MŰEGYETEM 1782

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Telecommunications and Media Informatics

Identification and modelling of traffic parameters for 5G application types

SCIENTIFIC STUDENTS' ASSOCIATIONS CONFERENCE PAPER

Author
Dániel Ficzere

Supervisors
Pál Varga, PhD
Gábor Soós

October 28, 2019

Contents

Kivonat

A 4. generációs hálózatok lassan már 10 éve elérhetőek a számunkra és jelenleg egy újabb nagy váltás küszöbén állunk. Az 5G hálózatok világszerte elindultak 2019-ben, egyelőre teszt üzemben. Itt az ideje az új, tömeges megjelenés előtt álló alkalmazások vizsgálatának, valamint megfelelő mérőszámokat kidolgozni ezen alkalmazások jellemzésére. A különböző szolgáltatások, hálózatok verifikációja folyamatosan fejlődött a korábbi évek során. A hálózat egyes részeinek önmagában való tesztelését felváltotották az end-to-end tesztelési módszerek. Továbbra is nehéz definiálni, hogy az adott szolgáltatás mennyire lesz megfelelő, mennyire elégíti ki a felhasználói igényeket. A machine-to-machine (m2m) felhasználások megjelenésével ez a kérdés pedig még összetettebbé vált, hiszen itt már közvetlen felhasználói elégedettség sem mérhető, ha a hálózat üzemel és nem keletkezett valamiféle hiba.

Az 5G hálózatok fő ígéretei a következők: 1 Gbps átviteli sebesség, 1 ms késleltetés és akár 1 millió $kliens/km^2$, 10-szeresen hatékonyabb energia-felhasználás, tömegesen csatlakozó berendezések. Ezen paraméterek együttes hatása a hálózatra jelenleg még nem ismert. A három fő alkalmazás-típus az 5G hálózatoknál az eMBB (enhanced Mobile Broadband), a cIoT (critical IoT) és az mIot (mass IoT) lesz, azonban egyelőre se a termékek között, se a készülékek között nem jelentek meg, mint elkülönülő irány. Elméletben számtalan új alkalmazást dolgoztak ki és mutattak be, azonban élő környezetben ezek többségét nem vizsgálták soha. Így nem ismert, hogy ezek közül valójában melyik lesz hosszú távon életképes.

Dolgozatomban mérések segítségével megvizsgálom a tervezett három fő alkalmazás-típusának - eMBB, cIoT, mIoT - forgalmi sajátosságait a jelenleg is elérhető mobilhálózatokon. Javaslatot teszek ezen alkalmazás-típusok forgalmi mintáinak leírására. Megfogalmazom, mit is jelent az előbb felsorolt alkalmazás-típusoknál a megfelelő szolgáltatás minőség, milyen igényeket kell kielégíteniük. A vizsgálati eredmények figyelembevételével forgalmi modelleket állítok fel a különféle alkalmazás-típusok forgalmának szimulációjához. A modelleket a segítségükkel generált mesterséges forgalmak és a valós forgalmi minták összehasonlításával validálom.

Mindezek ismeretében a már jelenleg is rendelkezésre álló egyik hazai teszt-5G hálózaton megvizsgálom, hogy milyen hatással lehet ezen forgalmi minták különféle eloszlású sokasága a hálózatra. Definiálok KPI-okat (Key Performace Indicator), melyekkel érdemes

ezen forgalmakat vizsgálni. Az eredmények alapján javaslatot teszek, milyen megfontolásokkal érdemes élni a mobilhálózati operátoroknak az 5G hálózati- és forgalomtervezési feladatai során ahhoz, hogy az újonnan megjelenő alkalmazások igényeit is ki tudják szolgálni a jövőben.

Abstract

4th generation mobile networks have been operating for almost 10 years now. Upcoming changes soon hit with 5G in the telecommunication industry. 5G networks are already operating all over the world in test mode. To analyze next-generation mobile networks properly, there is a need to define Key Performance Indicators (KPIs). The verification of networks and services have continuously evolved in the previous years. Testing signalling only or just partial domains of the network have been replaced with end-to-end testing methodologies. Nevertheless, it is still a tough task to define how adequate a particular service will be, and whether it can serve the user demands. With the appearing of machine-to-machine (m2m) applications, this question became even harder, since there is no direct user feedback. Quality of experience can not be measured accurately in m2m applications if the network operates correctly and without failures.

The upcoming promises of 5G include 1 Gbps data transmission, 1 ms latency, 1 million $devices/km^2$, and 10 times less energy consumption (compared to 4G UE). However, the combined effects of these attributes to the network are not fully known yet. At present, neither the networking services nor the User Equipment (UE) has been tuned for the three main use-cases, namely enhanced Mobile BroadBand (eMBB), critical Internet of Things (cIoT) and massive Internet of Things (mIoT). There are dozens of new - but theoretical - use-cases for 5G; however, these are not tested on a live network. It is hard to anticipate which ones will fulfil the expectations.

The first part of the paper examines the three application-types - eMBB, cIoT, and mIoT. Afterwards, I introduce the main traffic characteristics based on current mobile networks' traffic patterns and measurements. After that, I present some suggestions to describe and analyze the application-types and define what good enough service quality means for these use-cases. Considering the measurement results, I define traffic models for the simulation of different application-types. To validate these models, I compare the generated artificial traffic with real traffic patterns.

In the second part of the paper, I examine what the are the main effects of these traffic patterns on a domestic 5G test-network are. To characterize and describe these traffic types, I define appropriate KPIs. Finally, I make some considerations on the possible main impacts regarding 5G network design based on my measurement results.

Introduction

The rapid spread and popularity prove the success of cellular networks all around the world. Advanced mobile networks are critical components of the digital future. Boundaries between mobile and broader digital systems continue to blur. In this competitive scene, many operators are moving beyond traditional telecommunication businesses to explore new opportunities.

Traffic content is undergoing significant transformation driven by shifting consumer behaviour. The level of content consumption arises significantly; therefore, an increasing number of telco operators are entering the content space or strengthening their existing content offerings. They are creating content themselves or making partnerships with Over-The-Top (OTT) video service providers.

Artificial Intelligence (AI) could be essential to future business and digital transformation. To improve customer experience, identifying and learning customer behaviour with autonomous and intelligent networks are inevitable. Operators are focusing more and more on AI-based applications including digital assistants, customer care, advertising, network planning as a service.

Smartphones will remain the focal point of the consumer internet economy in the upcoming years — the range of connected devices higher than ever. Today's digital consumers will likely to become tomorrow's augmented customers, adopting emerging technologies such as virtual reality and AI. The telco sector is under pressure from slowing unique subscriber growth and regulatory interventions. However, upcoming new opportunities have the potential to provide the uplift to mobile operator revenue. Mobile operators are pursuing new incremental revenue opportunities in content, IoT, and 5G. They are looking and expanding their role in the value chain, from providing capabilities, tools, and support for partners to create IoT solutions and becoming end-to-end IoT solution providers themselves.

5G is now upon us, bringing with it the promise of multiple exciting new services and applications. Fifth-generation mobile networks are expected to transform the role of telecommunications in the society fundamentally. 5G is also expected to enable further economic growth and create a hyper-connected society, where there are not only all people connected to the network whenever needed, but many devices/things can be connected, too, creating the Internet of Everything. Therefore, 5G will enable a wide variety of new use-cases such as smart cities, smart transportation, home, and agriculture, etc.

5G is one of the most hyped technology currently. According to the Gartner Hype Cycle report (Figure 1), the International Telecommunication Union and other standards bodies are expected to ratify full 5G technical standards by 2020. Another Gartner report claims that 5G wireless network infrastructure revenue will reach \$4.2 billion in 2020, an 89% increase from 2019 revenue of \$2.2 billion.

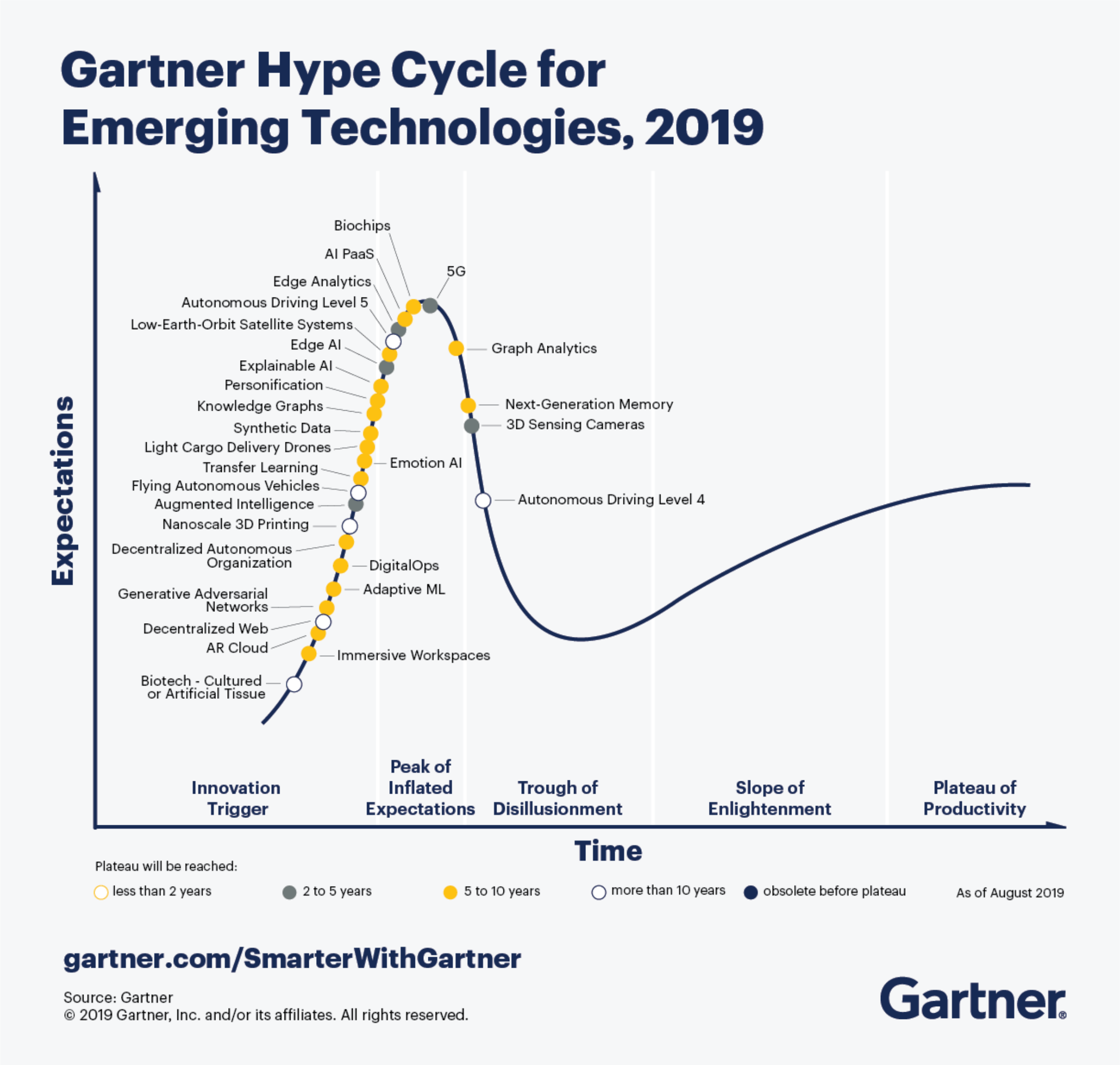


Figure 1: *Gartner Hype Cycle for Emerging Technologies, 2019 [1]*

With the emerging new applications, accurate traffic identification and measurements are crucial elements of creating high quality network business intelligence and network policy control. Content Service Providers (CSP) can not generate new subscriber services, optimize resource utilization, and ensure correct charging system without measuring and monitoring the traffic flow on their current networks. There are several techniques to identify traffic patterns, gain additional information and measure quantities, ranging from relatively simple to extremely complex. It is not fully understood how the various user demands can affect to the traffic flows, KPIs to quantify these effects are not defined yet. Every element of the cellular network need to co-operate to serve user demands, deal with traffic peaks, and have to obey Quality of Service (QoS) guarantees. In order to do this, traffic-related characteristics should be determined, and should be used during network planning, optimization and service shaping. This is exactly what motivates my work.

The outline of this work

In the case of 5G, so far there are just a few published measurement results on lab or live networks, regarding next-generation cellular traffic identification and use-case traffic pattern analysis. In addition to the measurements of the 5G infrastructure, the description of the upcoming use-cases need to be more accurate. Publications of next-generation protocols and even the general formalism of 5G are incredibly hard to find, hence posing a research gap.

It has not yet been examined how the new generation application traffic types or the currently existing ones will affect the live 5G network, and how the user experience will change in the future. As part of the preparation for the upcoming changes, the main use-cases regarding data and signaling resource allocation demands are to be defined and well understood. While the 5G white papers contain a lot of use-cases, this paper will extend the perspective with real-life network measurement results on the 4G live network. The results support the understanding of the current cellular network usage scenarios. Moreover, these results can help to define their traffic patterns and give a methodology to describe traffic of applications in general but in a bit more detailed terms.

Through the identification and modeling of given application traffic types, this current work supports application developers and service providers to prepare traffic-related strategies for network slicing, together with better orchestration of the cellular core. Furthermore, from a quite opposite point of view, this work provides an insight to characterizing traffic types that are the most suitable for the current HW and SW versions of the 5G network implementations.

The paper is structured as follows: Chapter 1. briefly presents the state of art and trends of the telecommunication industry. First, the key features of 4G application traffic types and their features are introduced. Afterwards, the main principles and expectations for 5G are presented. Finally, I give a brief overview of what the fundamental network traffic analysis methods are. Chapter 2. introduces the main 5G use-cases, then presents the architectural elements of the 5G test network and gives you an overview of the limitations and capacity considerations for each network element. In Chapter 3., I will explore the user-side modeling methodology and describe what parameters can be used to find the predecessors of 5G use-cases on existing 4G networks. During the modeling, I show the characteristics of 4 partly different user types and form the basis of a theory and model that describes its main features based on passive monitoring a Hungarian Service Provider's network. In Chapter 4, I present the 5G measurement environment, then design, implement, execute ,and analyze measurements based on the recorded use-cases. In Chapter 5, I summarize the measurement results, draw conclusions, and highlight the main findings of my work.

Chapter 1

Related Works

The growing number of users have the need for better QoS and new services with the explosion of mobile applications and the expanded reach of mobile connectivity. 4G became the leading technology across the world in the previous years, overtook 2G with 3.4 billion connections accounting. 4G's growth will continue, exchanging 3G and 2G in the following years. Meanwhile, 5G is now becoming reality. It will take time for 5G to hit critical number of subscribers, but some markets will grow rapidly. In this highly competitive telecommunication market service providers are busy with rolling out 4G networks to meet the growing user demands for faster and more secure connectivity and higher bandwidth. Many SPs have also started field trials for 5G, and preparing towards rolling out 5G deployments to capture new market opportunities.

1.1 Cellular network trends

Each generation of mobile technology has been motivated by the need to meet a requirement identified between that technology and its predecessor. The transition from 2G to 3G was expected to enable mobile internet on consumer devices, but while it did, added data connectivity. In 3.5G a giant leap in terms of consumer experience occurred, as the combination of mobile broadband networks and smartphones brought about a significantly enhanced mobile internet experience. It eventually led to the application centric world as we see today. From social media through music and video streaming to controlling your home appliances from anywhere in the world, mobile broadband has brought enormous benefits. It has fundamentally changed the lives of many people through services provided both by operators and third-party players.

The transition from 3.5G to 4G services has offered users access to considerably faster data speeds and lower latency [2]. The way that people access and use the internet on mobile devices continues to change dramatically. Operators often cite an increased level of video streaming by customers on 4G networks as a major contributing factor to this. The Internet of Things (IoT) has also been discussed as a key aspect for 4G, but in reality the challenge of providing low power, long ranging networks to meet the demand for widespread M2M deployment is not specific to 4G or indeed 5G.

There are multiple ideas exist on the purpose of 5G will be. One of them is that mobile operators would create a blend of pre-existing technologies covering previous generation technologies, Wi-fi, LPWA and others to allow higher coverage and availability, and higher network density in terms of cells and devices. Also, 5G will be an enabler to greater connectivity for Machine-to-Machine (M2M) services and the Internet of Things (IoT). This vision may include a new radio technology to enable low power, low throughput field devices with long life-cycles of ten years or more. Another one is more of the traditional 'generation-defining' view. It has specific targets for data rates and latency being identified, such that new radio interfaces and core architecture can be assessed against such criteria. This makes a clear demarcation between a technology that meets the criteria for 5G, and which does not.

According to Cisco's Global Mobile Data Traffic Forecast [3] on the global market of telecommunication, the relative share of 3G- and 3.5G-capable devices and connections surpassed 2G-capable devices and connections in 2018 (Figure 1.2). There were 30% 3G connections in 2017 compared to about 34% 2G connections, but by the end of the forecast period, there will be 20% 3G connections and 2G will only have 8% of connections. By 2022, there will be 3.4% of all devices and connections with 5G capability. The network evolution toward more advanced networks is happening both across the end-user device segment and within the M2M connections category.

Low Power Wide Area (LPWA) connectivity is meant specifically for M2M modules that require low bandwidth and wide geographic coverage [4]. It provides high coverage with low power consumption, module, and connectivity costs, thereby creating new M2M use cases for Mobile Network Operators (MNOs) that cellular networks alone could not have addressed. Examples include utility meters in residential basements, gas or water meters that do not have power connection, street lights, and pet or personal asset trackers. As the forecast states, the share of LPWA connections (all M2M) will grow from about 2 percent in 2017 to 14 percent by 2022, from 130 million in 2017 to 1.8 billion by 2022.

