to themselves than give it to an "infinitesimal" small bid (Myerson, 1981), is tough to achieve when the desired "appropriately chosen reserve price" is hard to define. However, using one in an experimental setting has been found to improve the revenue of a GSP auction (Edelman & Schwarz, 2010; Ostrovsky & Schwarz, 2010). This experimental setting was built on the Generalized English Auction, with the rising prices offered by bidders, and seems to be designed to force the bids of bidders upwards via the reserve price. This would imply that their bids without a reserve price are below their true valuation, going in conflict with the truthful valuation mechanism of second-price auctions, and thus is perhaps a bit outside my thesis scope. Such determination would require empirical research.

(Edelman et al., 2005) and (Varian, 2009) have shown that the full information game always has a pure Nash equilibrium, and that this equilibrium has the same outcome and payments as a VCG mechanism, the comparative mechanism to a GSP auction. In the comparative, unlike a VCG mechanism, a GSP auction does not have an equilibrium in dominant strategies and does not have truth-telling as an equilibrium either in a Bayesian game, when there is incomplete information (Edelman et al., 2005). Edelman et al. classify this category of equilibrium as envy-free equilibria: bidders do not prefer any other bidders' position in the mechanism *ex-post*. Critically, the value of an envy-free equilibrium is that the bid profiles are always efficient and the revenue is always greater than or equal to the VCG revenue (Lucier et al., 2012).

Another question we need to answer is the harm that comes to a bidder from the bundle they have won in a re-allocation round being taken away from them. This would happen if re-re-allocation was necessary to provide an entitled network slice resources, which would require the lowest valuation re-allocation bidder to lose their bundle. If the bidder loses a lot of utility from seeing their resources taken away from them by a higher bid before the end of the round, then they face a hazard rate they take into account for their *ex-ante* valuation. Were the individual rationality of the bidder be disrupted, i.e. not bidding at all having a higher utility than losing the resource halfway through the round, then the bidder would not wish to take part in the phase or might push them to bid strategically. A part of the harm is mitigated by the technological design of the mechanism: since the re-re-allocation of resources is done by software within the same network, the entitled network slice will see no change in service, thus removing this possible harm from the problem. However, the bidder in the re-allocation is definitely affected.

This hazard rate is driven by an *ex-post* valuation that the bidder holds: knowing that they might not get excess resources for the entire re-allocation phase might affect their valuation *ex-ante*. This might cause strategic behaviour, in that bidders would bid lower than their true valuation for the bundle. The mechanism of the re-allocation phase is uncertain in the resources it provides by design: even the highest bid might not get any resources, since all entitled network slices might want their resources at the same time. Bidders that want guaranteed resources beyond their minimal instantiation level would bid for a bigger network slice in the allocation phase, and would not take part in the re-allocation phase. However, as long as the up-Nash and down-Nash requirements of the equilibrium are met, hazard rate of losing a bundle is no problem. A bidder with this uncertainty would not hold envy for any position above or below them *ex-ante*, but only *ex-post* when the uncertainty is realised.

4.4.5. Social learning by orchestrator

With each round of subsequent allocation phases after the original round and its re-allocation phase, the orchestrator is better able to understand the value that resources, and their packaging, have on the market utility gained. The substructure of each MNO’s network can also be optimized to maximize utility in the following re-allocation phase, even if it causes lost revenue in the allocation phase. A constant higher demand for a certain resource within the re-allocation phase can also be used as a signal by the sellers to increase their capacity to profit from the higher demand.

Given that bidding based on your true valuation is a weakly dominant strategy in a second-price auction (Vickrey, 1961), the orchestrator is better able to understand the value of the resources on the market, based on the bids that are placed during re-allocation offerings. The actual valuation equations used by the bidders themselves is still private information, but the outputs that they yield for the re-allocation phase allow for the orchestrator to better understand what sort of bundles are perceived as valuable, thus what combination of resources is valued. At its core, the early network service offerings are based on heuristics and a pre-network slicing understanding of the market, centred around the service templates that all bidders provided for the allocation phase. Over time, the ability to collect truthful bids for a collection of bundles that would be forced to variate due to fluctuating availability of excess resources means that the orchestrator is able to track when and why bids went up and down during re-allocation.

5. Discussion

5.1. Fairness, Efficacy and Efficiency of Re-allocation

A re-allocation mechanism is both different to the commonplace in resource management, as well as instinctively unfair for seemingly re-using resources that have already been allocated to some actor in the market. Without a clear understanding why this sort of a mechanism is useful, has limited downsides, and is ultimately necessary, it is hard to justify the use of dynamic re-allocation. To do this, I will consider the value of the efficiency trade-off from adding more computation work for the mechanism in exchange of a higher usage rate of resources, thus showing the efficacy of the mechanism. Moreover, I will also look at why the re-allocation mechanism requires the allocation phase before it for the market to function fairly and effectively. To be clear, there will naturally be some bidders and sellers in the market that would benefit more from a static allocation mechanism, but I will show that their loss in the market is fair in that it fulfils the weighted proportional fairness criterion (Afèche, 2006).

The definition of a system's efficacy is that the intended result should be achieved: for us, the intended result is that at each point in time the available resources are allocated to those users that value them the most, to the best of the system's abilities. Thus, the system shouldn't be measured on its achieved allocation efficiency *ex-post*, but rather on the decisions made being the best possible ones *ex-ante.* Here, the allocated resource should end up with those that value them the highest in both the allocation and re-allocation phase. Although our system cannot guarantee this *ex-post*, such as in cases of a bidder arriving late for a re-allocation phase with a higher valuation, but the *ex-ante* valuation ranking in both phases guarantee that those with a higher valuation than the current marginal bidder never lose out. Furthermore, as was shown by (Maskin & Riley, 1989) and in equations (3-9), the bidder has no incentive to lie about their valuation. Thus, the valuations are the best responses bidders have for the resources on offer, and though these valuations might change or be inaccurate, this still satisfies the efficacy of the system *ex-ante*.

A question of efficiency can also be asked regarding a possible change in valuation between static and dynamic allocation: whether or not bidders change their reported valuation due to the system. Critically, when compared to a static allocation model like the one offered by (Ordonez-Lucena et al., 2018), the valuation of receiving their minimal instantiation level on both systems is the same. The key differences come from the way resources above the minimal instantiation level are valued: are they pre-defined during the allocation phase, or are they decided upon during the re-allocation phase. In Mierendorff's work on dynamic Vickrey auctions, he notes the issue of bidders and varying valuations arriving over time, as well as their *ex-post* participation constraints (Mierendorff, 2013). The latter one regards a bidder with an infinite negative utility if their *ex-post* payoff is below zero, i.e. they lost out in the bidding process (Ollier & Thomas, 2013). Critically, in a bidding process like network slicing will create, where the bidders have unclear understanding of how much they value network slice, they seek to avoid this negative payoff at all costs. Thus, a system that allows them to value their minimal participation and to bid for higher performance dynamically minimizes the hazard rate of overvaluing resources. Moreover, as Mierendorff notes, dynamic allocation allows for the orchestrator to seek out the bidder with the highest valuation, and thus maximize the efficiency of the network (Mierendorff, 2013).