Cisco's forecast mentions 4G already carried 72 percent of the total mobile traffic and represented the largest share of mobile data traffic by network type in 2017. In the future, it will continue to grow faster than other networks, however the percentage share will go down slightly to 71 percent of all mobile data traffic by 2022 (Figure 1.1). Currently, a 4G connection generates about three times more traffic than a 3G connection. There are two reasons for the higher usage per device on 4G. The first is that many 4G connections today are for high-end devices, which have a higher average usage. The second is that higher speeds encourage the adoption and usage of high- bandwidth applications, such that a smartphone on a 4G network is likely to generate significantly more traffic than the same model smartphone on a 3G or 3.5G network. By 2022, 5G will support 12 percent of mobile traffic. 5G connectivity with its very high bandwidth and ultra low latency is expected to drive very high traffic volumes.

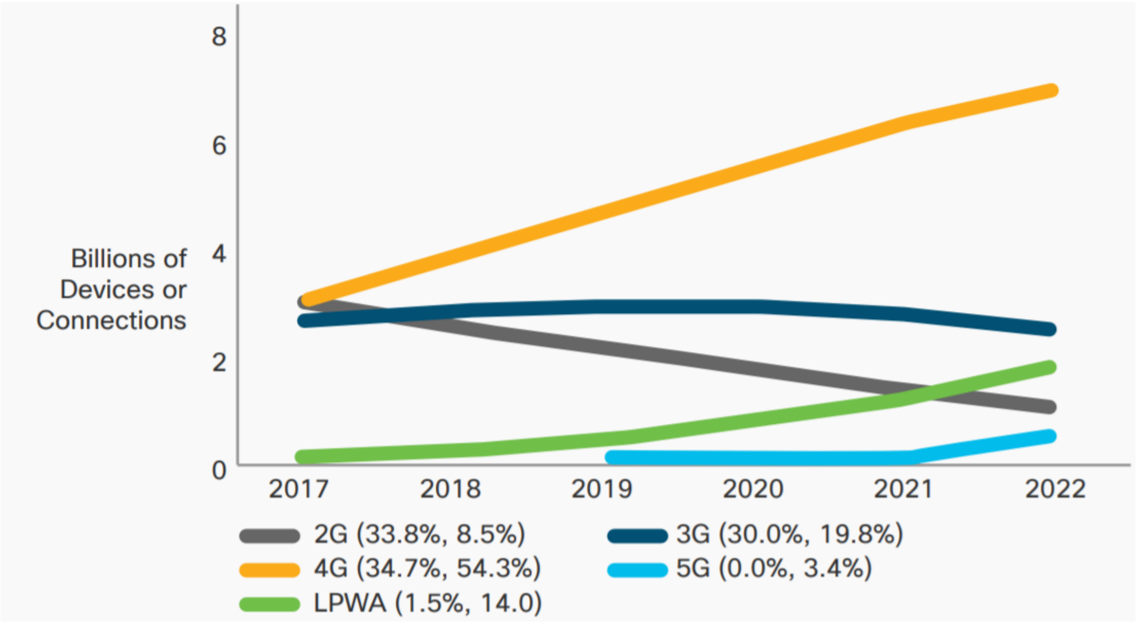


Figure 1.1: *Global Mobile Devices and Connections by Network Type [3]*

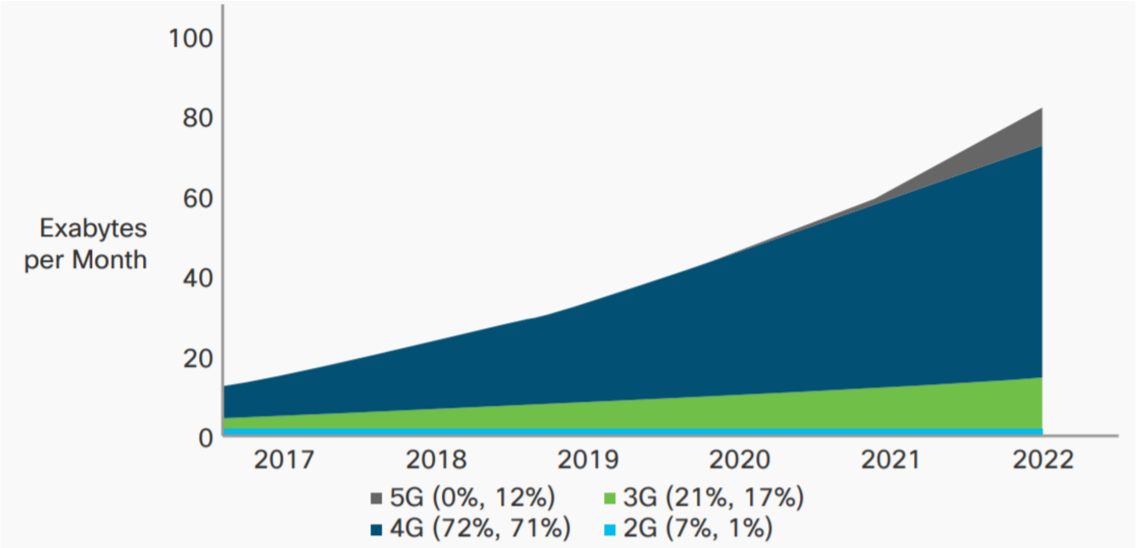


Figure 1.2: *Global Mobile Traffic by Connection Type [3]*

5G deployment probably will not be as quick as 4G's was. Slow expansion of 5G means that being first to market with 5G is less important than having a long-term strategy for 5G investment that creates value for customers. The rollout of 4G holds some important lessons. Because the quality of the mobile broadband experience relies heavily on network capability and capacity, network tests and consumer mobile broadband satisfaction tests will be even more important in the 5G world than for 4G or 3G. 5G devices will take time to roll out and become affordable.

1.2 The main features of 5G

5G is the next phase of mobile technology. 5G's primary improvements over 4G include high bandwidth, broader coverage, and ultra-low latency [5]. These features combined with enhanced power efficiency, cost optimization, high-precision positioning, massive IoT connection density and dynamic allocation of resources based on awareness of content, user, and location make 5G a flexible as well as a transformative technology. 5G will be able to accommodate IoT applications such as environment monitoring, various sensors and meters at the low end of the IoT spectrum. It will also support autonomous cars, smart grid, factory automation, and other tactile Internet-driven applications such as augmented and virtual reality. However, the actual value of 5G will lie beyond the connectivity and besides advanced application and massive IoT enablement. 5G can genuinely unlock the business value for the customers and create new revenue opportunities for the providers with enhanced network edge capabilities, data analytics, machine learning and artificial intelligence services. This technology is expected to solve or at least significantly improve frequency licensing and spectrum management issues. Currently, there are various standards, bodies, regulatory agencies, and industry consortiums focused on concerted efforts to resolve 5G issues, such as network standardization, spectrum availability and return of investment strategies to justify the investment associated with new infrastructure transitions and deployments. Given these evolving technology and business dynamics, we anticipate that some large scale commercial 5G deployments may not be executed until after the current forecast period (after 2022). A large number of mobile carriers perceive 5G as imperative for future growth and long-term sustainability.

As a result of this blending of requirements, many of the industry initiatives that have progressed with work on 5G identify a set of eight requirements [6]:

- 1-10Gbps connections to end points in the field (i.e. not theoretical maximum);
- 1 millisecond end-to-end delay (latency);
- 1000x bandwidth per unit area;
- 10-100x number of connected devices;
- (Perception of) 99.999% availability;
- (Perception of) 100% coverage;
- 90% reduction in network energy usage;
- Up to ten year battery life for low power, machine-type devices.

These requirements are specified from different perspectives, and they do not make an entirely coherent list. It is difficult to conceive of a new technology that could meet all of these conditions simultaneously, and 5G will not be capable satisfy all of these requirements at the same time.

Equally, while these eight requirements are often presented as a single list, no use case, service or application has been identified that requires all eight performance attributes across an entire network simultaneously. Indeed some of the requirements are not linked to use cases or services, but are instead theoretical goals of how networks should be built, independent of service or technology – no use case needs a network to be significantly cheaper, but every operator would like to pay less to build and run their network. It is more likely that various combinations of a subset of the overall list of requirements will be supported 'when and where it matters'.

ITU-R has defined the following main usage scenarios for IMT for 2020 and beyond in their Recommendation ITU-R M.2083:

Enhanced Mobile Broadband (eMBB) to deal with hugely increased data rates, high user density and very high traffic capacity for hotspot scenarios as well as seamless coverage and high mobility scenarios with still improved used data rates Massive Machine-type Communications (mMTC) also known as Mass IoT (mIoT) for the IoT use-cases, requiring low power consumption and low data rates for very large numbers of connected devices Ultra-reliable and Low Latency Communications (uRLLC) or Critical IoT (cIoT) to cater for safety-critical and mission critical applications.

1.3 Analyzing network traffic

Network traffic identification is essential for network management of ISPs to identify different applications, through which it is able to analyse, understand and predict user behaviour, ensure network security and control the network with high flexibility. Currently, internet traffic can be identified using different methods. The primitive application traffic identification technique is port-based, because traditional protocols, like email, use fixed port numbers which are assigned by Internet Assigned Numbers Authority (IANA). This method is simple to implement. However, it becomes limited because more and more private protocols start to use dynamic ports, and this type of applications has occupied a considerable portion of the total traffic. For example, the volume of Peer-to-Peer (P2P) applications has occupied almost 60% of the total traffic. Moreover, many applications try to allocate port number dynamically or encapsulate the original data into one of the traditional applications, like HTTP, to get through the firewall or the other defending systems, which makes traffic identification more difficult. Under this circumstance, the traditional port-based method is so vague that it can only be used for the rough classification.

In order to properly plan cellular data network services and to dimension required size of a serving system, it is necessary to know the possible users and to define their most relevant parameters. The parameters can be the required average or peak data throughput or the latency of data transmissions. Others can be based on higher-level requirements that are related to more complex applications. It is almost impossible to understand these needs well on the individual level, but with a group of users, the analysis can be very useful. The Service Providers (SP) main interest was always to serve first the interactive services and in the background, the less critical ones. As OTT services expanded and competition generated, eventually it had begun to use ports for non-critical applications which should have higher priority, such as OTT providers' image transfer on the HTTP port. However, there are cases where best-effort traffic are treated as critical ones. Typical examples are sending blank ID numbers with short message service, or voice over data traffic. Furthermore, there are a lot of different possibilities how to measure the user groups – one of the easiest method is to counter the transmitted data, or it can be based on Deep Packet Inspection (DPI) where even the application of the group is measured individually. Depending on the complexity of measurement, the results can be represented on detailed timelines. As I mentioned in my previous paper there are various traffic identification methodologies. The different type of analysis methods are discussed in the following subsections and summarized in Table 1.1, where I show the basis of recognition, with examples about the strengths and weaknesses of the used methodology [7].

1.3.1 Data traffic

Measuring the transmitted data of a users' group for a couple days with arbitrary resolution [8]. Transmitted amount of data can be measured at the packet core or at the connecting IP routers to the user's groups. The measurement results shall be the same as the Call Data Record files originating from the mobile packed core 'Ga' interface [9].

1.3.2 Recognition based on standard ports

Some application servers work with fixed ports which are specified by international standardization organizations such as IANA [10]. Based on the standardized recommendations, different applications traffic can be easily separated. For example, HTTP for 80, HTTPS for 443, FTP for port 21 or the extremely popular P2P applications with port 6346-6347. However, many newer applications do not have own, dedicated ports. Thanks to the rapid development of applications, ports allocation organizations can not always or can not fully properly track these changes. Application developers have a free choice on what ports they want to use for their application, they can freely decide to use the 80 and 443 ports for a peer to peer (P2P) application or starting HTTP services on port 21. In many cases new applications use unpublished protocols with a port which is used for a whole other purpose compared to the standard. Port-based traffic content recognition give us just a general picture from the actual content categories, however it can be extremely effective.

From another aspect, in a closed group communication the used port of an application or the actual role is not important, nevertheless the size and frequency of the messages are more significant aspects. It also means that the L4 protocol-based application detection is only usable for scaling the type of communication. To measure this is more than easy: from any part of the mobile packet core or connecting IP network, we can get individual IP or group level information, by well known monitoring services as openflow or netflow [11].

1.3.3 File extension type

In case the traffic is not encrypted, the transmitted traffic content can be recognized. In case of TCP port 80 traffic, one method for that is to recognize the HTTP field, and there the name of the file. It can be read from the extension what kind of file is that without any problem. Table 1.1 shows some easily recognizable image and application file types. The process is similar in case of video files or other kind of applications, as well.

1.3.4 Deep Packet Inspection

As we see, port based recognition is ambiguous. To identify an application based on traffic, it is needed to examine and search unique fields in the higher layers of OSI model; this methodology is often called as Deep Packet Inspection (DPI) [12]. An example for this method is recognizing HTTP SYN and SYN_ACK messages, or FTP port commands. Knowing the most popular protocols, traffic can be examined not only based on ports, more examining packets' header or a couple bytes of the traffic applications can be recognized also, even though they operate on other ports.

1.3.5 Statistical analysis

In most cases nowadays, the end-to-end data transmission layer is encrypted, hence content identification is getting a really tough task. Above the IP layer all data content looks like randomly filled bits, therefore any port or DPI based recognition is impossible. Traffic can be identified only from source and destination IP addresses, timing or other behaviour patterns. Naturally this method also can be supported by machine learning to increase precision.

1.3.6 Recognition with help of the OTT provider

From all recognition methods this is the most accurate one, where the OTT application developer gives the exact details and parameters for the application to the SP. This could be a simple IP-based recognition, when the destination IP address and port pool of the servers will define the application traffic. Naturally these can not be fully open in case of P2P connections, or when the application should be controlled within some kind of data center group. In other cases the OTT developers share the exact protocol description with the SP. Then the provider is able to detect the presence of the OTT application precisely, without knowing the IP pool of servers or ports. The advantage of this method

Type	**Basics of recognition**	**What to recognize**	**Unknown and bad recognition**
Data only	IP packet data	only the size of transmitted data	no details about traffic type
Port based	standards of IANA	HTTP(S), FTP, Telnet, ssh, some p2p	not standardized protocol usage
File extension type	file/video/program name open database	exe, bat, jpg, gif, png, txt, 3gpp,mpeg, mp4, stream, javascript, json	if URL not contains extension
DPI	standard apps. even on non standard ports	HTTP(S), FTP, Telnet, ssh, some p2p	when the application is not standardized
DPI with Machine Learning	applications based on closed protocols	skype, facebook, viber, p2p, etc.	in case of fast changing applications
Statistical analysis	behaviour analysis	applications inside encrypted traffic	in case of changing applications
With the help of OTT	IP address, port, protocol IDs	OTTs sharing protocols with SP	OTTs not sharing protocol type

Table 1.1: *A summary of features for traffic identification methods*

is to detect even P2P traffic between subscribers. However it is necessary to frequently exchange information between the SP and OTT, and also the legal contracts with national regulation shall support this, as well.

Chapter 2

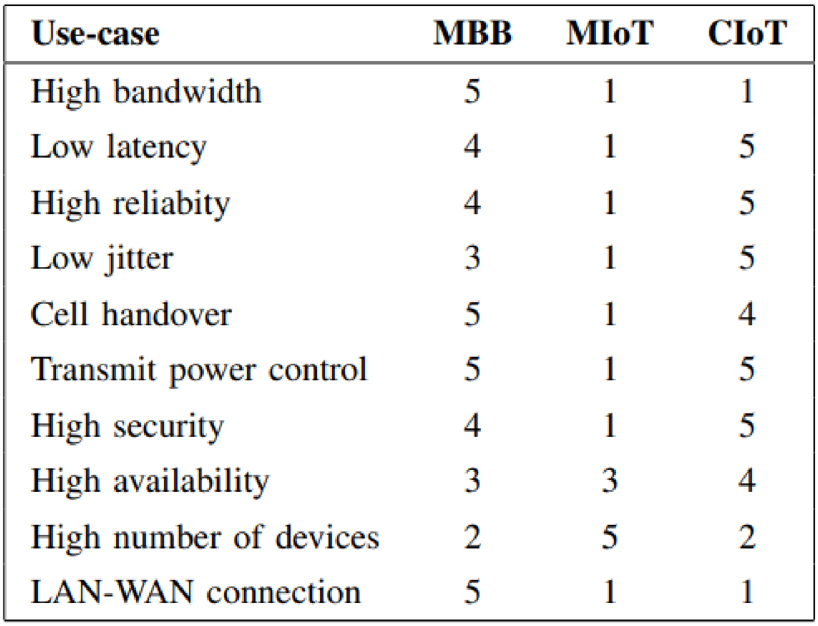
Service Provider perspective on 5G traffic and the network

Standardization bodies continuosly keep an eye on the new opportunities, enabler functions and technologies that van define the upcoming use-cases. The user expectations keep growing every year – as well as the performance of cellular devices. 5G delivers basic connectivity as well as mobile data services that are always available to customers when they need them. Networking services for applications with high computing requirements will be enabled by 5G. The growing sophistication of Artificial Intelligence and the availability of cheap computing power will drive the widespread adoption of digital assistants, intelligent IoT nodes and automated industrial processes.

Industry stakeholders and standardization bodies have identified several potential use cases for 5G, which pose demands for very diverse requirements. Most of these can be identified within three primary categories[13]: enhanced Mobile Broadband (eMBB), massive IoT (mIoT) and critical IoT (cIoT) (Figure 2.1 and 2.2).

2.1 Enhanced Mobile BroadBand (eMBB)

eMBB will be the key proposition in early 5G deployments and will drive increased performance, functionality and efficiency. Out of these three use-cases, "Massive Broadband" is currently the most widely spread – subscribers are humans. The main requirement is to serve the user with as high data rate as possible, while keeping the latency and end-to-end response time low. It will support high definition video (TV and gaming), and smart city services (video cameras for surveillance). eMBB is targeting mainly general subscribers, their primary interest is to provide better Quality of Service (QoS) and user experience. Furthermore, it provides higher bandwidth and higher speeds for densely populated urban areas and event locations, such as stadiums. Broadband connectivity is expected from application services delivering augment and virtual reality, 4K/8K video streaming, and seamless cloud services on demand.