Next, let's briefly touch upon the technological efficiency of the re-allocation mechanism. The operation of dynamic re-allocation mechanism requires both the constant availability of all resources on the network and the operation of the dynamic SFC model suggested in this paper (Csoma et al., 2014). However, with the requirement of resource availability in the alternative static allocation model, due to the guarantee of bundles to each slice provided (Ordonez-Lucena et al., 2018), the re-allocation mechanism does not have any significant effect on cost in comparison. There is a cost in a virtualized resource being active, as opposed to idle, but this seems both intuitively small and is outside of the scope of this thesis. On the cost of the SFC model used, the authors of the model have based it on an OpenSource language and it can be run on any generic hardware, and therefore we can assume that the extra cost for the sellers in our mechanism for operating this tool are likely outweighed by even a small increase in resource usage and thus revenue.

5.2. Ex-post MNO Network Composition

The problem with a naïve allocation of network slices to the network resources of all participating MNOs is the conflict of efficiency between the allocation and re-allocation phase: the allocation phase might cause the re-allocation to be less efficient *ex-post* than another division at the original allocation level. As an example, this could mean that one MNO's network slices all use their network services heavily from midnight to noon but have a low network service valuation for the other half of the day, and vice versa for the network slices of another MNO's network. *Ex-post*, we would want to split the network slices between these two networks to allow higher instantiation levels to be ran by all network slices. Also, a poor balance of fixed and dynamic demand network slices would mean that the customer satisfaction of the network slices would be lower, since some fixed demand slices would not have the ability to run on higher instantiation levels, while some dynamic demand slices would see their resources being constantly re-allocated away to cause a need for constant re-scaling of the network slice.

Since a solution like this would require existing knowledge of the market operation, it is a problem that goes beyond the scope of this thesis but would still be an interesting subject for further analysis. The difficulty of creating an algorithm to achieve this is a familiar one to this paper: the mismatch in efficiency between *ex-ante* and *ex-post* allocation. The critical question for the algorithm that would plan this division would be the expected total valuation a bidder will have for the combined allocation and re-allocation phase network slice they received. This would require knowledge of previous rounds to make accurate estimations. The alternative way of doing this would be heuristics regarding the estimated demand structure that each network slice would have during the re-allocation phase.

5.3. Net Neutrality

The vision of network slicing envisioned in this paper goes contrary in principle to the laws that concern net neutrality in the EU, although the technology behind it does not necessarily break these laws. The large concern is the way that prioritization and differentiation of network slices might account to discrimination under the net neutrality laws, causing a fear that the development of the technology might be hindered by the lack of regulatory clarity (Morris, 2018). The intent of the laws are both clear and desirable, but the ability for network slices to customize their own network and for our mechanism to divide up resources based on the valuations by bidders is a necessary component that makes network slicing efficient. In this section, I will look through how the current EU net neutrality laws conflict with both network slicing in general and my own mechanism.

The crucial differentiation that EU's regulation and the Body of European Regulators for Electronic Communications (BEREC) guidelines make is between internet access service (IAS) and specialized service providers (SpS): IAS providers offer users a connection to the distribution networks of the internet, while specialized service providers offer other services than IAS that is "optimized for specific content, applications or services" ("Open Internet and Net Neutrality," 2018). On one end of the scale, an internet service provider like Verizon would obviously be categorized as an IAS provider, while Spotify would clearly be SpS provider. However, the issue with network slicing is the customization that is possible under the technology, causing uncertainty in which category it fits under.

The BEREC guidelines note both that specialized services cannot be used or offered as a "replacement for IAS" (Body of European Regulators for Electronic Communications, 2016, para. 102) or used to circumvent the stricter rules on IAS as a "potential substitute for the IAS" (Body of European Regulators for Electronic Communications, 2016, para. 126). For the purpose of network slicing, the key component here is that slices cannot act for the sole purpose of internet access, but rather be a vehicle of internet access through some application. An example of this would be video streaming: the user is in end effect accessing the internet, but only through a specialized service that seeks to optimize the quality of your access. However, the regulations refer that SpS provider should provide services "other than internet access services" (EU/EC, 2015, sec. 3(5)) and the guidelines that specialized services "do not provide connectivity to the Internet" (Body of European Regulators for Electronic Communications, 2016, para. 110). These two would then imply that the opposite is true, and that network slices shouldn't be allowed to serve as an access to the internet. Although it seems likely these contradictions are only intended to deny the use of SpS to avoid the more stringent regulation on IAS and to stop SpS to be used as a proxy for direct internet access, this is not yet clear in the text. It is unlikely that any mobile network operator (MNO) would allow for any bidder to provide a direct IAS under its own physical network due to its ability to cannibalize their own IAS business, while allowing this competitor to pick and choose its customer base due to the "as-a-Service" -nature of network slicing (Kotakorpi, 2006; Vogelsang, 2017). Therefore, any network slice is unlikely to evolve into this area of service, meaning that the regulatory framework only serves to threaten the innovation that can occur under specialized services. This section of the regulation would need a resolution that lifts the uncertainty from businesses that provide an indirect access to the internet through their network slice.

The second thing we will look at is the Quality of Service (QoS) differentiation that is at the heart of the network slicing value proposition, what the EU regulation calls traffic management. The regulation allows for differentiation to take place on the Internet, in a limited capacity: *"...such measures shall be transparent, non-discriminatory and proportionate, and shall not be based on commercial considerations but on objectively different technical quality of service requirements of specific categories of traffic. Such measures shall not monitor the specific content and shall not be maintained for longer than necessary."* (EU/EC, 2015, para. 9) Applied to network slicing, we can clearly see that different use cases like VR video content, smart grid solutions and automated cars fit into having "objectively different" QoS demands, and thus discrimination between them would be within the language of the regulation. Also, the "transparent and non-discriminatory" operation of traffic management can be controlled through 5G QoS indicators (5QI), which look at the "QoS forwarding treatment" for the QoS flow in areas like scheduling weights and admission thresholds (3GPP, 2018). 5QIs allow for network slices and the MNOs to prove to regulators that traffic is not handled differently from the valuations of the system. This requires extra costs to the MNO and network slices, but helps show compliance to the regulations and thus lessens the uncertainty.

The more difficult situation arises from understanding the language regarding “commercial considerations” not being the basis of decisions, and the differentiation between network slices underneath the same use case, e.g. Netflix and HBO. Under our model in the creation phase, firms value the impact of a network slice to their business as either the demand of its customer and their willingness-to-pay for the service, or the additional projects or savings in existing projects (see equation 1). Let us assume that Netflix and HBO have exactly the same network request in terms of what resources their network slices need. However, the value Netflix customers gain from the service is higher than HBO customers, and thus Netflix’s valuation is higher. This could be modelled as Netflix customers accepting a higher subscription price than HBO customers for the service under a network slice.

Now, let us assume that both network slices are seeking more C-RAN resources to handle a high customer load and to guarantee latency rates, a key Quality of Service measurement for both service providers. All other parameters the same in valuation in the re-allocation phase, Netflix would naturally offer more for these resources than HBO. This could cause a situation where only Netflix gets these resources in the re-allocation, causing the QoS of Netflix to be higher than HBO. This would be done partially according to “commercial interests”, in that the resource providers would gain more from allocating the resources to Netflix rather than HBO. However, it would be transparent in terms of Netflix winning based on observable, clear reasons. Also, this allocation would provide the greater net social welfare, since Netflix consumers benefit more from the same resources. An equal level of QoS is therefore unlikely to be achieved, as well as being an inefficient use of resources. The regulatory language would thus need to be clarified on what sort of differentiation is allowed.

(Frias & Pérez Martínez, 2018) looked into the disparity between QoS and the guaranteeing of fair competition between “equivalent services of a certain vertical industries” with specific QoS that make them specialized services, noting that this requires the categorization of the application-layer services. However, the reality is that most of the network slices will seek differentiation from each other, even if they are in the same industry. Frias and Pérez Martínez agree on this, noting that such services “are in constant evolution and become substitutive in a very dynamic way" (Frias & Pérez Martínez, 2018). Thus, there's a difficulty in trying to define what the fair competition should look like. The value of the mechanism described in this thesis is its ability to distinguish the reason for QoS difference between different network slices: those that get the resources are the most efficient users of the resources given the weakly dominant strategy of truthful bidding. However, the aversion against paid prioritisation by the current net neutrality regulation might mean that such efficient allocation would be seen as discriminating on traffic on "commercial interests", and thus illegal. The harm of such aversion to efficient allocation is explained well by (Frias & Pérez Martínez, 2018): "the potential of network virtualisation… cannot be unleashed under a regulatory vision that presumes that best-effort is always best". Given the customization of 5G networks, the best-effort internet is not the best option available, and thus requires amendment to be suitable for a future NSaaS world.

5.4. Herding and Social Learning

An issue with a market mechanism that calls upon principals and agents to provide private valuations for the sold resources, especially the bidders that seek to value the resources on offer, is the possibility of conforming towards the price signals that other agents provide. If the bidders in our mechanism have uncertainty in their own valuation of the value of a network service, they might be willing to forgo their own private information for the noisy signals that other bidders provide with their allocation and re-allocation requests. Especially in a fear of having underestimated the value of a network service combination, bidders might end up bidding inefficiently for the available network services, causing an *information cascade*, where bidders end up making decisions independently of their private information. If such a situation arises, the decisions made are not based on any valuable information from a succession of bids, and therefore there is no social learning. (Vives, 2008, p. 200).

The value of our re-allocation mechanism is premised on the network slices learning about the value of each bundle and the resources within them. We cannot expect the first allocation round to provide a fully efficient market, simply because the information on bundle value is expected, not realised. With each round of allocation and re-allocation, we are able to learn more about what is valuable: the orchestrator can see the change in valuation for bundles with different geographical, resource-balance elements. This can be seen as one actor in the market finding out previously hidden information, thus changing their private information. If, for example, they found out that the geographical grouping of physical radioheads (as seen in figure 13) would be more efficient by organizing them along user traffic hotspots like roads, so moving from the red clusters to the green clusters, this information would allow them to garner more value out of the bundles allocated to them. This information being available to the orchestrator through a change in bidding habits would also work as a signal to make the network service packaging more efficient in the next allocation round, thus generating more value for the entire market. Moreover, changes in the demand on the network slice side can also cause changes in the way network services are bid for. With the end user of the network slice better understanding the capabilities of the network, their valuation can evolve enough to push the bidder to change their demands to the orchestrator for a different way of packaging bundles to be required. We thus see that social learning is a valuable tool in the long-term for making the market mechanism better.

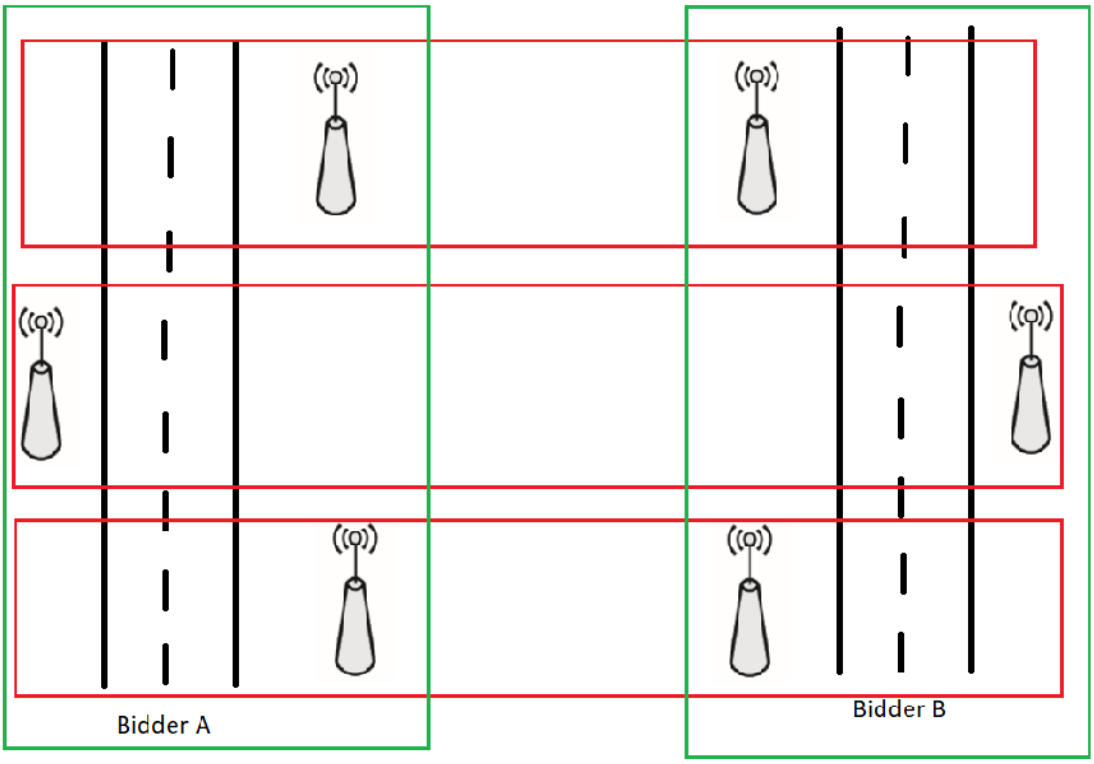


Figure 13: Example of more efficient radiohead network service packaging

How about a network slice that is the first slice to learn the value of a specific combination of resources, or an innovation in efficiency brought about by combining some network services? This agent now has private information that it will wish to keep private, while reaping the benefit of this information. It will thus want to win the bids for the network services it seeks in both the allocation and re-allocation mechanisms, without other bidders understanding the true value of these bundles. Let's suppose that bidder A has a new valuation of $2x$ for the bundle B_1, when combined with two other bundles, B_2 and B_3. Meanwhile, let us assume that the expected value that other bidders hold for this combination is $E(\pi)\sim N(0, x)$, but the realized value with this combination is $\pi\sim N(0, 3x)$. Thus, under the GSP mechanism, bidder A knows that in the first round, any bid above x is going to place it at the top of the priority list. Since the information rent from this knowledge will come from continued undervaluing by other bidders, bidder A has an incentive to hide the private information they have.