Use-case	MBB	MIoT	CIoT
High bandwidth	5	1	1
Low latency	4	1	5
High reliabity	4	1	5
Low jitter	3	1	5
Cell handover	5	1	4
Transmit power control	5	1	5
High security	4	1	5
High availability	3	3	4
High number of devices	2	5	2
LAN-WAN connection	5	1	1

Figure 2.1: *Use-cases and their corresponding, important parameters (scale-importance: tolerant/low: 1, normal/medium: 3, major/high:5)*

The push for eMBB in 5G is a continuation of the 4G mobile broadband transformation. The emphasis on eMBB will sustain the 5G business case, and drive 5G network investment, in the absence of any new mobile use cases. However, under current operator business models, building a ubiquitous 5G network to support the latency that autonomous vehicles require remains as unrealistic as funding the level of small cell deployment needed to support a seamless augmented reality experience.

2.2 Ultra Reliable Low Latency communications (uRLLC)

The user base of cIoT on the current cellular networks is negligible. This is due ti the fact that service providers can not guarantee sub 20 ms latency and high level of QoS in case of packet delivery. However, the main goals of uRLLC include very low latency, packet delivery guarantees, and accurate user localization. SPs offer various opportunities for the possible use-cases, but concrete business needs do not exist yet. From a technical point of view, examining cIoT is maybe the most labour-intensive task among the three main 5G use-cases, as concrete user demands are not known. These use cases serve the real-time interactions for mission critical communications, such as autonomous driving, robotic control for industrial automation, drones and remote surgery medical care systems. The tactile response time is expected to be less than 1 ms. Public safety and emergency services supporting disaster response and location services are critical for lifeline communication and recovery [14].

2.3 Massive IoT (mIoT)

The need for machine-to-machine and machine-to-human communication is growing rapidly. Therefore, massive IoT will be a key application type in 5G. In mIoT, 5G network have to be capable of dealing with the high density of IoT devices. It will serve billions of low-cost, long range, ultra-energy efficient devices, machines and things that need connectivity from remote locations as well as cloud applications with periodic, infrequent communication. It is important that devices have to be able to connect as groups to the network without individual authentication. MIoT devices need to be cheap but unlike eMBB devices, they do not have high calculation capacity processors neither memory capacity. Also, their radio receiver/transciever module have to be as primitive as possible. Furthermore, mIoT features include 10 years battery life-time, which can be achieved only with highly efficient communication strategies in order to minimize power consumption. Fortunately, these kinds of applications and traffic patterns already exist on cellular networks. LPWA cellular technologies were introduced with Narrow Band IoT (NB-IoT) and LTE-M in 3GPP Release 13 [15] with further enhancements in Release 14. they are aligned to the improvements of the 5G architecture. NB-IoT use-cases do not hit the critical amount of end-devices yet, but traffic patterns from currently operating live network are more than enough for further examinations.

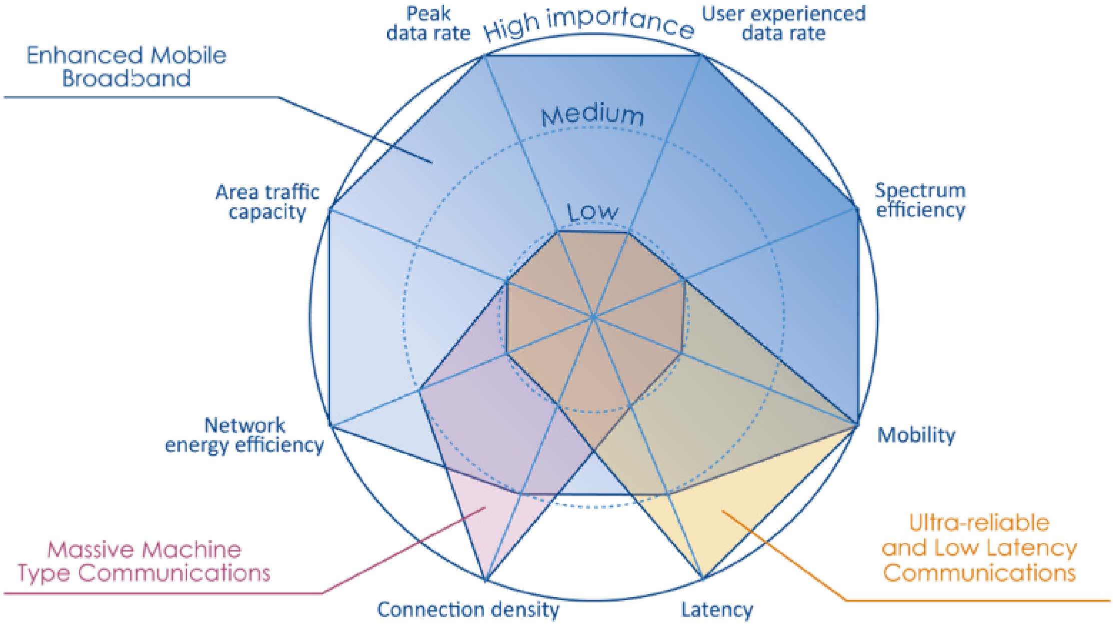


Figure 2.2: *The expected requirements for 5G application-types [16]*

2.4 Network considerations for 5G

In Chapters 2.1 through 2.3, I briefly outlined which are the main use-cases types that 5G networks can support. However, from a network perspective, I have not yet shown what use-cases are network-friendly or easiest to service operators. It is clear that like all technical systems – including 5G infrastructure with the terminal equipment and server – has an operating point, an operational boundary that is easy to serve and is not optimized. Since the 5G network is primarily designed for data transmission, it is also worth considering what data format or transmission structure the network can comfortably serve. In other words, what kind of transmission you can transmit with the best efficiency and what you can not between the two endpoints. To better understand the parts, I have plotted the entire data transmission path in Figure 2.3, and each element is detailed here, going from the user equipment to the server.

2.4.1 User Equipment or subscriber endpoint

The equipment connected to the cellular network can be divided into a large number of components. It is indisputable that its essential parts are the human or other machine-connected interfaces, the internal data processing and structuring element, and the modem connected to the cellular network. The human/machine interface is mainly located in the presentation layer and is one of the essential places in the whole network because the use-cases and user experiences are realized here. We can say that the purpose of the whole network architecture and transmission technology is to provide a perfect user experience for the connected machines or human users. Of course, defining user experience is not a straightforward task either.

The user experience depends on the entire network, to provide for each layer and its elements the best possible quality, and it may hide the random error of other devices or layers. Another critical part of the end-device is the internal processing and structuring element. The task of this element is to communicate with the network to guarantee the best user experience and decide which user request has priority and which has not. For example, if a mobile phone had a higher priority on displaying a website rather than receiving an incoming call, the user could even consider it as a network error, while a poorly implemented priority management is the source of the problem on our end-device. However, this decision centre has to face increasingly complex tasks as technology advances. End-devices have several different interfaces towards the network, such as 2-3-4 and 5G interfaces, as well as two types of WiFi at 2.4 and 5GHz, possibly Bluetooth and NFC. This implicates to go through a fairly complex decision tree. Especially when you combine the user experience with multiple applications at the same time. There might be complex cases where the decision of the end-device is questionable from the user's point of view.

From the third point of view, the end-device must be able to handle several different network interfaces. These may include interfaces that support 3GPP technologies as described in well-structured standards 2-3-4-5G or IEEE supported by Bluetooth and WiFi []. During this paper, I am focusing on 4G and 5G network interfaces and their user experience. When controlling these interfaces, both the end-device and the user should consider the feedback of the network and interface properties. The network equipment send signals its status to the end-device. These messages contain basic information about whether the network has been successfully contacted, is the network available, and what kind of permission has been granted to the end-device. Network interfaces know what kind of network resource they have, how much bandwidth is available for them. However, they do not know what types of packages can be easily serviced. The end-device logic has the ability to test for the status of a connection, but can only obtain indirect information about the current transmission properties. Even if you know something about it, it may not be able to pass on any information to the application layer. It would be interesting to publish a standard or feedback process that indicates to the application itself that the network does not currently have sufficient capacity for this data stream, but could support a lower version or quality. As an example example, Youtube would only display audio, or a series of much worse quality frames, rather than a paused frame.

2.4.2 Radio Access Network

The system between the end-device's mobile network interface and the base station is determined by a great number of parameters. This can be leased frequency system (2-3-4-5g) or unleased frequency band transmission technology too. Standards [17] define the essential elements, however, manufacturers are often plan arbitrary the packages scheduling management. These parameters are often freely adjustable, to a specific base station or user group.

The vendors do not make these parameters publicly available and since they sometimes change their back-testing unclearly, so neither the service providers nor vendors put emphasis on producing reports and analyzes based on these parameters. But for the 5G use-cases, this is essential. It is clear that without measurement, it is not possible to provide feedback on the quality of the network to either the device or the service provider base stations. There is a clear need for different scheduling and load balancing between serving when the terminals are sending many small packets and even if sometimes sending large packets. There are different needs depending on the traffic direction (downlink or uplink).

2.4.3 The Evolved Packet Core (EPC) and the IP transmission core

After the radio base stations, service networks typically have dedicated fiber or leased lines for IP data transmission up to EPC. It is often magnified that the bandwidth in these areas can be considered infinite. However, many base station's traffic is not served individually by one IP and one Core network; these traffic are aggregated, resulting 1 IP and Core device have to deal with millions of user's traffic. Manufacturers and service providers use equipment with high capacity, but it is clear that they have negative (besides many positive) effect on network traffic. Manufacturers specify the types of interfaces on the device's datasheet – which can currently be up to 100 Gbps and mention how much packet transfer capacity they have. So, you can see that datasheet categorize the interface

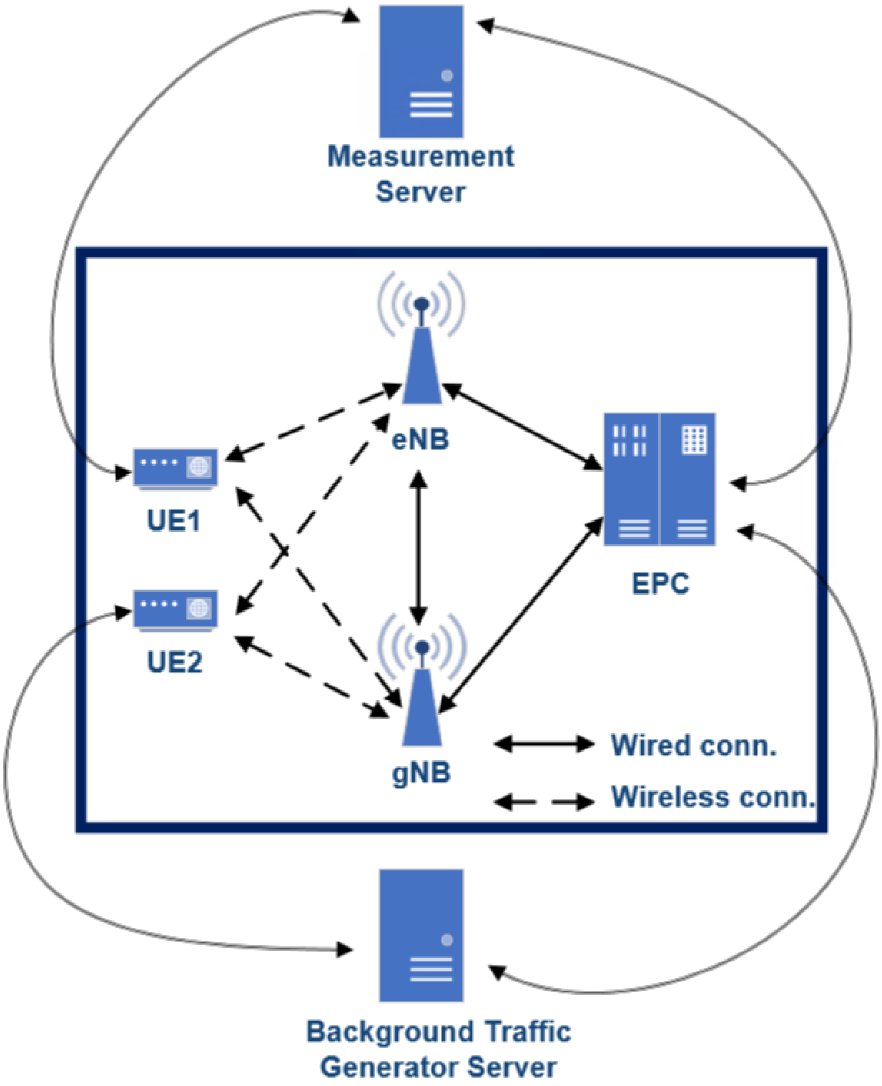


Figure 2.3: *The designed architecture of the 5G testbed – as Option 3.X (discussed in Chapter 4)*

by bandwidth, but they mention the packet capacity for the transfer capability of these devices. It follows that transmission can only be guaranteed if the packets are the right size. [6]

In network equipment dealing with data streams management is quite a comlicated task. Manufacturers provide several different prioritization methods [18], but hardly discuss the of inbound and outbound packet storage capabilities. However, this is a basic need for a well-tuned system. A centralized equipment that handles the traffic of millions of users at once will surely need to decide on a priority basis at the expense of one or another data stream. Wireless service providers often use QCI classes as recommended by 3GPP, which accept and manage all devices between RAN and Core. However, they can only distinguish between classes, but if a critical amount of packets arrive from the same class, then standard processes must be triggered due to packet loss.

2.4.4 Server-side services

The documentation of server-side services is perhaps the most detailed and published part of the user experience line [19], [20]. Packet congestion and service strategy are key elements to serve user and server needs. User preferences are also worth to take into consideration here; for example, an online video request is worth serving faster than an FTP file transfer connection. From the analysis of traffic patterns, we can provide certain estimated for the distribution of traffic over time, as well as its scaling. By separating the application-types and the traffic of the underlying servers, we can analyze the data transfer characteristics of each application-types.

It is clear from the number of end-devices that we need to look at a great variety of directions to determine what a user experience will be when using next-generation cellular networks. Therefore, it was our primary task to measure the type of traffic passing through current networks and to develop a methodology based on complex characteristics. Nevertheless, when this new methodology, complex data can be captured, where the user experience regarding data bandwidth and latency. Still, this can support us to determine what kind of user experience, improvement, or degradation can be expected in future networks.

Chapter 3

Analysis and modeling of current mobile network traffic

After the mass launch of smartphones and applications, the hunger for mobile data traffic has increased significantly. Mobile network operators were well aware that applications having video sharing functions, such as Youtube, Facebook, and Instagram would generate a high amount of data. Operators had developed different tariff packages in order to increase their revenues. The competition among SPs forced them to find out which applications and services are those that users most emotionally attached to. Previously, it was enough to determine about user groups how much data and when would they like to transmit, but this approach has been replaced by much more profound analytical methods. SPs were curious about which applications, websites are the most visited and when. This information is obtained through Deep Packet Analysis and mainly used to meet business needs. Due to the evolving business requirements, researchers have made significant improvements in relation to traffic analysis and Deep Packet Inspection (DPI). Nowadays, it is easy to provide a fair estimate about which application generated a specific network traffic flow – even for encrypted traffic.

3.1 Live network measurments

Regarding network design, the packet arrival rate and the size of data packets by a user or group of users have exceptional significance. In the L2 and L3 network layers switches, and routers are responsible for processing packet headers and deciding which direction to forward them. Received packets have to be saved in memory then forwarded to the chosen outgoing interface. This requires different sized buffers and transport tables. An L2 or L3 transmission device will react differently and transmit data packet at different speeds depending on their size. This paper does not cover the internal packet transfer properties of routers and switches, which are considered as Device or Network Under Test (DoNUT) and defined as a black box. By examining the inputs and outputs, conclusions can be drawn about the operation of the network and network elements [21].

One of the important features of IP-based networks is that they can dynamically handle the traffic and traffic-directions. This flexibility and scalability also means that by analyzing traffic at a node, we can make estimations about the whole network traffic. From our analysis point of view, this is a disadvantage, since I would like to examine and draw conclusions separately for each upcoming use-case of 5G. To gather such accurate information, we have contacted one of the largest Internet providers in Hungary. The measurements were carried out within this Service Provider's network. Measurements took place during September and October, 2019. Analyzing live network measurements is the best method to predict future traffic, even for cellular networks. By examining the current SP's network, several conclusions can be presented regarding the live network traffic patterns, and generic estimations can be made on the upcoming 5G traffic.

3.2 Identifying application types based on traffic flows

The distributions of packet size and packet interarrival-time are essential parameters for the live network. Network operators aim to meet customer needs as quickly, efficiently, and properly as possible. In the previous section, I have shown that different usage demands require different packet size distributions and, thus, service requirements. However, the user experience, whether human or machine, is not necessarily the same as the packet size distribution.

If you are looking for a higher-level interpretation of customer experience, you might examine the network-side reactions to different kinds of user activities. For example, when a user opens a website, the time between sending the request and the arrival of the response, all of its related packages defines a flow [22], which has a related user experience. Flows can be used to track the activity of a user or group of users. From another aspect, a flow can be a mixture of various traffic types (such as http or/and ftp samples) traversed between the same IP source/destination addresses and ports. This was already detailed in Chapter 2.

In a Service Provider's DPI system, it is typically easy to separate anonymized user groups to different kinds of flows. Flows can be characterized by two parameters: the sum of transmitted data, and the total number of packets, where you can see how many packets have been handled by that flow. Of course, we can separate them to up- and downlink traffic as well.

The purpose of this paper is to provide models for the expected traffic types of future 5G use-cases through the identification of the essential properties of these traffic-types. The analysis takes place on a pre-5G network where the generated traffic approximates the traffic patterns of the identified application types. In order to draw the right conclusions and to make any predictions about the behavior of 5G applications, it is necessary to know the current mobile traffic trends. To determine what features eMBB, mIoT or cIoT will have, we need to know what trends are driving the current mobile networks.

In our previous work [23] we classified users based on their content nature into three categories. These categories – eMBB, mIoT and cIoT – already exist on 4G networks, as some kind of forerunner of future 5G use-cases.

With 10 different features it can be described (see Figure 3.1) to which group a certain use-case belong to, and then identified these groupings within a live mobile network, where the subscription was bought by a service partner because their primary interest is these kinds of use-cases. With the help of Access Point Names (APNs), we examined these distinct user groups without their user identification. Although the exact borders of the three APNs can not be drawn precisely, the distinguishing features are shown by Figure 3.1 are more than enough for our examination. Nevertheless, strictly speaking, these forecasted 5G use-cases are not existing in current 4G networks; there are only groups with similar purpose can be found. Parts of the following brief description of APN scenarios are from our previous work [13].

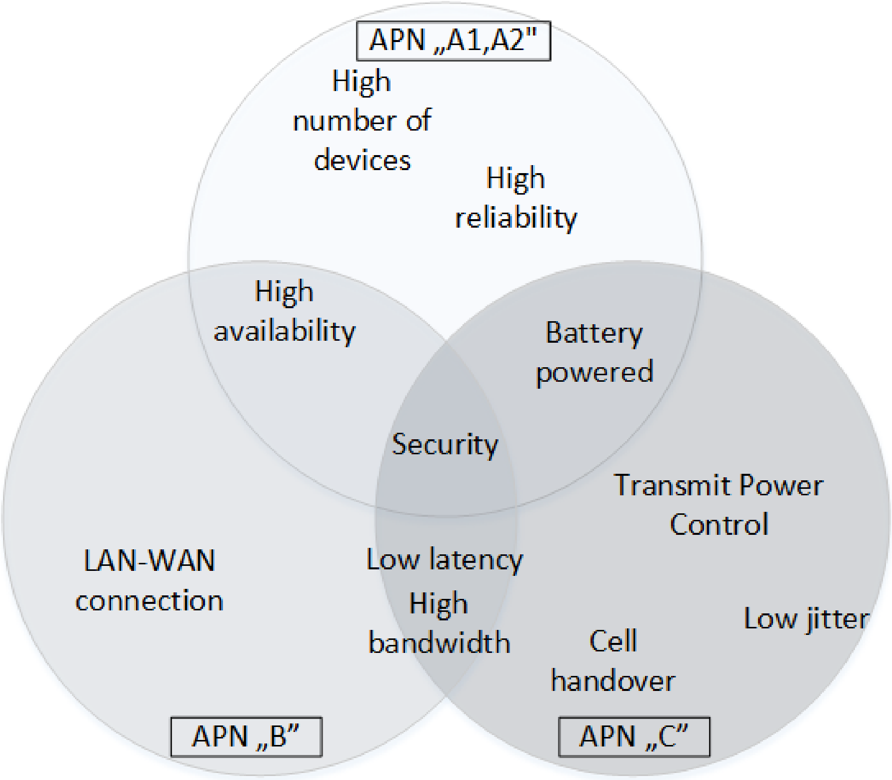


Figure 3.1: *Mapping 5G use-case characteristics to APN traffic scenarios (marking some distinguishing features in this diagram)*

I had the opportunity to monitor the traffic of a Hungarian SP, and its various APN's traffic. There are certain traffic types that the operator aims to gather at given APNs. The traffic I analyzed was monitored in flow separation, each flow representing a user or user group experience. The APNs themselves are nationwide, so the random 10,000 sample is sufficiently representative, not too distorted by the habits of particular user groups. The report includes the number of packets of each flow and the total number of bytes per flow for up- and downlink on each APN.

For each APN, I plotted the size of the flows as a function of packet numbers. This method can be used to identify the traffic type that a given APN typically serves and the characteristic features that can be observed in traffic patterns. Also, I had prior knowledge about the type of traffic on each APN. For privacy reasons, I do not label each APN with their real name; it is not relevant to the content of this paper in what name they are called or to which service provider they belong.

3.2.1 M2M communication

APN “A1 and A2” serves mainly Machine to Machine type communication. These are the forerunners of critical or massive IoT type use-cases. There are lot of subscriptions, with very-low-cost devices. The main business-case for the owners of such APNs are companies working with sensors of public services, or alarm devices of houses and cars. The UEs send keep-alive messages regularly, asking for status update in a planned interval to one or more central server.

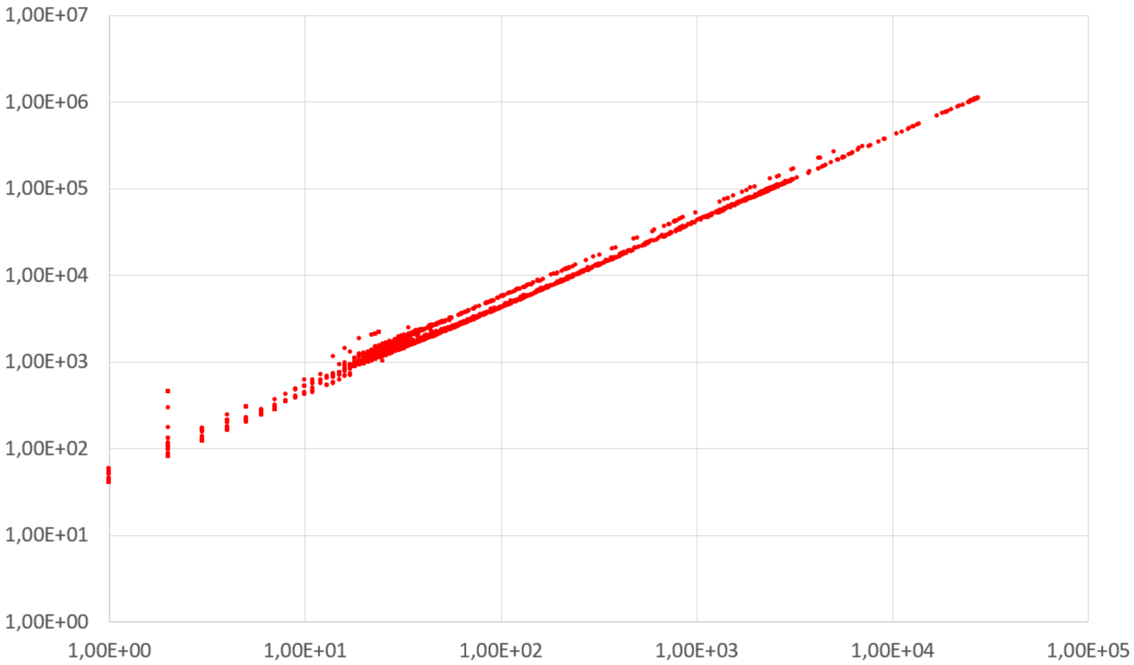


Figure 3.2: *Downlink traffic flow distribution of APN "A1" (x axis: number of packets, y axis: Size of flows [Byte])*

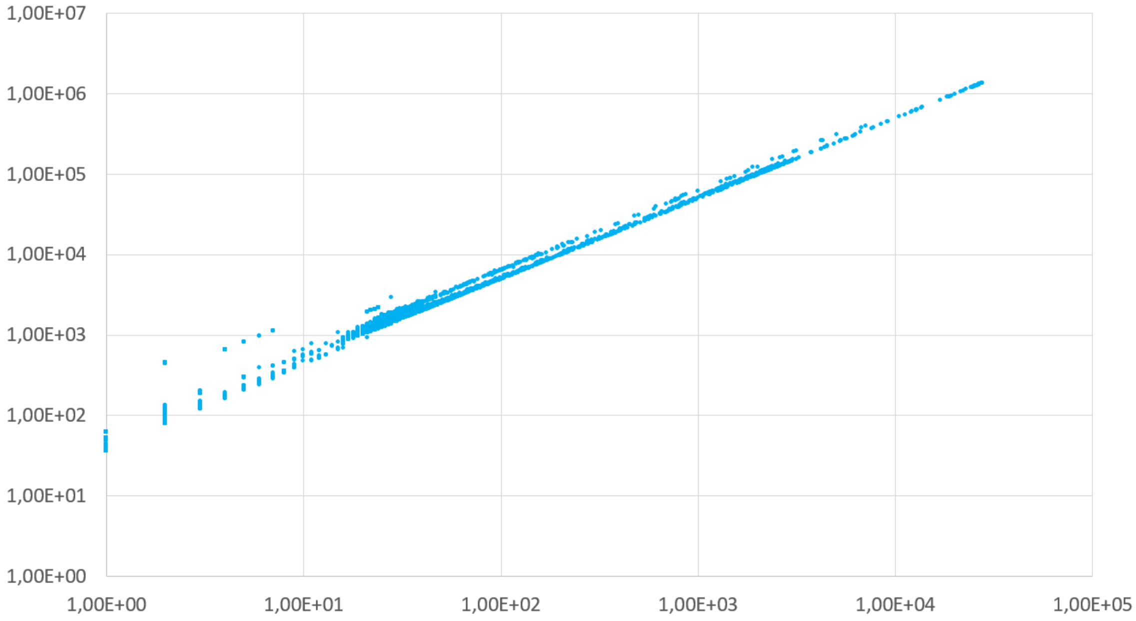


Figure 3.3: *Uplink traffic flow distribution of APN "A1" (x axis: number of packets, y axis: Size of flows [Byte])*

The APN "A1" traffic shown in Figures 3.2 and 3.3, while APN "A2" in Figures 3.4 and 3.5. In these figures, a few "vertical line" can be seen, which represent specific events with a well-defined number of packets. This indicates that these APNs serve mainly M2M traffic. In the case of M2M applications, there are strictly defined transmission strategies, where the size and the number of packets sent by end-devices are determined.

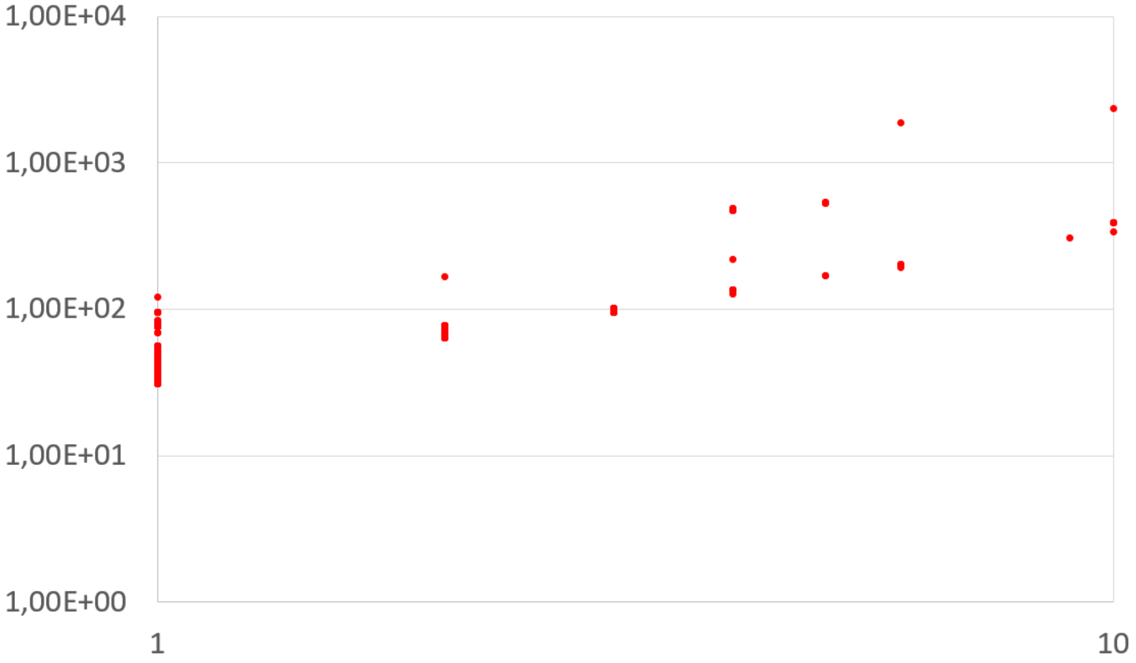


Figure 3.4: *Downlink traffic flow distribution of APN "A2" (x axis: number of packets, y axis: Size of flows [Byte])*

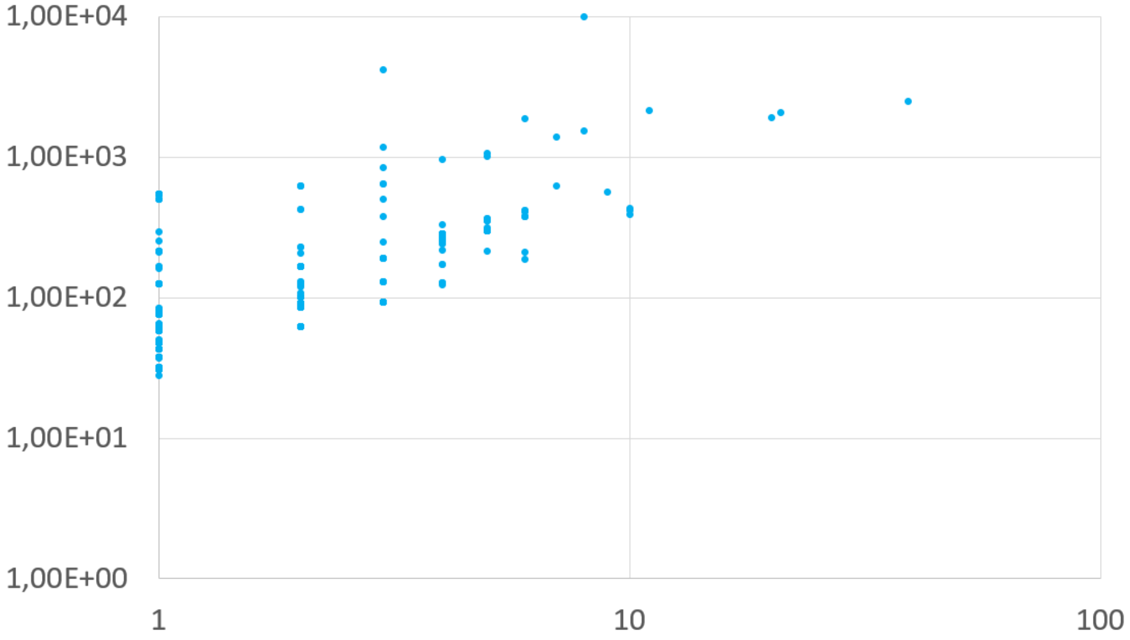


Figure 3.5: *Uplink traffic flow distribution of APN "A2" (x axis: number of packets, y axis: Size of flows [Byte])*

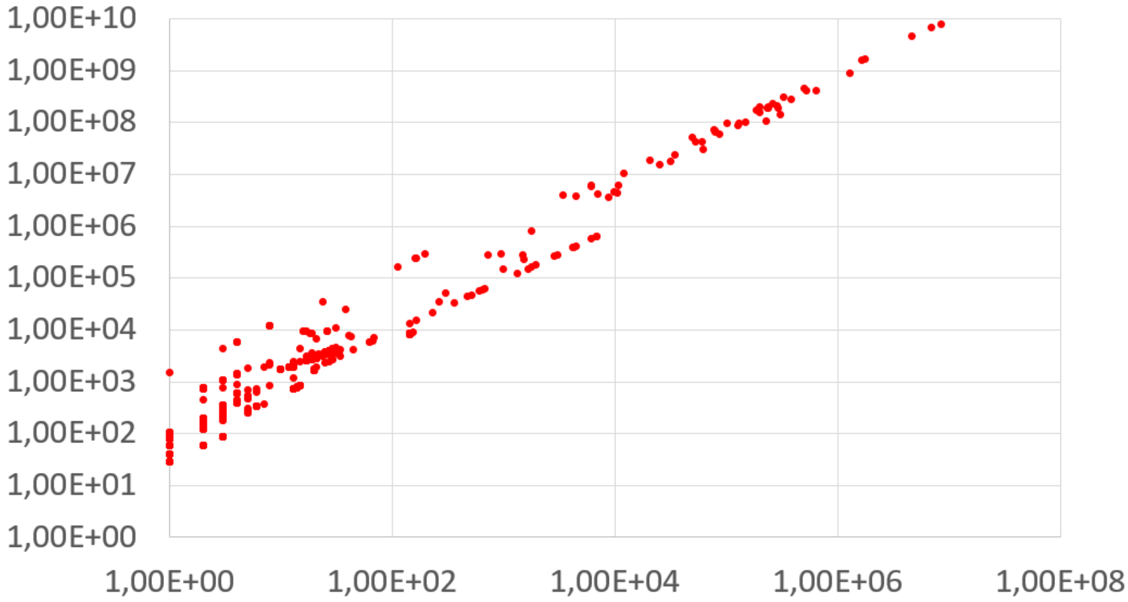


Figure 3.6: *Downlink traffic flow distribution of APN "B" (x axis: number of packets, y axis: Size of flows [Byte])*

3.2.2 More subscribers behind Cellular routers

APN “B" is a network segment connected by cellular routers [24]. Companies or individuals using these subscriptions to utilize one connection to the SP and then via network address translation (NAT) connect more local users via cable or wifi. Main use-cases include surfing (web), communicating (e.g., Facebook, Skype), or watching online videos (e.g., youtube). The owner of these subscriptions can use IP data tunnelling between end subscriptions and centralized servers. Here the data consumers are mainly humans, but the operator of the devices connecting to a cellular network can be a company, not allowing the unnecessary reboot cycles. Figures 3.6 and 3.7 show the flow distribution of APN "B" traffic.