Let us look at how the discrepancy of private information can make the bidder make strategic bids for the resources. The signal bidder A is concerned about now is the size of the bid they provide, thus the signalled valuation on the bundle. We can model the value this information has to other bidders as a value from a normal distribution curve: most information will provide value to a part of the bidders. Bidder A doesn't know the value of this information to others but fears higher competition for the network service combination she seeks. This is especially pronounced, if the value generated is highly dependent on one specific resource, such as a location-specific edge computing service. Using the traffic offloading as-a-Service (TOFaaS) example in figure 13, bidder A has private information that the green combination on the left is a superior option to any of the red options available. When this combination is available during the allocation and re-allocation phases, it will seek to win the bid, and given it has a significantly higher valuation for the key network service, this would be the efficient option. The orchestrator will use this growing demand to redesign the network services, packaging more of them in the green combination, since this is more efficient. However, bidder A fears that this signalling by the orchestrator, who can view all the submitted bids and seller minimum prices, gives information to the other bidders to change their behaviour in the same way. This then makes the bids for the green combination more competitive, reducing the surplus bidder A garners from their private information. All this might change bidder A's *ex-ante* incentives to provide a truthful valuation to the orchestrator.

However, since the past prices are a consistent estimator of the true valuation, prices will eventually converge through increased knowledge by other bidders. This information is based both on understanding of other bids and of the realized value of a network composition. The speed of this convergence is dependent on the amount of informed bidders on the market

(Vives, 2008, p. 267). The key mechanism that the orchestrator can use to increase convergence in the market is anonymized bid information for each round. This allows for bidders to place more weight on the public information from past bids and less on their noisy private valuation, thus utilizing the private information of bidder A to understand the true value of the combination (Vives, 2008, p. 339). To counter this, bidder A consider information leakage in its bids: they will seek to bid when the market is deep, i.e. when noisy bids don't change the prices too much and avoid bidding when it's not. However, the trade-off between hiding and bidding at desired moments is going to cause the private information to gradually seep into the price (Kyle, 1985).

5.5. Analysing bidder types

We briefly looked into the types of bidders that would take part in our market mechanism in the creation phase, section 4.2, but we had to make some assumptions and simplifications regarding the preferences of bidders to make a functional mechanism. Given the market mechanism's semi-permanent design in allocation and re-allocation, and the necessity for the rules of the mechanism to be the same for all, there are bound to be bidders that are not ideally served by the design. The fear in creating different levels of rules and allocation timeframes is that it incentivizes bidders to strategically identify their types in order to classify for different rules. In this section, we'll look deeper into the different alternative bidder types likely seeking network slices from our market, and how we can try to serve their needs.

Firstly, we need to look into network slice seekers with very periodic valuation, i.e. bidders with a need for a network slice on occasions only. These sorts of bidders would want to have a network slice to use at times, while not needing it at other times, thus not being suitable for the timeframe designed for the allocation phase. The benefit of our market mechanism does come from the flexibility of the creation phase during discovery, however: the bidder is able to create itself a network slice structure during the allocation phase that does not have any resources tied to it. Rather, they have a network topology created for them, i.e. what network, computation and storage resources it would use, and how it would use them, if it was allocated some resources. The cost of this would therefore only be the costs to design this network topology to the mobile network operator. Thus, this group of bidders is able to take part in the re-allocation phase of the market mechanism, whenever it needs resources for its network slice. It can therefore take part in the market for one week every month, for example, although it would not have resources guaranteed to it under this system. This is the main drawback of our system for this type of bidder, although accommodating them in any other way would compromise the optimization the orchestrator seeks.

Next, let's look at bidders with a one-off desire for a network service, e.g. the telecommunications network created for the Pyeongchang Olympics in 2018. This type of bidder does not have any lock-in effects to worry about, the cost of switching that hinders changing suppliers due to unexpected price changes, since it is using the network for such a limited amount of time (Farrell & Klemperer, 2007). However, as with the earlier bidder type, the timeframe of the allocation phase is too long. The option of creating a network topology that they could utilize during a set of re-allocation phases exists for them, but the lack of continued usage would make this less viable. The alternative option within our network would be to sublet a network slice from an entitled network slice, but this network slice would likely be sub-optimally designed for them. The more realistic option is for them to go directly to an MNO to create a private network slice, since this would allow for the contractual flexibility that they need, and our mechanism cannot provide.

Finally and most importantly, let's look at bidders with a much longer timeframe they require the network slice to be guaranteed than it is in our mechanism. An example of this would be a company building a factory and seeking a network slice to connect all its devices within the factory together, and to connect the entire factory to its larger network. They would seek assurances from the orchestrator for a much longer timeframe than the market mechanism can accommodate for, such as 20-30 years. If they cannot get guarantees of a minimal instantiation level to lock in their weighted average cost of capital (WACC), the risk to cashflow of a factory without the required network slice causing the entire factory not to be worth it. The disruption to their business from not having an operative network slice can be inhibitive. Specifically, in the case of cloud computing as we have here, the problem of vendor lock-in can be severe due to the difficulty in migration from one network to another of their applications and data, with possibly incompatible software or architectures hindering movement (Opara-Martins, Sahandi, & Tian, 2014).

This is the most difficult bidder type to accommodate within our market mechanism: a bidder seeking assurances for a network slice that might not be optimal in the long-term. The solution within the mechanism would come from the valuation of the factory being so high that they would gain a network slice throughout the 20-30 years, each time there's a new allocation phase. Thus, the end outcome would be the same as a guarantee, although there would be *ex-ante* risk for the bidder that they would have to factor into their profitability calculations. This would be likelier to work in later allocation rounds, once the approximate value of the network slices has risen, mitigating this bidder's risk. Given a bidder with a similar timeframe demand, but with a marginal valuation within the allocation phase, the risk involved for them would likely be so high that the investment would not be worth it without guarantees. Such a bidder would go against the *ex-ante* efficiency rule for the orchestrator (Mierendorff, 2013), with a higher valuation bidder likely to arrive for the following allocation phases to take the place of the factory bidder, thus making the guarantees a sub-optimal option. Realistically, some of this type of bidder is likely to exit our mechanism entirely, seeking their own network slice with dedicated resources, separate from our market, rather than bidding for this network slice as-a-Service. However, this type of bidder cannot be fully accommodated in our system without creating knock-on effects to other bidders.

5.6. Market Information Structure

A critical question that deserves further pondering is the amount of information other bidders should have of each other's bids. On one extreme of the spectrum, the orchestrator organizes the mechanism as a black box, where bidders can only share their own valuations, and the only information is their own actions, and the collective structure of the market. On the other end, bidders would get information after each round on what each bidder received, and how much they paid for this. The trade-off from information in a market as nascent as ours is the effect that it has on bidders' valuations: new information may help bidders adjust their valuations to be more accurate, but it also disincentivizes truthful bidding for those with more certain private information. In this section, we'll look at what this balance of information should be in our mechanism, reflecting on existing literature.