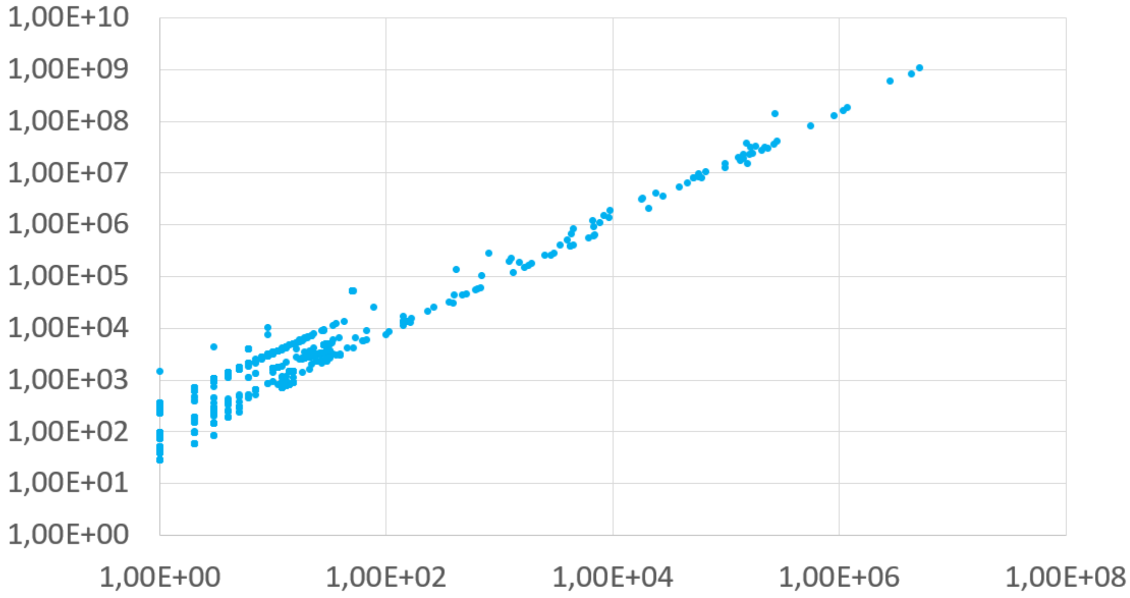


Figure 3.7: *Uplink traffic flow distribution of APN "B" (x axis: number of packets, y axis: Size of flows [Byte])*

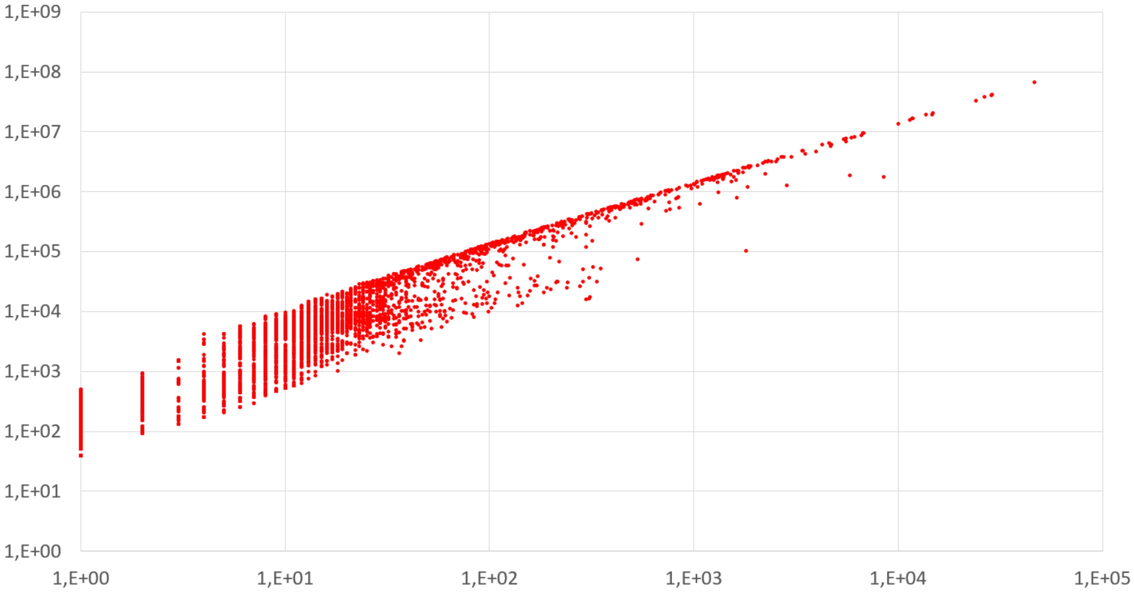


Figure 3.8: *Downlink traffic flow distribution of APN "C" (x axis: number of packets, y axis: Size of flows [Byte])*

3.2.3 Smartphone

APN “C” is utilized by average cellular mobile subscribers: mainly smartphones, and USB sticks used by humans. These customers are also called mass-market at the SPs, who are selling these for anyone entering a customer representative shop. Based on the uplink and downlink traffic of the APNs in Figures 3.8 and 3.9, it can be seen that the flow sizes and the value of packet numbers are much more varied. It is clear that discrete lines are formed at certain packet numbers, each line representing a specific event. The varied flow of traffic indicates that the APN "C " serving regular mobile users; thus, it be perfectly capable of simulating eMBB traffic.

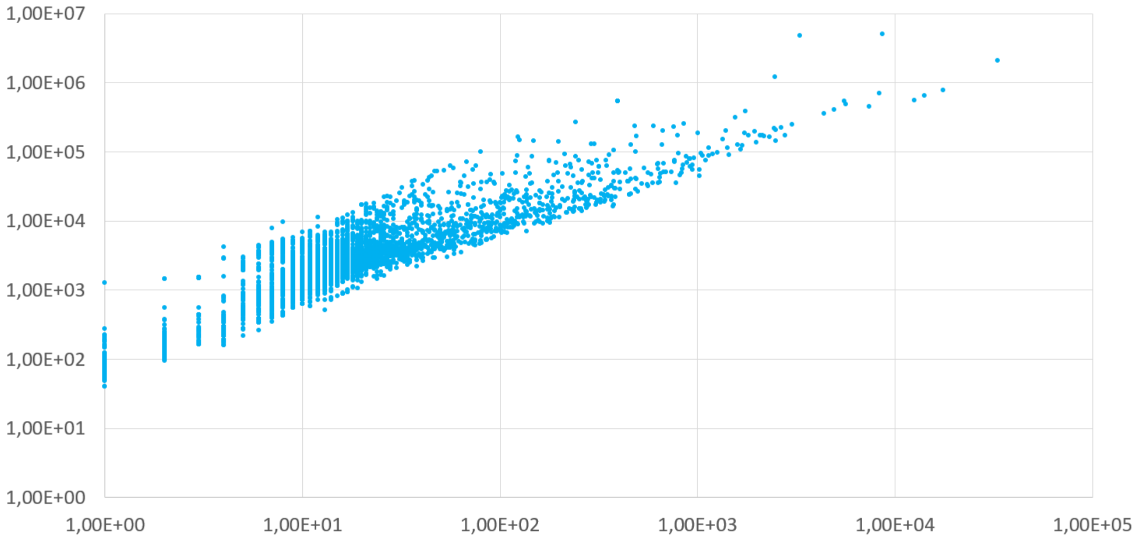
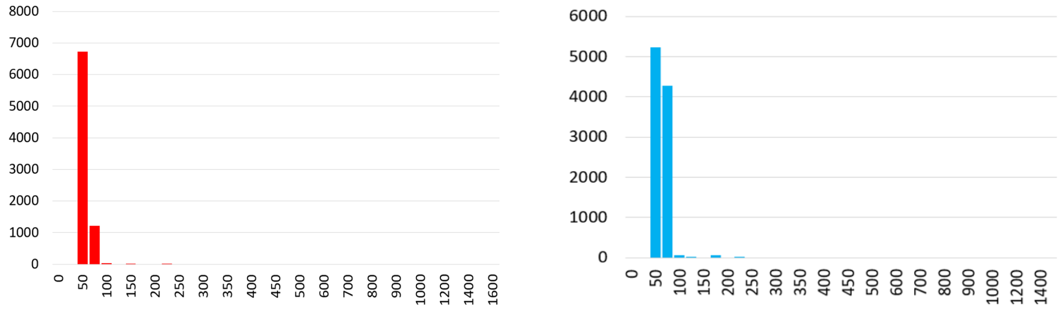


Figure 3.9: *Uplink traffic flow distribution of APN "C" (x axis: number of packets, y axis: Size of flows [Byte])*

3.3 Defining measurement scenarios

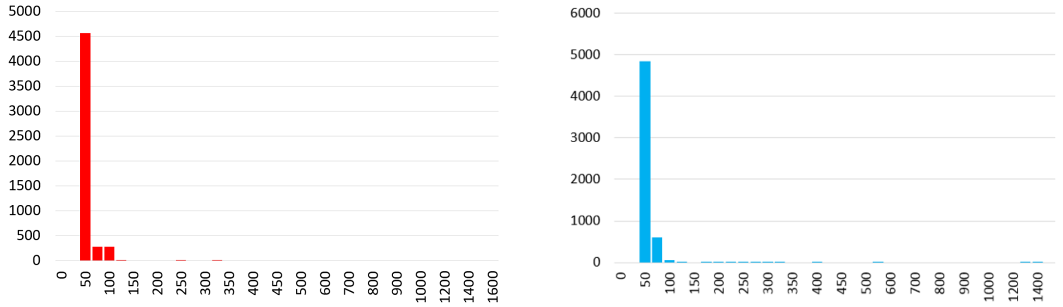
To determine the different application traffic types, I analyzed the measured traffic of the APNs on two different kinds of histograms. First, I examined the distribution of packet sizes of the flows. Second, I plotted the number of packets that each flow consists of. I have created both types of histograms in separate up- and downlink directions. These histograms approximate the density function of the packet numbers and packet sizes of the flows for each APN. Furthermore, they help to define the key parameters related to the discussed application-types traffic for the simulations. Once I have identified these parameters, the background traffic and the application traffic types can be simulated. It is important to note that I have only measured downlink traffic. The reason for this is that the contacted SP's 5G deployment has 4G architecture for uplink direction as a part of the current Option 3.X architecture [25]. However, the methodology for the uplink is completely identical with measurements that I made for the downlink channel.

Note, that unfortunately, there are no real live representation of the cIoT traffic yet, as its significant benefits such as sub-ms latency and 99.99999% network availability can not be guaranteed at current mobile networks. Therefore, cIoT traffic is not to be included in my measurements. However, latency can still be measured, which is one of the essential requirements for these applications.



(a) downlink packet size distribution **(b)** uplink packet size distribution

Figure 3.10: *APN "A1" traffic flows [Byte]*



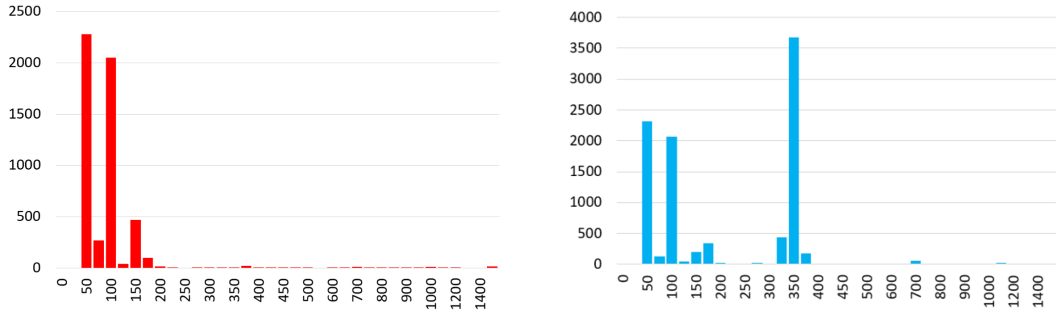
(a) downlink packet size distribution **(b)** uplink packet size distribution

Figure 3.11: *APN "A2" traffic flows [Byte]*

APN "A1", APN "A2" and APN "B" can be used to model mIoT traffic. It can be observed that for APN "A1" and APN "A2", the packet size distribution can be derived from a single value (50 Bytes packets – Figure 3.10 and 3.11) with a good approximation. APN "B" is a bit different from this pattern (Figure 3.12); there are significant numbers of 100 Bytes packets besides the 50 Bytes packets. For APN "A2" and APN "B", the number of packets that create flow are between 0 and 5. The APN "A1" traffic is slightly different here, while the 5-packets flows are dominating the other two APNs' traffic – APN "A2" and APN "B" –, here the 10 and 40 packet flows are also significant.

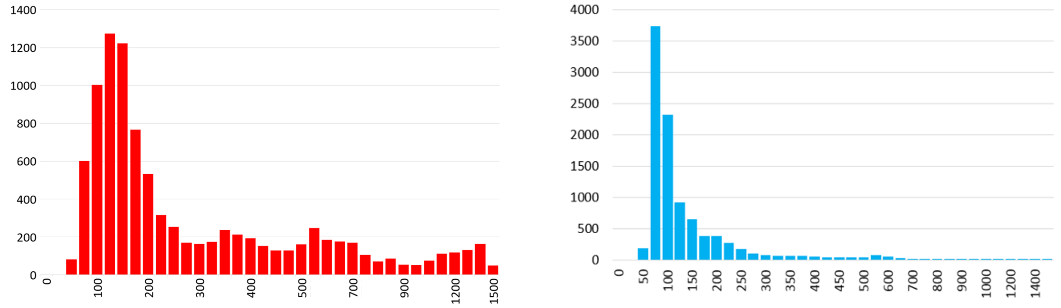
APN "C" traffic is used to model eMBB traffic (Figure 3.13). You can see that traffic on this APN cannot be modeled as easily as traffic on APN "A1", "A2" and "B" (altogether: on mIoT). Here, the packet sizes that create the traffic are varied, unlike in the case of mIoT. For simpler modeling, I have identified the essential values of the packet size distribution and defined the modes (150, 375, 600, 1400 Byte). Based on these values, I have developed a traffic modeling function that seems to model the live network traffic according to my measurements accurately. However, it should be noted that this method may not always give an accurate approximation of live network traffic.

The packet distribution of traffic models simulating traffic patterns of different APNs are shown in Table 3.1 and the flows' packet number distribution are summarized in Table 3.2. Based on these values, I create an active measurement methodology and traffic scenarios for our testbed architecture, which will be discussed in Chapter 4.



(a) downlink packet size distribution **(b)** uplink packet size distribution

Figure 3.12: *APN "B" traffic flows [Byte]*



(a) downlink packet size distribution **(b)** uplink packet size distribution

Figure 3.13: *APN "C" traffic flows [Byte]*

APN name	Packet size	Percentages
A1	50 Byte	89.10%
	75 Byte	5.5%
	100 Byte	5.4%
A2	50 Byte	84.62%
	75 Byte	15.38%
B	50 Byte	44.99%
	100 Byte	40.45%
	150 Byte	9.23%
C	150 Byte	69.56%
	375 Byte	13.03%
	600 Byte	10.8%
	1400 Byte	6.61%

Table 3.1: *Flows packet size distribution on different APNs*

APN name	**Number of packets**	**Percentages**
A1	5	94.10%
A2	5	44.43%
	20	24.75%
	40	25.48%
B	5	92.58%
C	5	62.50%
	10	14.19%
	20	17.74%

Table 3.2: *Flows packet number distribution on different APNs*

The measurement set-up consists of two parts in terms of traffic. One is the background traffic, which aims to model the aggregated network traffic. The other part is the concept of a single user or group of users traffic. The amount of packets or bytes in a flow is irrelevant in defining background traffic, only the distribution of packet sizes and the total number of bytes are the important parameters. To describe the background traffic, I primarily considered traffic through the general network. However, I model the packet distribution composition of background traffic with APN "C"'s traffic as eMMB application-type traffic. This assumption may be correct, because, based on our previous work [7], we have found that current machine-to-machine communication is negligible *in volume* compared to human communication. Furthermore, predominant application-type and a significant portion of network traffic will be eMBB traffic. Moreover, the data packet distribution shown in Figure 3.8 is correct, even if there is only minimal bandwidth available through our network. This allows us to produce arbitrary bandwidth for network equipment or network section testing.

Chapter 4

Simulating traffic patterns

Earlier, I have shown the cellular networks main characteristics through two approaches. In Chapter 3, I have presented the packet distributions of the 4 examined APNs based on a passive measurment method. Then, I have shown what kinds of flow parameter value ranges can describe a given networking application – and its user experience. In this chapter, I examine how the user experience is affected in the different use-case scenarios by varying background traffic on a live 5G test network. To execute that, I create an active measurement set-up and generate the traffic flows based on the 4 different APNs traffic patterns.

4.1 Testbed architecture

While building the testbed, I aimed to create a setup that resembles to a live network architecture as much as possible. The layout is shown in Figure 4.1. Background traffic was generated by one device (UE2), and measurements were made with another client device (UE1). The UE1 is connected to the measurement server on one side and the EPC on the other side via 4G and 5G base stations [26]. The EPC was then connected directly to the measurement server. It should be mentioned that the traffic in this set-up followed the 3GPPP reference model 3.X [25], the traffic from the UE to the EPC used 4G devices, whereas downlink from the EPC to the UE used 5G network.

The UE2 generated the background traffic and used wired connection. It is connected to the background traffic generator server's client-side interface with a wired connection. UE2 is wirelessly connected to eNB and gNB, then it accessed the background traffic generator server, server-side interface via EPC.

The wired connections were 10 Gbps Ethernet in the testbed, which is similar to a live network. This way our measurement can offer a good approximation to a live network scenario. With this architecture, those active parts of the network are examined, which influence the user experience the most. Measurements were carried out in a closed box, with

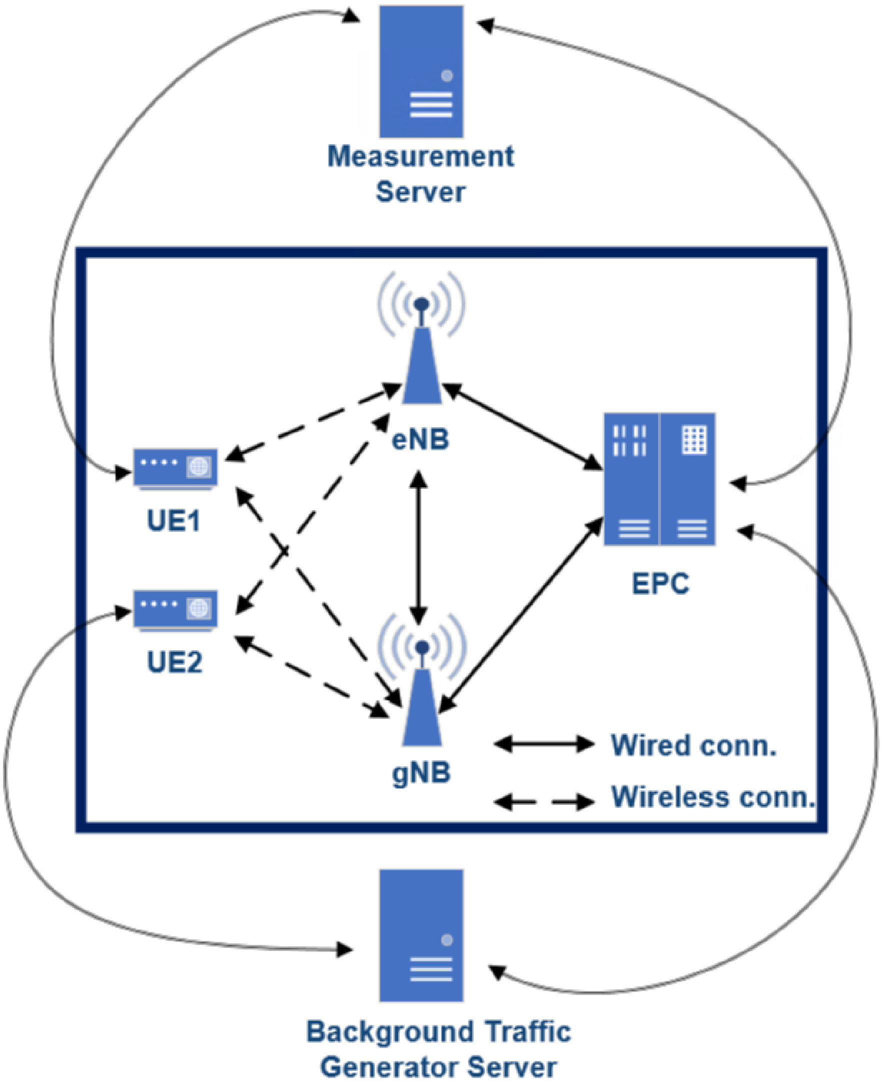


Figure 4.1: *The architecture of the 5G testbed – as Option 3X*

set signal levels, under stable radio conditions for the measurement period – as suggested by [27], [28].

4.1.1 Direction of the examined traffic

Measuring one-way delay

In this setup the measurement traffic flows from the EPC interface of the measurement server to the EPC and then through the gNB via a wireless 5G connection to the UE1. Then, on the wired interface of UE1 there is a connection to the client-side interface of the measurement server.

Measuring Round Trip Time

In this setup the measured traffic flows from the UE-side interface of the measurement server to UE1 and then from UE1 to the eNB on a 4G radio access network. It accesses the measurement server from the eNB via the EPC via a wired connection. The measurement server records the arrival of packets and returns the data stream on its wired interface to the EPC. Passing through the EPC, it reaches gNB and then accesses the UE1 client device from the gNB via a 5G wireless connection. Finally, on the UE1 wired connection, it moves to the measurement server.

4.1.2 Measurement equipment

Measurement and Background traffic generator servers were x86-based machines with 12-core i7 processors, and had 8-8 GB of memory. The network interfaces were Intel cards capable of real 2x1 Gbps Ethernet throughput on the client-side, while the server-side had 2x10 Gbps optical connections. The server traffic was provided by a software package based on Ubuntu 18.04. The major measurement, background traffic generator, and monitoring programs were the following: NMAP software packages [29], Iperf [30], Cisco T-rex [31], Ethernet, and Smooth ping. We used dedicated processor mapping and kernel compiled programs for the smooth running of the software. The measurement results were recorded for traceability purposes.

The EPC is a 2x64 x86 server with 124 GB of memory and has a 2x10 Gbps LAG connection in all wired directions, towards gNB and eNB. The mobile network software consisted of 4G MME-SGW-PGW with option 3.X architecture [32], where I implemented a minimal architecture design with only having the S6a interface outside of this. The system did not include any interface other than the minimal design.

The eNB and gNB were running in separate HWs, with an x86-based architecture, dedicated with a 2x24 core processor and 8 GB of RAM. The eNB was responsible for the 4G network, for the uplink channel and the controlling S1AP interface. It operated on 1.8 GHz band with 10 MHz bandwidth in 2x3 MIMO mode. The gNB operated on 3.6 GHz with 100 MHz bandwidth, but due to software limitations, it was safely used only in 2x2 MIMO mode. Clients were 2x3 MIMO mode devices conforming to the Option 3.X architecture and were simultaneously connected to 4G and 5G networks. It was supported by a multi-core processor and several GBs of RAM.

4.2 Results

The main parameters of the 5G testbed have been defined through a series of measurements, and for reference comparison we have carried out similar measurements on 4G. The aggregated measurement results (Table 4.1) show that for 5G, even 1 UE downlink peak rate heavily outperforms the 4G values. However, 5G uplink transmission is an exception as it has similar performance as 4G. As expected, latency is also significantly lower at 5G than at 4G.

	4G- Cat 3	**4G- Cat-12**	**5G - 1 UE**	**5G - 2 UE**
Downlink Peak Rate	75 Mbps	420 Mbps	885 Mbps	1465 Mbps
Uplink Peak Rate	22 Mbps	87 Mbps	92 Mbps	91 Mbps
Latency (one-way)	13 ms	12 ms	3.71 ms	4.96 ms
Packet Error Rate	0.3%	0.2%	0.39%	0.67%

Table 4.1: *Aggregated (median) results of measurements*

The traffic of different application types was measured with different amounts of background traffic. The measurement cases are as follows:

- NO background traffic,
- 25% – of the available bandwidth – background traffic,
- 50% background traffic (to measure 1 direction latency),
- 100% background traffic.

4.2.1 Latency results

A key promise of 5G is that latency experienced with 4G will be reduced drastically to around 1 ms. Furthermore, latency can be kept low even if the traffic load is high in the network.

To measure 5G network latency I examined how the DoNUT affect latency deviation. It was tested with 100 Byte and 1000 Byte packets. During the tests, packet-sending frequency were between 10 ms and 1000 ms, in 4 different scenarios. After these tests we measured latency with continuous background traffic, with its volume being approximately 50% of the network's capacity.

One-way latency results are depicted by Figure 4.2. Figure 4.3 shows latency results when the background traffic took 50% of the link capacity. Again, Table 4.1 shows the aggregated results of the measurements.

As Figure 4.2 shows, the minimum values for the latency were relatively low, at all sending rates. Furthermore, the packet size does not really affect the minimum latency. The variance of the delay showed similar values for all the 8 measurement scenarios. In Figure Figure 4.3, where the measurement was taken with different transmission intervals and background traffic, the maximum value of the latency has increased when compared to Figure 4.2. It is especially interesting that the latency of sending bigger, 1000B packets at each 100ms shows smaller variance than the latency for 100B packets at the same rate. This shows us that the DoNUT can handle the bigger-sized packets more smoothly.

Figure 4.4 depicts our measurement results on sending different-sized packet-flows sent at different rates, where the background traffic was set to 25%. When taking a closer look on the maximum values of the packet-flow lengths, it seems that the network can handle the middle-sized flows smoothly. Interestingly the mean values of the latency did not change significantly throughout this scenario.

Figure 4.5 and Figure 4.6 shows measurement results where significant, 50% and 100% background traffic was put on the cell, respectively. There are various, interesting conclusions can be drawn by comparing the two figures. It is clear that the advantage of 5G networks are visible here: small packet flows are handled by the network with low latency. While the variance was low, latency sometimes reaches those very low, expected values of 5G - such as 3-4ms. This does not increase much at the case of 100% background traffic, either. Figure 4.7 differs from these, because there is no background traffic injected. The latency sometimes almost reached 1 s, but this was always due to the very first packet of the flow, which needed the radio channel to set up first.

From Figure 4.8 to 4.11 show the user experience of "A1", "A2" and "B" type APN-s. Since these traffic types are more sterile – merely two kinds of packet sizes seem to appear –, their results do not differ that much. Minimum values are close to constant. The average values of latency grow when the cell is under high (background) load, although this growth is not extremely high; should can be handled well by the applications.

Surprisingly, the observed variance with 0 background traffic is much more significant. One possible explanation is that improving load balancing is one of the main goals of 5G. Because the examined 5G architecture is an early test deployment, this effect may change in the future. However, the packet numbers that make up the flows do not significantly affect the latency values; differences can be detected, but the measurement results do not show a clear pattern.

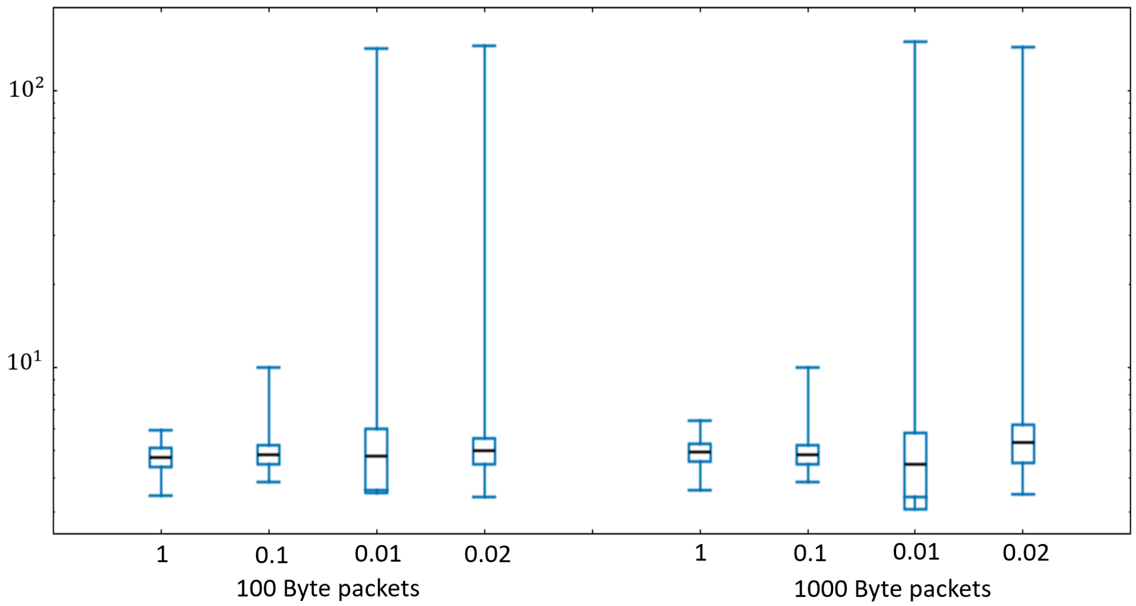
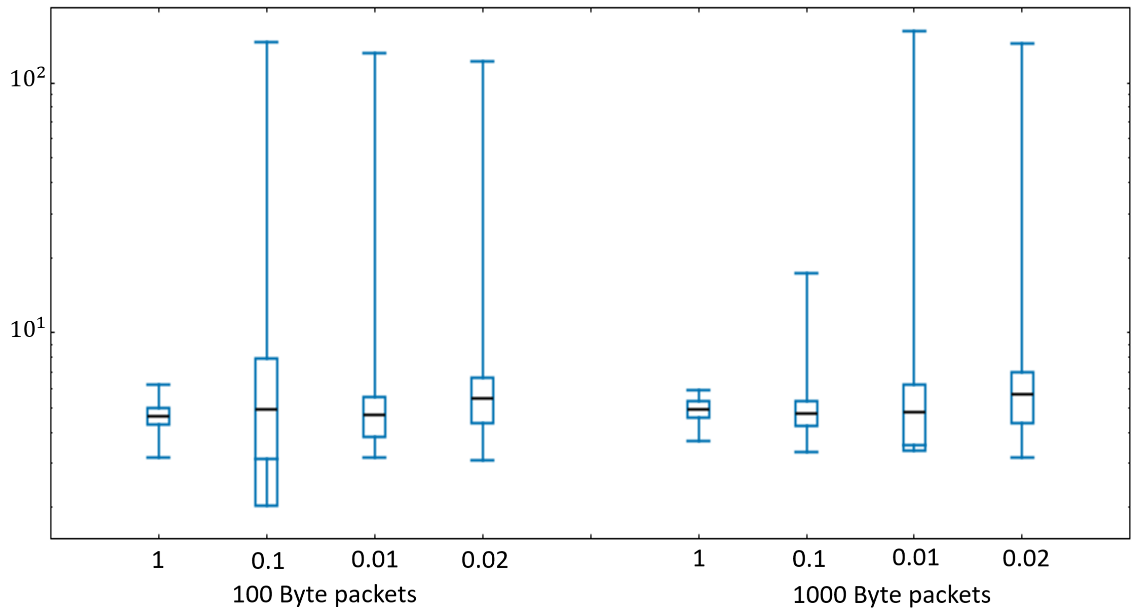


Figure 4.2: *Latency of 5G links, measured with different transmission intervals [ms] – NO background traffic*



4.2

Figure 4.3: *Latency of 5G links, measured with different transmission interval [ms] and background traffic*

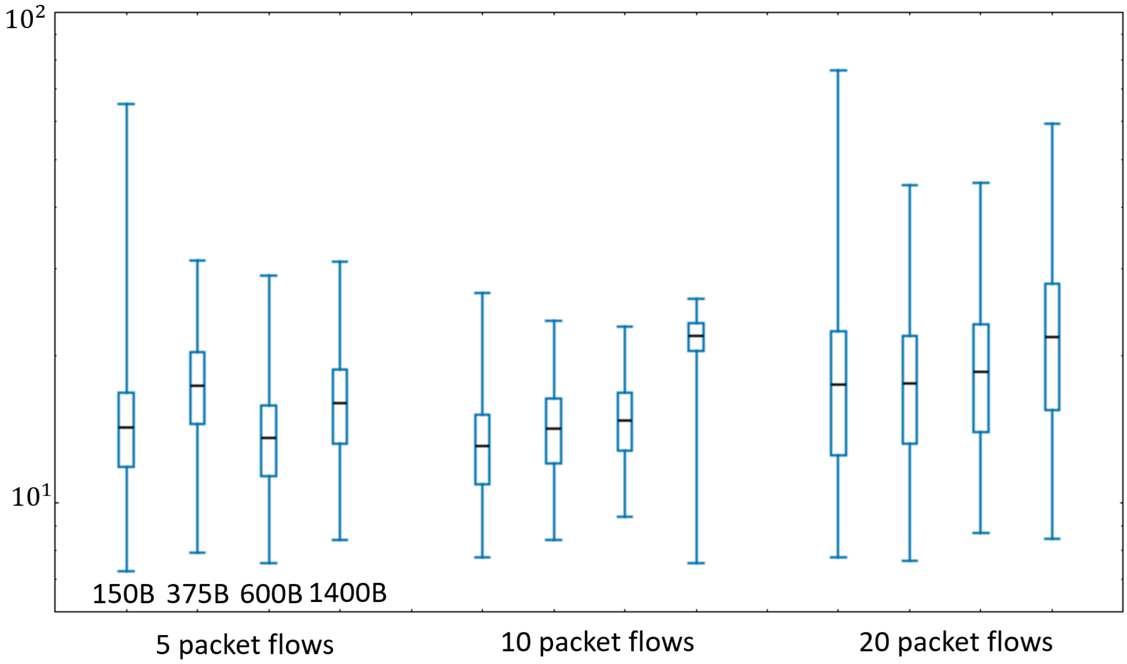


Figure 4.4: *APN "C" latency results with different size and number of packets [ms] (25% background traffic)*

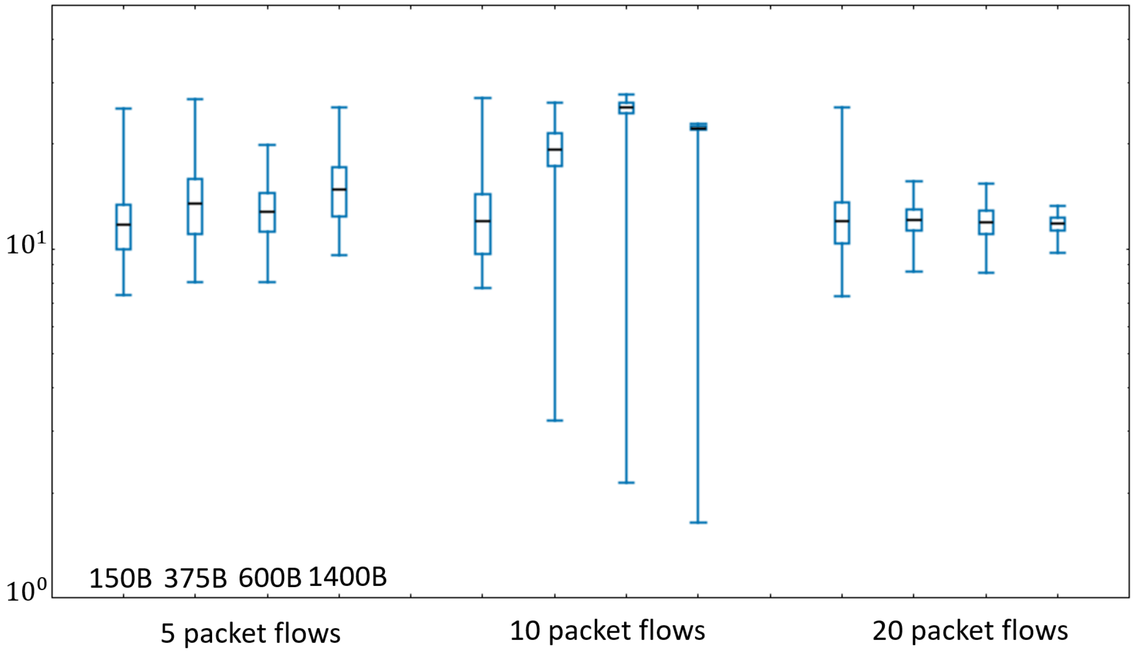


Figure 4.5: *APN "C" latency results with different size and number of packets [ms] (50% background traffic)*