During the allocation phase of the mechanism, bidders bid for a network slice designed for them, without information on what other bidders are bidding for and what they are offering for it. By design, this limits the bidder's options within the market to the alternatives the orchestrator designs in section 4.3.1.2, outside of collusion between bidders. For the orchestrator, awareness on all other bids would allow the bidder to practice bid shading, reducing their offer from their true valuation (Krishna, 2010a). In practice, this could happen in the form of a bidder becoming aware about the lack of demand for one network service that their network slice requires. Building on the assumption that the orchestrator designs the optimal network slice and alternative, given the bidder's QoS requirements and the total resource demand, we assume that awareness of the network slice structure for other bidders would not make the market more efficient, and thus would have no value for the orchestrator to provide. The phase is therefore best kept as a game of incomplete information, with bidders reliant on defining a strategy as a best-response to a Bayesian game against other bidders. This allows us to rely on the literature of Bayes-Nash incentive compatible games and the strategy structures they offer (e.g. Myerson, Maskin and Riley).

Besides understanding the bids of other bidders, another key area of information concerns the acquisition of information by an uninformed bidder to make their private valuation more accurate. One way to model this is for the bidder to have access to statistical experiments that would yield further information on their private valuation from a set of probability distribution. Given that the orchestrator seeks to create a socially optimal mechanism, it is easy to see why the bidders have an incentive to acquire information in a socially optimal manner (Bergemann & Välimäki, 2002). Effectively, this would mean a mechanism with a costly entry, where bidders gain a private signal about their network slice upon market entry (Mcafee & Mcmillan, 1987). If used in our mechanism, this cost would not be required for all bidders, but would rather be something an uninformed bidder might use. Key here is that the orchestrator itself is not parcelling out information regarding the true value of a network slice. Firstly, the orchestrator is likely to have only limited information in the first round of allocation in the mechanism. Secondly, any sort of information parcelling by the orchestrator would risk mitigating the benefits of a neutral actor, with an entry barrier limiting the entrance to the market of smaller bidders.

As mentioned in section 4.4.4, our re-allocation mechanism is best modelled as a full information game, with a market that becomes more predictable in its prices over time due to fairly static bundle composition and predictable total demand. Since the re-allocation mechanism would be repeated often, the game quickly transitions towards one with full information. The differentiation that must be made is between information of what was sold for what price, and who won which bundles. The former is information that would be impossible but also impractical to hide from bidders, since the orchestrator wants to generate social learning between bidders to minimize the risk they factor into their valuations (Vives, 2008, p. 200). The bidders want to learn about the true value of the bundles on offer, but are reliant on information signals it gains from the market, such as the value of winning bids at certain points in time. On the other side of the coin, the orchestrator has limited ability to observe the signals that it provides to bidders and the value this information has to them. The orchestrator controls the information on the market, but has limited ability to learn the value of the information, thus creating the trade-off of information that I'm talking about (Esö & Szentes, 2007). There is a greater issue with free-riding when bidders have awareness of the winning bids within the market. Since bidders can strategically bid low, and still garner the information within the market, the need for collective action in figuring out the true value of these bundles is lost. If the "social marginal value of evidence" is pushed higher by forcing bidders to bid truthfully to understand the marginal winning bid for a bundle, i.e. at what price are they just losing, this information goes to those that need it the most. Thus, the market is made more accurate for those bidders that put effort into information collection, mitigating the free-riding problem (Li, 2001).

The latter option, of bidders learning the identity of winning bidders would create a second trade-off: this knowledge would vastly improve other bidders' private information by making information signals less fuzzy, but it would also further disincentivize presenting this information, as shown in section 5.4. Given that the information structure in this game is dynamic, and that bidders are likely to learn new private information at different rates, the revelation of their identity would inhibit them from the acquisition of information. The mechanism should seek to reward truthful information with information rent, in order to induce a socially efficient system. Even if the *ex-post* optimal solution would be for all bidders to learn from their peers for the most efficient bidders to always win, the *ex-ante* considerations of information acquisition incentives makes it untenable in the long-term. This is also the way in which Google operates its generalized second-price advertisement auctions, with bidders got given information on others' identities, only estimated slot positions and the average price they should expect to pay (Edelman et al., 2005, p. 6).

6. Conclusion

To reiterate the core purpose of this thesis, I add to existing literature a dynamic allocation mechanism for network slicing that allocates required resources based on the valuation that bidders have for these resources, and thus by extension the end user welfare created through these resources. Far from intended to be an exhaustive exploration, this work creates a new framework that a future dynamic allocation mechanism can draw from. To create this comprehensive mechanism, I have drawn on technical works in computer science as a basis for a solution that resolves incentive problems through existing economic literature and modified economic theory to suit the technology behind network slicing. The problem faced in the network slicing market is one of vastly different requirements for networks by bidders that have private valuations for their desired network slice functionalities and an uncertainty for the orchestrator of this market on what to offer. The allocation therefore benefits from having incentives for truth-revealing in an uncertain environment, limiting the ability for a menu-based offering with pre-set prices. Rather, the bidding mechanism should make a truthful valuation incentive compatible.

My designed mechanism starts off with the collection of bidders and sellers into one marketplace run by a neutral orchestrator. This orchestrator then takes the requests bidders have for their network slice to create individual, prospective network slices based on the resources offered by the sellers. Since bidders have such wide-ranging demands for their own network slice, any mechanism design that would require bidders to bid for the same packages would clash in a trade-off between efficiency and tractability. Also, this way allows us to create an allocation mechanism built upon payment through the Clarke pivot rule, guaranteeing truthful valuations by the bidders. Finally, we also add a second-chance bid extension to the mechanism to avoid individual sought-after resources causing bottlenecks that would limit the *ex-post* allocation efficiency.

Rather than measure efficiency on the maximization of service guarantees, I have extended the existing literature by maximizing the total valuation of the granted network slices. Thus, the orchestrator's role doubles from just maximizing the utilization of resources to maximizing the value created by them. Since bidders can create vastly different levels of utility from the same resources, it makes sense for the orchestrator to discriminate in favour of the value-adding network slices. I further extend this model from one-off static allocation of resources to a dynamic allocation mechanism that seeks to use the same value maximization principle to resources that are unused in the network. Rather than guaranteeing a minimum-maximum range of resources to network slices, the mechanism guarantees their desired minimum level Quality of Service, with extra resources bid for through the dynamic re-allocation phase. This guarantees bidders what their network slice needs, while allowing bidders to seek more resources when they need it. This seeks to maximize the utility created within the entire network by minimizing the chance of a network slice that would create high utility from its services not being allocated the necessary resources.

Bibliography

3GPP. (2018). *ETSI Release 15: System Architecture for the 5G System* (Vol. 0).