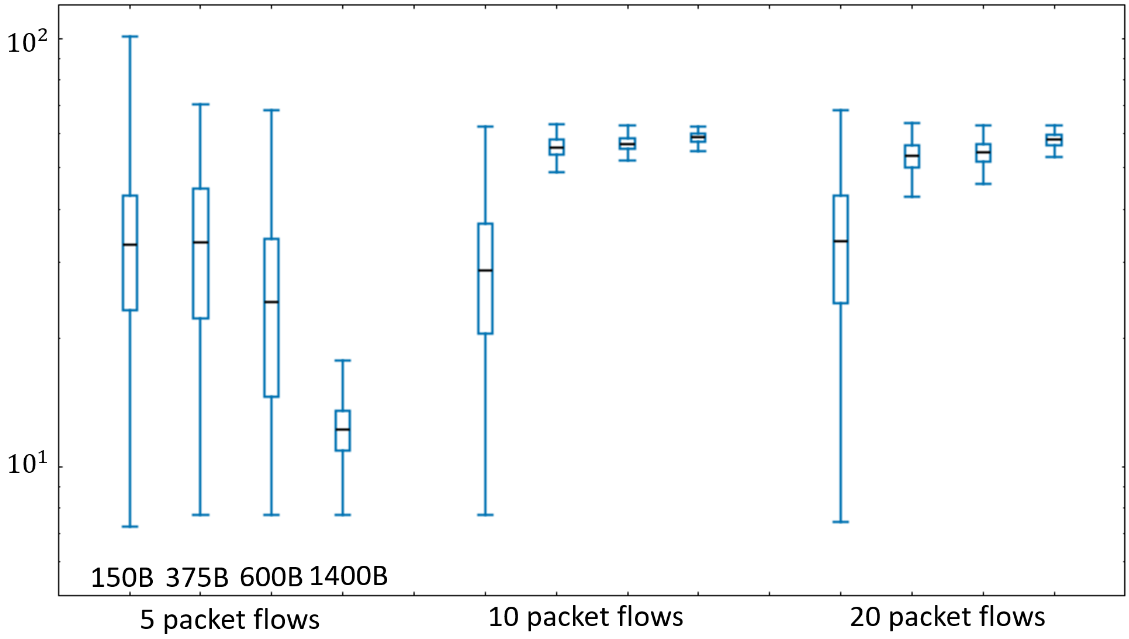


Figure 4.6: *APN "C" latency results with different size and number of packets [ms] (100% background traffic)*

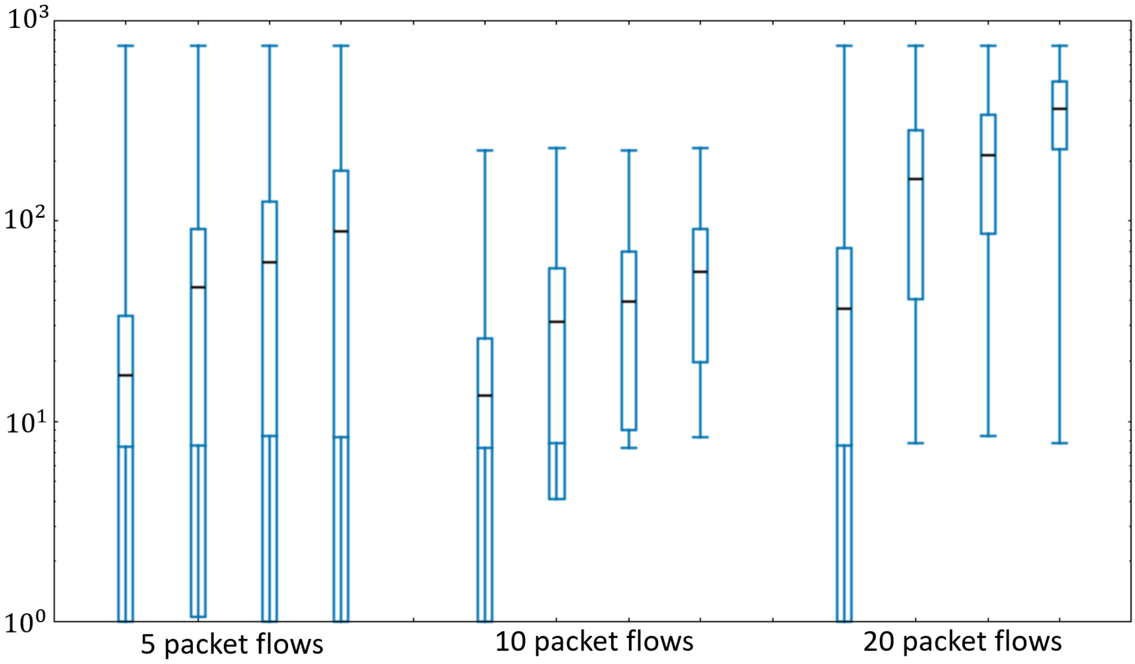


Figure 4.7: *APN "C" latency results with different size and number of packets [ms] (0 background traffic)*

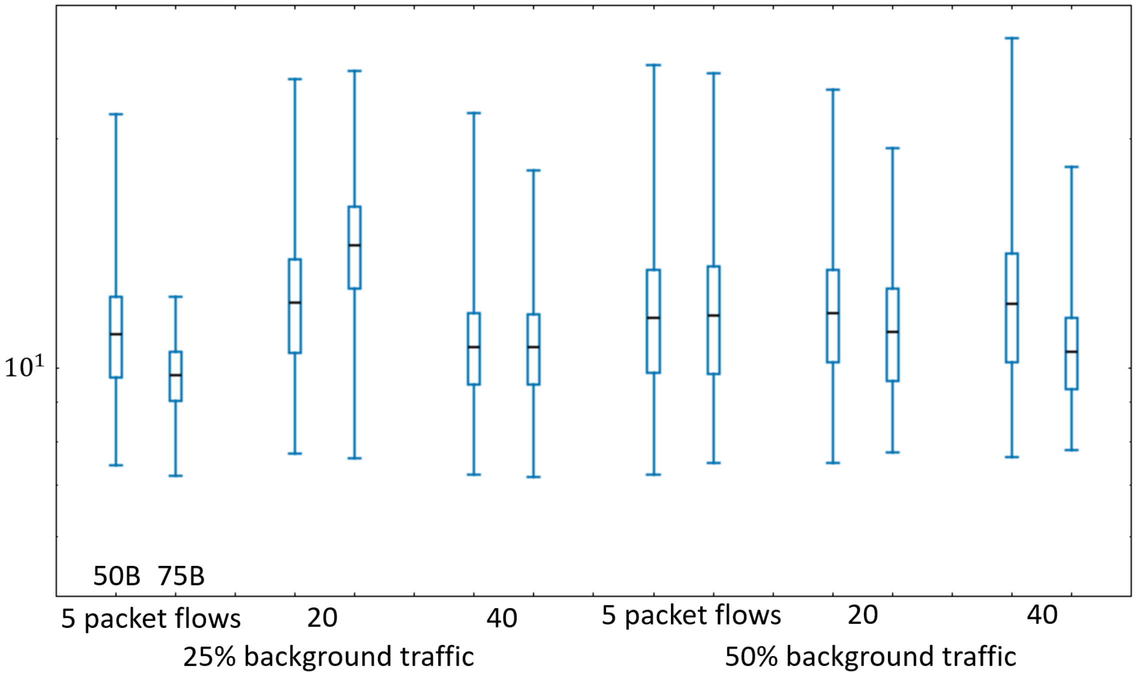


Figure 4.8: *APN "A1" latency results with different size and number of packets [ms] (25% and 50% background traffic)*

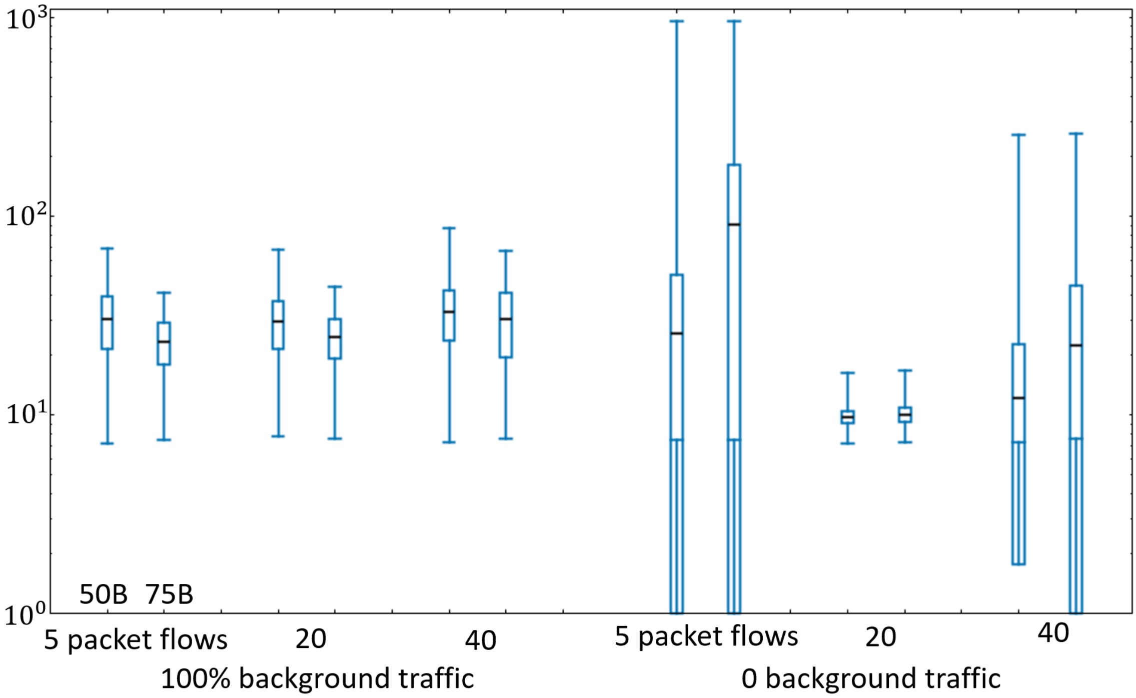


Figure 4.9: *APN "A1" latency results with different size and number of packets [ms] (100% and 0 background traffic)*

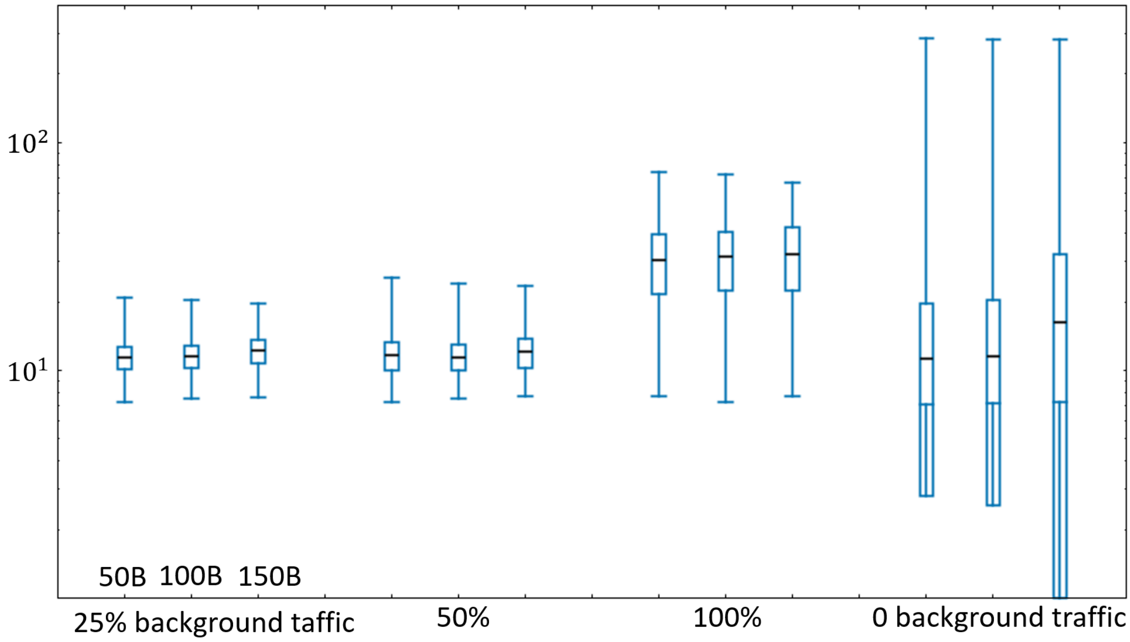


Figure 4.10: *APN "B" latency results with different size and number of packets [ms] (25%, 50%, 100% and 0 background traffic)*

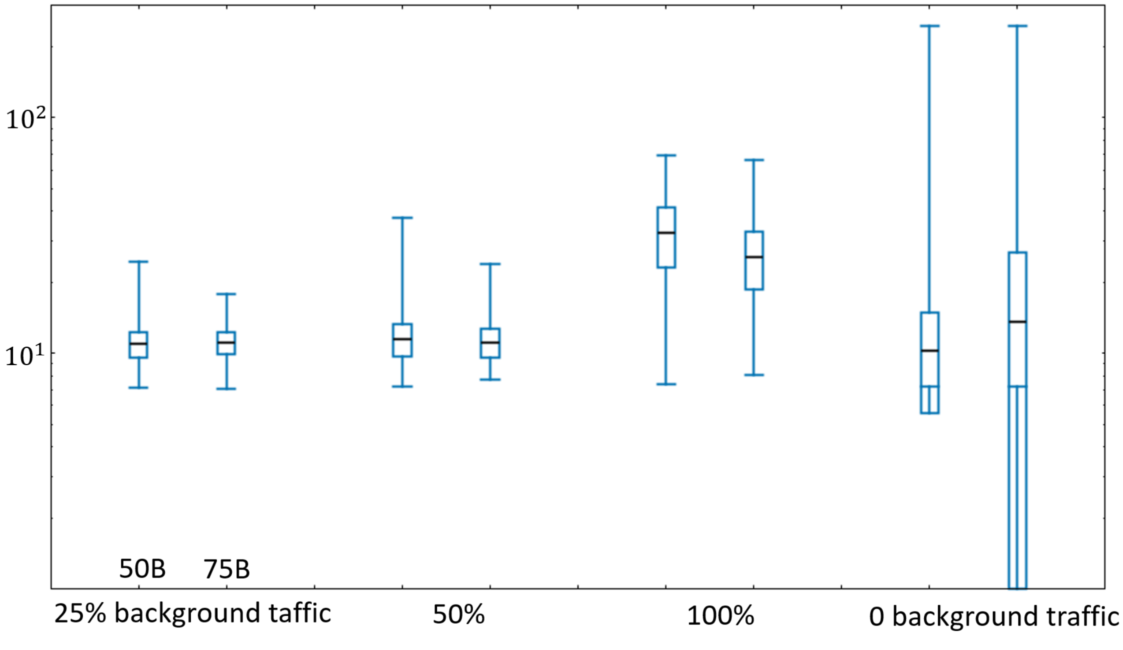


Figure 4.11: *APN "A2" latency results with different size and number of packets [ms] (25%, 50%, 100% and NO background traffic)*

4.2.2 Throughput

Higher data transmission rate is achieved with increasing bandwidth, different coding and better modulation. In downlink direction even 100 MHz bandwidth can be used; the theoretical throughput limit is 1.5 Gbps. Naturally, transmission rate can be tuned with the number of receiver and transceiver antennas as well.

4.2.3 Packet Error Rate (or data loss)

The data connection layer of the radio channel should correct the transmission errors if packet loss has occurred. We measured Packet Error Rate at a higher layer, so we detected PER only when lower layers were not able to restore packets.

4.3 Jitter results

In this measurement setup I were curious about the variance of the packet inter-arrival time during the transmission. The examined packets were 100 Bytes and 1000 Bytes long and also in 2 different frequency (sending-rate) scenarios between 1 and 0.1 seconds. Figure 4.12 shows the received packets interarrival times with 0.1 and 1 s packet sending periods (in-between packets). The difference from the expected theoretical results is minimal. It is worth to mention that at 0.1 s packet sending period some packets arrive earlier occasionally. The reason could be the variance of transmission path latency. When examining the results of bigger packets' transmission at Figure 4.13 shows how 1000 Byte packets jitter changed during test period. The received packets timestamp mostly differ from reference in the negative direction. Specific channel allocation could be a reason for that. As a part of the 1000 Byte was transferred with the previous packet, and the rest of them later.

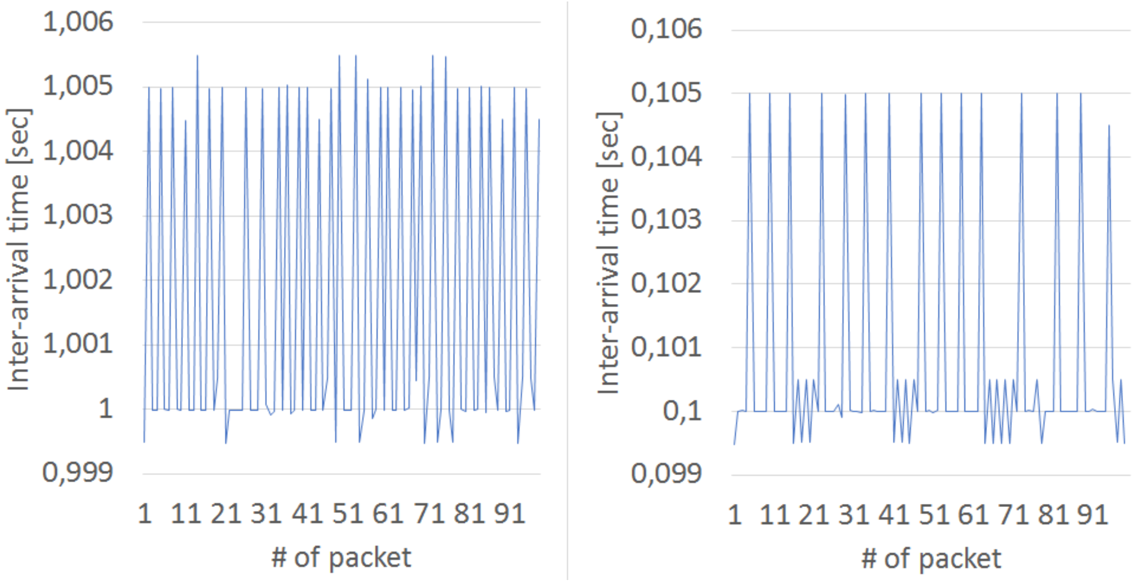


Figure 4.12: *Jitter results (left): 100 Byte packets, 1000 ms sending period, (right): 100 Byte packets, 100 ms sending period*

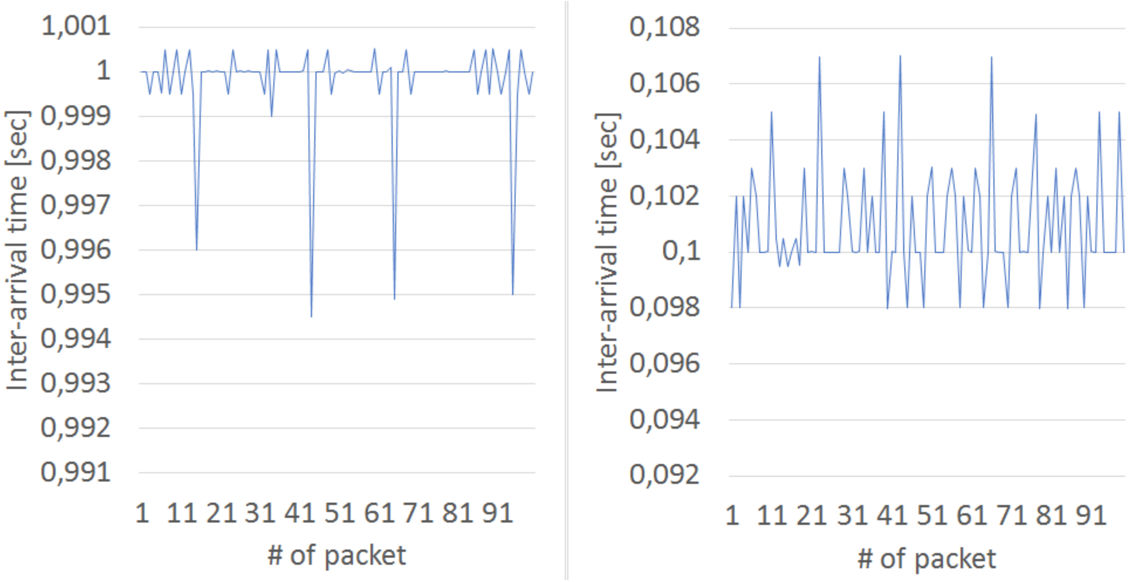


Figure 4.13: *Jitter results (left): 1000 Byte packets, 1000 ms sending period, (right): 1000 Byte packets, 100 ms sending period*

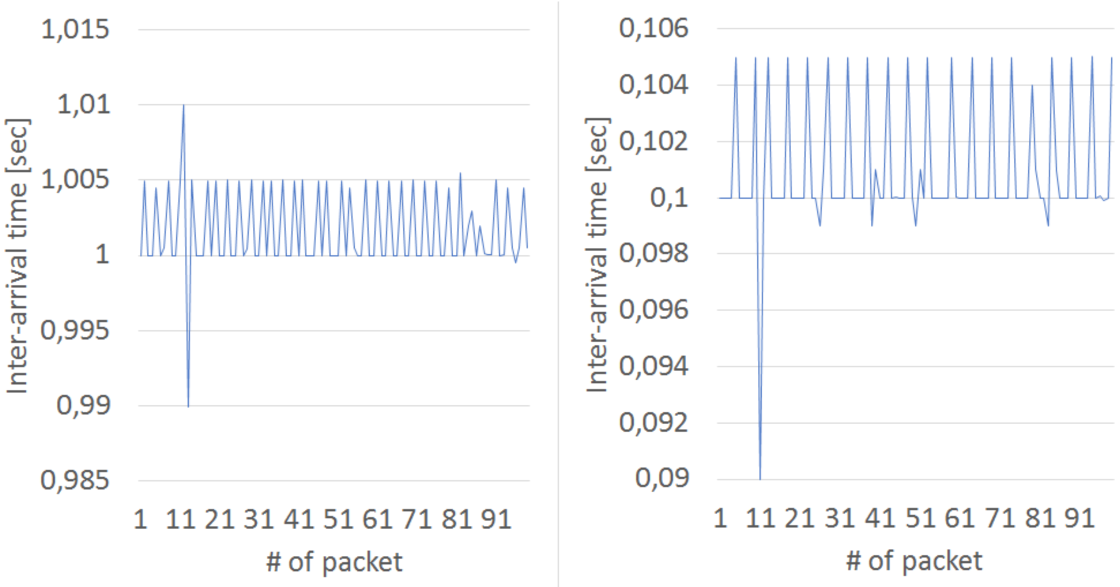


Figure 4.14: *Jitter results (left): 100 Byte packets, 1000 ms sending period, (right): 100 Byte packets, 100 ms sending period with background traffic*

4.3.1 Jitter results with background traffic

During these measurements I wanted to examine how packet inter-arrival times vary in case of a loaded cell. Based on the previous measurements I knew that cell capacity is around 1400 Mbps, when PER remains negligible, so I transmitted some background traffic during our latency measurement. The results are as follows.

In the Figure 4.14 there was 50% Mbps background traffic. When we compare these results with Figure 4.12, there is no considerable difference, except a few packets.

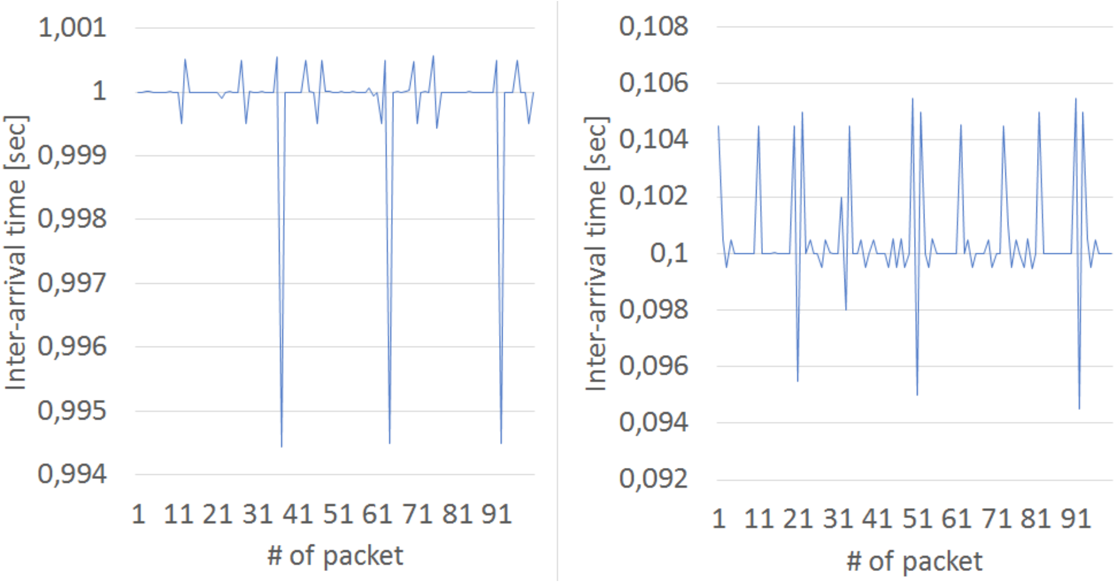


Figure 4.15: *Jitter results (left): 1000 Byte packets, 1000 ms sending period, (right): 1000 Byte packets, 100 ms sending period with background traffic*

The 1000 Byte packets and the background traffic's correlated effect can be seen on Figure 4.15. The results show that latency is not changed drastically but the effect of cell load is perceivable.

Chapter 5

Summary

An important part of my work is the refinement of the known methodology for interpreting network traffic not only as a sterile stream of packets of the same size and frequency but also for its periodicity and packet size diversity. As part of this refinement, besides the packet size distribution, I have defined the frequency and the distribution of the data packet sets of a group of users. This correlates with the user experience of the subscribers for given application types. Describing this refined methodology, I have isolated, tested, and modeled user traffic similar to 5G use-cases. As a practical result, different application-types – including those eMBB, cIoT and MIoT that are distinguished for 5G – can be described using this method. The main parameters of the different use-cases were recorded and then fitted to the 5G network to investigate network effects. For the work, I got partly ready HW and SW elements:

- The mass user network management system for DPI measurements;
- 5G HW and SW equipments.

While analyzing the traffic of different application types, I have seen that applications used by people tend to use medium to large packets, where the up- and downlink directions show significantly different characteristics. The access network utilized merely by machines usually uses small packet sizes with minimal or zero variance, for the up- and downlink, as well.

Based on this, we can conclude that our network should rather be optimized for fast transfer of small numbers of packets when the packet is received from the user – uplink. Most of the packets going to the user – downlink – are also small but there may be packets of a larger size and frequency. I used these results to model general background traffic in the network. Furthermore, it can be known by experience and seen from the measurement results that the network transmits larger packets more easily, more efficiently.

Limitations during the measurements were included in both the User Equipment, RAN, and EPC network while vendors running early-released software. Nevertheless, I received from the vendors much more support than I expected for the analyses. It was evident that there is still room for improvement at the software level on the 5G RAN and related networks. Note, that by using an experimental software, I was able to achieve much more excellent results in terms of service guarantees than those published in the current paper. However, they are not published here because such experimental work has not always been achieved during standard operating mode. Furthermore, it was evident that the various client-RAN-Core software did not always co-operate well with each other, thus significantly reducing the reliability of such measurement results.

As I mentioned in the introduction, the documentation for use-case traffic appearing in 5G is still quite incomplete, especially if one is interested in live network measurement results. There are various opportunities for me to take my measurements further in the future. First of all, if the 5G deployment reference model 3.X will be changed, where the uplink channel will be also based on 5G architecture, then my analysis can expanded. It may be worth examining the effects of network slicing by making measurements and comparisons between vendors. I have already identified traffic patterns for the main 5G application types, and further predictions also can be made. It is worth to examine what will change in these trends in the future, and measure their impact on new generation networks.

Acknowledgement

I would like to render my sincere thanks to my supervisor, Dr. Pál Varga for his guidance and suggestions. I would also wish to express my gratitude to PhD student Gábor Soós for his expert advice which has been invaluable throughout all stages of work. Last but not least many thanks to Rebeka Hordós and Dorottya Balogh for review comments and their valuable suggestion regarding ways to improve this paper.

Bibliography

[1] Gartner, “5 trends appear on the gartner hype cycle for emerging technologies, 2019.” https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/, 2019.

[2] GSMA Intelligence, “Global mobile trends what’s driving the mobile industry?.” https://www.gsmaintelligence.com/research/?file=8535289e1005eb248a54069d82ceb824download, 2018.

[3] Cisco, “Cisco visual networking index: Global mobile data traffic forecast update.” https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.pdf, 2017.

[4] Ericsson, “Internet of Things forecast,” tech. rep., Ericsson, 2017.

[5] GSMA Intelligence, “Road to 5g: Introduction and migration.” https://www.gsma.com/futurenetworks/wp-content/uploads/2018/04/Road-to-5G-Introduction-and-Migration_FINAL.pdf, 2018.

[6] 3rd Generation Partnership Project, “Summary of Rel-15 Work Items, TR 21.915 version 0.5.0 Release 15,” version 15, 3GPP, 2018-12.

[7] G. Soós, D. Ficzere, and P. Varga, “User Group Behavioural Pattern in a Cellular Mobile Network for 5G Use-cases,” in *IEEE/IFIP Network Operations and Management Symposium*, - submitted - 2020.

[8] M. Slanina, L. Klozar, and S. Hanus, “Practical measurement of data throughput in LTE network depending on physical layer parameters,” in *2014 24th International Conference Radioelektronika*, pp. 1–4, IEEE, 2014.

[9] 3GPP TS 32.295, “Charging management; Charging Data Record (CDR) transfer,” version 15, 3rd Generation Partnership Project, 2008-06.

[10] IANA, “Internet assigned numbers authority.” https://www.iana.org/, 2019.

[11] Cisco, “Netflow 9 - Technical Support and Documentation.” Cisco IOS NetFlow, 2019.

[12] X. Lu, W. Cao, X. Huang, F. Huang, L. he, W. Yang, S. Wang, X. Zhang, and C. Hongsong, “A real implementation of dpi in 3g network,” pp. 1–5, 12 2010.

[13] G. Soós, F. N. Janky, and P. Varga, “Distinguishing 5G IoT Use-Cases through Analyzing Signaling Traffic Characteristics,” in *42nd International Conference on Telecommunications and Signal Processing (TSP)*, pp. 562–565, 2019.

[14] Q. Zhang and F. H. Fitzek, “Mission critical IoT communication in 5G,” in *Future Access Enablers of Ubiquitous and Intelligent Infrastructures*, pp. 35–41, Springer, 2015.

[15] 3GPP TS 21.101, “Technical Specifications and Technical Reports for a GERAN-based 3GPP system,” version 13, 3rd Generation Partnership Project, 2015-12.

[16] Etsi, “Why do we need 5g?.” https://www.etsi.org/technologies/5g, 2019.

[17] 3GPP TS 23.501 version 15.2.0 Release 15, “5G; System Architecture for the 5G System,” version 15, 3rd Generation Partnership Project, 2018-06.

[18] 3rd Generation Partnership Project, “Evolved Universal Terrestrial Radio Access Network (E-UTRAN); S1 Application Protocol (S1AP),” 36.413 release 8, 3GPP, 2016.

[19] S. Molnár, Z. Móczár, and B. Sonkoly, “How to transfer flows efficiently via the internet?,” in *2014 International Conference on Computing, Networking and Communications (ICNC)*, pp. 462–466, Feb 2014.

[20] S. Molnár, P. Megyesi, and G. Szabo, “Multi-functional traffic generation framework based on accurate user behavior emulation,” in *2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 13–14, April 2013.

[21] D. Kozma, G. Soós, D. Ficzere, and P. Varga, “Communication Challenges and Solutions between Heterogeneous Industrial IoT Systems,” in *1st International Workshop on Analytics for Service and Application Management (AnServApp 2019), together with IFIP/IEEE CNSM 2019*, 2019.

[22] J. Chung, Y. Won, B. Park, and J. Won-Ki Hong, “Mass-count disparity in mobile traffic,” *IEEE Communications Letters*, vol. 20, pp. 65–68, Jan 2016.

[23] G. Soós, D. Ficzere, and P. Varga, “A 5G Testbed and its Initial KPI Measurements,” in *IEEE/IFIP Network Operations and Management Symposium*, - submitted - 2020.

[24] H. Matsue, S. Kubota, H. Hojo, K. Watanabe, T. Saito, S. Aikawa, and A. Sato, “Future systems and technologies for broadband wireless access services,” pp. 156 – 157, 11 2003.

[25] GTI, Ericsson, ZTE, Nokia, “5G Network Architecture White Paper.” https://www.3gpp.org/keywords-acronyms/1612-ue-category, 2018-02.

[26] K. Kondepu, F. Giannone, S. Vural, B. Riemer, P. Castoldi, and L. Valcarenghi, “Experimental Demonstration of 5G Virtual EPC Recovery in Federated Testbeds,” in *IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, pp. 712–713, 2019.

[27] A. Shorov, “5g testbed development for network slicing evaluation,” in *2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pp. 39–44, Jan 2019.

[28] K. Kousias, K. Katsalis, D. Stavropoulos, T. Korakis, and L. Tassiulas, “Building virtual 802.11 testbeds towards open 5g experimentation,” in *2016 IEEE Wireless Communications and Networking Conference*, pp. 1–7, April 2016.

[29] Nmap.org, “Nmap: the Network Mapper - Free Security Scanner.” https://nmap.org, 2019.

[30] iPerf, *iPerf - The ultimate speed test tool for TCP, UDP and SCTP.*

[31] Cisco.com, “TRex - Realistic traffic generator .” https://trex-tgn.cisco.com/, 2019.

[32] D. Lake, G. Foster, S. Vural, Y. Rahulan, B. Oh, N. Wang, and R. Tafazolli, “Virtualising and orchestrating a 5g evolved packet core network,” in *2017 IEEE Conference on Network Softwarization (NetSoft)*, pp. 1–5, July 2017.