Abdelwahab, S., Hamdaoui, B., Guizani, M., & Znati, T. (2016). Replisom: Disciplined Tiny Memory Replication for Massive IoT Devices in LTE Edge Cloud. *IEEE Internet of Things Journal*, *3*(3), 327–338. https://doi.org/10.1109/JIOT.2015.2497263

Afèche, P. (2006). Ch. 2:Economics of Data Communications. In T. Henderschott (Ed.), *Economics and Industrial Systems* (pp. 53–135).

Akyildiz, I. F., Lin, S., & Wang, P. (2015). Wireless software-defined networks (W-SDNs) and network function virtualization (NFV) for 5G cellular systems : An overview and qualitative evaluation. *Computer Networks*, *93*, 66–79. https://doi.org/10.1016/j.comnet.2015.10.013

Alexiadis, P., & Shortall, T. (2017). The Advent of 5G : Should Technological Evolution Lead to Regulatory Revolution ?, 1–13.

Ausubel, L. (2004). An Efficient Ascending-Bid Auction for Multiple Objects. *American Economic Review*, *94*(5), 1452–1475.

Ausubel, L. M., & Milgrom, P. (2005). *The Lovely but Lonely Vickrey Auction*. *Combinatorial Auctions*. https://doi.org/10.7551/mitpress/9780262033428.003.0002

Barbarossa, S., Sardellitti, S., & Di Lorenzo, P. (2014). Communicating while computing: Distributed mobile cloud computing over 5G heterogeneous networks. *IEEE Signal Processing Magazine*, *31*(6), 45–55. https://doi.org/10.1109/MSP.2014.2334709

Bartal, Y., Gonen, R., & Nisan, N. (2003). Incentive compatible multi unit combinatorial auctions. *Proceedings of the 9th Conference on Theoretical Aspects of Rationality and Knowledge - TARK '03*, 72. https://doi.org/10.1145/846247.846250

Bergemann, D., & Välimäki, J. (2002). Information in Mechanism Design. *Econometrica*, *70*(3), 1007–1033.

Bhamare, D., Erbad, A., Jain, R., Zolanvari, M., & Samaka, M. (2018). Efficient virtual network function placement strategies for Cloud Radio Access Networks. *Computer Communications*, *127*(April), 50–60. https://doi.org/10.1016/j.comcom.2018.05.004

Bhushan, N., Li, J., Malladi, D., Gilmore, R., Brenner, D., Damnjanovic, A., & Teja, R. (2014). Network Densification : The Dominant Theme for Wireless Evolution into 5G, (February), 82–89.

Body of European Regulators for Electronic Communications. (2016). BEREC Guidelines on the Implementation by National Regulators of European Net Neutrality Rules, (August), 45. Retrieved from http://berec.europa.eu/eng/document_register/subject_matter/berec/download/0/6160-berec-guidelines-on-the-implementation-b_0.pdf

Bourreau, M., Cambini, C., & Doğan, P. (2012). Access pricing, competition, and incentives to migrate from old to new technology. *International Journal of Industrial Organization*, *30*(6), 713–723. https://doi.org/10.1016/j.ijindorg.2012.08.007

Bourreau, M., Cambini, C., & Hoernig, S. (2012). Ex-ante regulation and co-investment in the transition to next generation access. *Telecommunications Policy*, *36*(5), 399–406.

Brito, D., Pereira, P., & Vareda, J. (2010). Can two-part tariffs promote efficient investment on next generation networks? *International Journal of Industrial Organization*, *28*(3), 323–333. https://doi.org/10.1016/j.ijindorg.2009.10.004

Brito, D., & Tselekounis, M. (2016). Access regulation and the entrant's mode of entry under multi-product competition in telecoms. *Information Economics and Policy*, *37*, 20–33. https://doi.org/10.1016/j.infoecopol.2016.10.005

Caballero, P., Banchs, A., Veciana, G. De, Costa-Pérez, X., & Azcorra, A. (2018). Network Slicing for Guaranteed Rate Services : Admission Control and Resource Allocation Games, *17*(10), 6419–6432.

Caragiannis, I., & Kaklamanis, C. (2011). On the efficiency of equilibria in generalized second price auctions. *... ACM Conference on ...*, 1–41. https://doi.org/10.1145/1993574.1993588

Cave, M. (2018). How disruptive is 5G? *Telecommunications Policy*, *42*(8), 653–658. https://doi.org/10.1016/j.telpol.2018.05.005

Chatras, B., Tsang Kwong, S., & Bihannic, N. (2017). NFV Enabling Network Slicing for 5G, 219–225.

Choyi, V. K., Abdel-hamid, A., Shah, Y., Ferdi, S., & Brusilovsky, A. (2016). Network Slice Selection , Assignment and Routing within 5G Networks, 1–7.

Cisco. (2017). Full-Text, 2016–2021. https://doi.org/1465272001663118

Clarke, E. H. (1971). Multipart Pricing of Public Goods. *Public Choice*, *11*(1), 17–33.

Csoma, A., Sonkoly, B., Csikor, L., Németh, F., Gulyás, A., Tavernier, W., & Sahhaf, S. (2014). ESCAPE: Extensible Service ChAin Prototyping Environment using Mininet, Click, NETCONF and POX. *Proceedings of the 2014 ACM Conference on SIGCOMM - Demo*, 125--126. https://doi.org/10.1145/2619239.2631448

Dimitri, N. (2018). Combinatorial advertising internet auctions. *Electronic Commerce Research and Applications*, *32*(October), 49–56. https://doi.org/10.1016/j.elerap.2018.10.005

Dorsch, N., Kurtz, F., & Wietfeld, C. (2018). On the economic benefits of software-defined networking and network slicing for smart grid communications. *NETNOMICS: Economic Research and Electronic Networking*, pp. 0–3. https://doi.org/10.1007/s11066-018-9124-3

Economides, N., & White, L. J. (1995). Access and Interconnection Pricing: How Efficient is the "Efficient Component Pricing Rule"? *Ssrn*, *508*(1994). https://doi.org/10.2139/ssrn.15114

Edelman, B., Ostrovsky, M., & Schwarz, M. (2005). Internet advertising and the GSP auction: selling billions of dollars worth of keywords. *American Economic Review*, *02138*, 1–25.

Edelman, B., & Schwarz, M. (2010). Optimal Auction Design and Equilibrium Selection in Sponsored Search Auctions. *American Economic Association*, *100*(2).

Esö, P., & Szentes, B. (2007). Optimal Information Disclosure in Auctions and the Handicap

Auction. *The Review of Economics Studies*, *74*(3), 705–731.

Esswie, A. A., & Pedersen, K. I. (2018). Opportunistic Spatial Preemptive Scheduling for URLLC and eMBB Coexistence in Multi-User 5G Networks. *IEEE Access*, *6*, 38451–38463. https://doi.org/10.1109/ACCESS.2018.2854292

ETSI GS NFV. (2013). NFV-Architectural Framework. https://doi.org/DGS/NFV-0011

EU/EC. REGULATION (EU) 2015/2120, Official Journal of the European Union § (2015).

Farrell, J., & Klemperer, P. (2007). Coordination and Lock-In: Competition with Switching Costs and Network Effects. *Handbook of Industrial Organization*, (3).

Foros, O. (2004). Strategic Investments with Spillovers , Vertical Integration and Foreclosure in the Broadband Access Market. *International Journal of Industrial Organization*, *22*(1), 1–24.

Foukas, X., Patounas, G., Elmokashfi, A., & Marina, M. K. (2017). Network Slicing in 5G: Survey and Challenges. *IEEE Communications Magazine*. https://doi.org/10.1109/MCOM.2017.1600951

Frias, Z., & Pérez Martínez, J. (2018). 5G networks: Will technology and policy collide? *Telecommunications Policy*, *42*(8), 612–621. https://doi.org/10.1016/j.telpol.2017.06.003

Halpern, J., & Pignataro, C. Service Function Chaining (SFC) Architecture, Internet Engineering Task Force (IETF) (2015). https://doi.org/10.1017/CBO9781107415324.004

Han, B., Gopalakrishnan, V., Ji, L., & Lee, S. (2015). Network function virtualization: Challenges and opportunities for innovations. *IEEE Communications Magazine*, *53*(2), 90–97. https://doi.org/10.1109/MCOM.2015.7045396

Houngbonon, G. V., & Jeanjean, F. (2016). What level of competition intensity maximises investment in the wireless industry? *Telecommunications Policy*, *40*(8), 774–790. https://doi.org/10.1016/j.telpol.2016.04.001

Hsieh, H., Chen, J., & Benslimane, A. (2018). Journal of Network and Computer Applications 5G Virtualized Multi-access Edge Computing Platform for IoT Applications. *Journal of Network and Computer Applications*, *115*(May), 94–102. https://doi.org/10.1016/j.jnca.2018.05.001

Intel. (2013). Smart Cells Revolutionize Service Delivery, 1–7. Retrieved from https://www.intel.com/content/www/us/en/communications/smart-cells-revolutionize-service-delivery.html

ITU-T. (2008). E.800: Definitions of terms related to quality of service. *ITU-T Recommendation*, 1–30. Retrieved from http://www.itu.int/rec/dologin_pub.asp?lang=e&id=T-REC-E.800-200809-I!!PDF-E&type=items

Jeanjean, F., & Houngbonon, G. V. (2017). Market structure and investment in the mobile industry. *Information Economics and Policy*, *38*, 12–22. https://doi.org/10.1016/j.infoecopol.2016.12.002

Jiang, M., Condoluci, M., Mahmoodi, T., & Guijarro, L. (2017). *Economics of 5G Network*

Slicing: Optimal and Revenue-based allocation of radio and core resources in 5G. Retrieved from https://nms.kcl.ac.uk/toktam.mahmoodi/files/TWC-16.pdf

Keene, I., Fabre, S., & Takiishi, K. (2019). Market Guide for 5G New Radio Infrastructure.

Kelly, F. P., Maulloo, A. K., & Tan, D. K. H. (1998). Rate control for communication networks, 1–16. Retrieved from papers2://publication/uuid/06B5BEDE-C730-47A6-A8B8-259338FBC976

Kotakorpi, K. (2006). Access Price Regulation, Investment and Entry in Telecommunications. *International Journal of Industrial Organization*, *24*(5), 1013–1020.

Krishna, V. (2010a). Equilibrium and Efficiency with Private Values. In *Auction Theory* (pp. 185–202). https://doi.org/10.1016/B978-0-12-374507-1.00013-3

Krishna, V. (2010b). Nonidentical Objects. In *Auction Theory* (pp. 227–238). https://doi.org/10.1016/B978-0-12-374507-1.00006-6

Ksentini, A., & Nikaein, N. (2017). Towards enforcing Network Slicing on RAN : Flexibility and Resources abstraction. *IEEE Communications Magazine*, *55*(6), 102–108.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, *53*(6), 1315–1335.

Laffont, J.-J., & Tirole, J. (1991). Auction Design and Favoritism. *International Journal of Industrial Organization*, *9*(1), 9–42.

Lehmann, D., Oćallaghan, L. I., & Shohman, Y. (2002). Truth Revelation in Approximately Efficient Combinatorial Auctions. *Journal of the ACM*, *49*(5), 577–602.

Leivadeas, A., Falkner, M., Lambadaris, I., & Kesidis, G. (2017). Computer Standards & Interfaces Optimal virtualized network function allocation for an SDN enabled cloud. *Computer Standards & Interfaces*, *54*(January), 266–278. https://doi.org/10.1016/j.csi.2017.01.001

Lemstra, W. (2018). Leadership with 5G in Europe: Two contrasting images of the future, with policy and regulatory implications. *Telecommunications Policy*, *42*(8), 587–611. https://doi.org/10.1016/j.telpol.2018.02.003

Lestage, R., & Flacher, D. (2014). Infrastructure investment and optimal access regulation in the different stages of telecommunications market liberalization. *Telecommunications Policy*, *38*(7), 569–579. https://doi.org/10.1016/j.telpol.2014.01.003

Li, H. (2001). A Theory of Conservatism. *Journal of Political Economy*, *109*(3), 617–636.

Lucier, B., Paes Leme, R., & Tardos, E. (2012). On revenue in the generalized second price auction. *Proceedings of the 21st International Conference on World Wide Web - WWW '12*, 361. https://doi.org/10.1136/jcp.42.10.1092

Mach, P., & Becvar, Z. (2017). Mobile Edge Computing: A Survey on Architecture and Computation Offloading. *IEEE Communications Surveys and Tutorials*, *19*(3), 1628–1656. https://doi.org/10.1109/COMST.2017.2682318

Marsch, P., Silva, I. Da, Bulakci, Ö., Tesanovic, M., Eddine, S., Ayoubi, E., … Boldi, M. (2016). 5G Radio Access Network Architecture : Design Guidelines and Key Considerations. *IEEE Communications Magazine*, *54*(11), 24–32.

Marsden, R., & Ihle, H. (2018). Mechanisms to incentivise shared-use of spectrum. *Telecommunications Policy*, *42*(4), 315–322. https://doi.org/10.1016/j.telpol.2017.07.001

Maskin, E., & Riley, J. (1989). Optimal Mulit-unit Auctions. In *The Economics of Missing Markets, Information, and Games* (pp. 312–335). Oxford University Press.

Mcafee, R. P., & Mcmillan, J. (1987). Auctions and Bidding. *Journal of Economic Literature*, *25*(2), 699–738.

Medhat, A. M., Taleb, T., Elmangoush, A., Carella, G. A., Covaci, S., & Magedanz, T. (2017). Service Function Chaining in Next Generation Networks: State of the Art and Research Challenges. *IEEE Communications Magazine*, *55*(2), 216–223.

Mierendorff, K. (2013). Games and Economic Behavior The Dynamic Vickrey Auction. *Games and Economic Behavior*, *82*, 192–204. https://doi.org/10.1016/j.geb.2013.07.004

Mijumbi, R., Serrat, J., Gorricho, J., Bouten, N., Turck, F. De, & Boutaba, R. (2015). Network Function Virtualization : State-of-the-art and Research Challenges. *IEEE Communications Surveys & Tutorials*, *18*(1), 236–262.

Milgrom, P. (2000). Putting Auction Theory to Work : The Simultaneous Ascending Auction. *Journal of Political Economy*, *108*(2), 245–272.

Morgado, A., Huq, K. M. S., Mumtaz, S., & Rodriguez, J. (2018). A survey of 5G technologies: regulatory, standardization and industrial perspectives. *Digital Communications and Networks*, *4*(2), 87–97. https://doi.org/10.1016/j.dcan.2017.09.010

Morris, I. (2018). Ericsson CEO: Net Neutrality Threatens 5G. Retrieved January 10, 2019, from https://www.lightreading.com/net-neutrality/ericsson-ceo-net-neutrality-threatens-5g/d/d-id/740854

Myerson, R. B. (1981). Optimal Auction Design. *Mathematics of Operations Research*, *6*(1), 58–73.

Myerson, R. B. (1983). Mechanism design by an informed principal. *Econometrica*, *51*(6), 1767–1797.

Nakao, A., Du, P., Kiriha, Y., Granelli, F., Gebremariam, A. A., Taleb, T., & Bagaa. (2017). End-to-end Network Slicing for 5G Mobile Networks. *Journal of Information Processing*, *25*, 153–163. https://doi.org/10.2197/ipsjjip.25.153

Nisan, N., & Ronen, A. (2007). Computationally feasible VCG mechanisms. *Journal of Artificial Intelligence Research*, *29*, 19–47. https://doi.org/10.1613/jair.2046

Nunes, B. A. A., Mendonca, M., Nguyen, X., Obraczka, K., & Turletti, T. (2014). A Survey of Software-Defined Networking : Past , Present , and Future of Programmable Networks. *IEEE Commun. Surv. Tutor., 16 (3), 1617-1634.*, *16*(3), 1617–1634. https://doi.org/10.1109/SURV.2014.012214.00180>

Ollier, S., & Thomas, L. (2013). Ex post participation constraint in a principal-agent model with adverse selection and moral hazard. *Journal of Economic Theory*, *148*(6), 2383–2403. https://doi.org/10.1016/j.jet.2013.07.007

Opara-Martins, J., Sahandi, R., & Tian, F. (2014). Critical Review of Vendor Lock-in and Its Impact on Adoption of Cloud Computing. *International Conference on Information*

Society, 92–97.

Open Internet and Net Neutrality. (2018). Retrieved from https://ec.europa.eu/digital-single-market/en/open-internet-net-neutrality

Ordonez-Lucena, J., Adamuz-Hinojosa, O., Ameigeiras, P., Munoz, P., Ramos-Munoz, J. J., Chavarria, J. F., & Lopez, D. (2018). The Creation Phase in Network Slicing: From a Service Order to an Operative Network Slice. In *2018 European Conference on Networks and Communications, EuCNC 2018* (pp. 31–36). https://doi.org/10.1109/EuCNC.2018.8443255

Ostrovsky, M., & Schwarz, M. (2010). Reserve Prices in Internet Advertising Auctions: A Field Experiment. *EC*, (November). https://doi.org/10.2139/ssrn.1573947

Oughton, E. J., & Frias, Z. (2018). The cost, coverage and rollout implications of 5G infrastructure in Britain. *Telecommunications Policy*, *42*(8), 636–652. https://doi.org/10.1016/j.telpol.2017.07.009

Panzar, J. C. (1989). TECHNOLOGICAL DETERMINANTS OF FIRM AND INDUSTRY STRUCTURE. In R. Schmalensee & R. D. Willig (Eds.), *Handbook of Industrial Organization*.

Pedersen, K. I., Pocovi, G., Steiner, J., & Khosravirad, S. R. (2018). Punctured scheduling for critical low latency data on a shared channel with mobile broadband. In *IEEE Vehicular Technology Conference* (Vol. 2017–Septe, pp. 1–6). https://doi.org/10.1109/VTCFall.2017.8287951

Pianese, F., Gallo, M., Conte, A., & Perino, D. (2016). Orchestrating 5G Virtual Network Functions as a Modular Programmable Data Plane, 1305–1308.

Pindyck, R. S. (2005). *Pricing Capital under Mandatory Unbundling and Facilities Sharing. National Bureau of Economic Research*. https://doi.org/10.3386/w11225

Samdanis, K., Costa-Perez, X., & Sciancalepore, V. (2016). From Network Sharing to Multi-tenancy: The 5G Network Slice Broker. *IEEE Communications Magazine*, *54*(7), 32–39.

Sciancalepore, V., Cirillo, F., & Costa-Perez, X. (2017). Slice as a Service (SlaaS) Optimal IoT Slice Resources Orchestration. *IEEE Access*.

Tadayon, N., & Aissa, S. (2018). Radio Resource Allocation and Pricing: Auction-Based Design and Applications. *IEEE Transactions on Signal Processing*, *66*(20), 5240–5254. https://doi.org/10.1109/TSP.2018.2862398

Taleb, T., Ksentini, A., & Jäntti, R. (2016). Anything as a Service for 5G Mobile Systems. *Ieeexplore.Ieee.Org*, (December), 84–91. Retrieved from https://ieeexplore.ieee.org/abstract/document/7593429/

Taleb, T., Samdanis, K., Mada, B., Flinck, H., Dutta, S., & Sabella, D. (2017). On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration. *IEEE Communications Surveys and Tutorials*, *19*(3), 1657–1681. https://doi.org/10.1109/COMST.2017.2705720

Tang, L., Yang, H., Ma, R., Hu, L., Wang, W., & Chen, Q. (2018). Queue-aware dynamic placement of virtual network functions in 5G access network. *IEEE Access*, *6*, 44291–44305. https://doi.org/10.1109/ACCESS.2018.2862632

The Next Generation Mobile Networks Alliance. (2015). {5G} white paper. https://doi.org/10.1021/la100371w

Varian, H. (2009). Online Ad Auctions. *American Economic Review*, *99*(2), 430–434. Retrieved from https://www.jstor.org/stable/25592436

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, *16*(1), 8–37.

Vives, X. (2008). *Information and Learning in Markets*. *Princeton University Press*. Princeton University Press.

Vogelsang, I. (2017). The role of competition and regulation in stimulating innovation – Telecommunications. *Telecommunications Policy*, *41*(9), 802–812. https://doi.org/10.1016/j.telpol.2016.11.009

Zhang, G., Yang, K., Jiang, H., Lu, X., Xu, K., & Zhang, L. (2017). Equilibrium Price and Dynamic Virtual Resource Allocation for Wireless Network Virtualization. *Mobile Networks and Applications*, *22*(3), 564–576. https://doi.org/10.1007/s11036-016-0766-9

Zhou, X., Li, R., Chen, T., & Zhang, H. (2016). Network Slicing as a Service: Enabling Enterprises' Own Software-Defined Cellular Networks. *IEEE Communications Magazine*, *54*(7), 146–153.