5G standalone network's reliability, one-way latency and packet loss rate analysis for URLLC implementation

Master of Science Thesis
University of Turku
Department of Computing
Information and Communication Technology
2022
Ville Ritola

Supervisor:
Jyri Lamminmäki (Nokia)

Examiners:
Petri Sainio (University of Turku)
Jouni Isoaho (University of Turku)

UNIVERSITY OF TURKU
Department of Computing

Ville Ritola: 5G standalone network's reliability, one-way latency and packet loss rate analysis for URLLC implementation

Master of Science Thesis, 76 p.
Information and Communication Technology
July 2022

5G is the fifth generation technology standard for cellular networks. It has three main application demands, which are Enhanced Mobile Broadband (EMBB), Massive Machine-Type Communications (MMTC) and Ultra-Reliable Low-Latency Communications (URLLC). URLLC is a very challenging demand to implement, with strict reliability and latency requirements. It has been highly specified by 2022 and 5G vendors are starting to implement basic URLLC features in the near future.

The motivation for this thesis is to find ways to make measurements on how a 5G standalone (SA) network performs on key URLLC performance indicators, analyse and visualize these measurements, find reasons for certain network behavior and make estimates on what kind of impact different URLLC features will have when implemented. Furthermore, another motivation is to find a way to detect packet loss and reasons behind it, because packet loss impairs reliability significantly and should be minimized before deploying URLLC features.

To measure 5G SA network's performance, four different kind of test cases were identified, in which URLLC type of network traffic is generated. There are static tests done in good coverage and bad coverage from the 5G cell, and mobility tests done by moving from good coverage to bad coverage while attached to the same 5G cell, and a handover test in which the 5G cell is changed. All tests are done in a 5G field verification environment, for both downlink and uplink.

For downlink, coverage and mobility inside a cell did not have a meaningful impact to one-way latency. This was mainly because there was no need for packet retransmissions, which would have increased latency. This is promising especially for mobility URLLC use cases such as Vehicle-To-Everything communications (V2X). Uplink performed much weaker, mainly because of uplink resource scheduling and packet retransmissions. Handover was problematic for both downlink and uplink, because of the brief but massive increase in latency caused by the cell change.

All packet loss in the measurements happened in uplink transmission, and this thesis includes a case study where different potential factors causing packet loss were consistently eliminated. In the end, the cause for packet loss indicates towards the 5G chipset used for the tests.

Keywords: 5G, New Radio, URLLC, reliability, latency, packet loss

Contents

Abbreviations

3GPP	3rd Generation Partnership Project
5GC	5G Core
5GQI	5G Quality of Service Identifier
ACK	Acknowledgement
AMF	Access and Mobility Management Function
BLER	Block Error Rate
CCDF	Complementary Cumulative Distribution Function
cmW	Centimeter Wave
CQI	Channel Quality Indicator
DAPS	Dual Active Protocol Stack
dB	Decibel
DL	Downlink
DNS	Domain Name System
EMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FDD	Frequency Division Duplex
FR	Frequency Range
gNB	gNodeB
GUI	Graphical User Interface
HARQ	Hybrid Automatic Repeat ReQuest
IoT	Internet of Things
ISD	Inter-site distance
KPI	Key Performance Indicator
LDPC	Low-Density Parity Check
LTE	Long Term Evolution

MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input and Multiple-Output
MMTC	Massive Machine-Type Communications
mmW	Millimeter Wave
MS	Millisecond
NR	New Radio
NSA	Non-Standalone
NSSAI	Network Slice Selection Assistance Information
NSSF	Network Slice Selection Function
OFDM	Orthogonal Frequency Division Multiplexing
PCAP	Packet Capture
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PHY	Physical layer
PPS	Packets Per Second
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
QoS	Quality of Service
RBG	Resource Block Group
RLC	Radio Link Control
RRC	Radio Resource Control
RRH	Remote Radio Head
RSRP	Reference Signal Received Power
RSU	Road-Side Unit
SA	Standalone
SCS	Subcarrier spacing
SDAP	Service Data Adaptation Protocol

SDU	Service Data Unit
SINR	Signal to Interference and Noise Ratio
TDD	Time Division Duplex
TRxP	Transmission Reception Point
TTI	Transmission Time Interval
UE	User Equipment
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communications
V2X	Vehicle-To-Everything
VoIP	Voice over IP

1 Introduction

5G is the fifth generation technology standard for cellular networks. URLLC (Ultra-Reliable Low-Latency Communications) is one of 5G's main demands, with strict reliability and latency requirements.

3GPP has specified URLLC to very high level by release 17, and 5G vendors are starting to implement the basic URLLC features in 2022. These basic features increase reliability and reduce latency with techniques used in frame structure, scheduling, link control and gNB processing time. Based on that, it is necessary to find out on what level the current 5G standalone network's reliability and latency are, how much there is packet loss and what can be done to prevent it.

Therefore, the motivation for this thesis is to find ways to make measurements on how a 5G standalone network performs on key URLLC performance indicators, analyse and visualize these measurements, find reasons for certain network behavior and make estimates on what kind of impact different URLLC features will have when implemented. For this, a 5G field verification environment will be used, in which URLLC type of network traffic will be generated. Furthermore, another motivation for this thesis is to find a way to detect packet loss and causes for it, because packet loss has a big impact for reliability and should be minimized to reach URLLC requirements.

This thesis consists of general theory about 5G networks in second chapter, and in-depth discussion of URLLC in third chapter. Fourth chapter includes background about network field verification and some practical information about the test setup used for this thesis, relevant concepts to the tests and testing procedures. Fifth chapter explains how the tests are done, then the test results will be visualized and analysed. Sixth chapter is a case study to find out reasons for packet loss, which happened in the fifth chapter's tests. Finally, the seventh chapter concludes the key observations of this thesis and discusses some future work.

2 5G New Radio

The fifth generation mobile network technology (5G), also known as New Radio (NR), is the latest generation wireless cellular network, specified by third-generation partnership project (3GPP). 5G has many benefits and new technologies compared to older generations, such as new spectrum, Massive Multiple Input Multiple Output (MIMO) Beamforming, Network slicing, support for cloud and edge computing and Long Term Evolution (LTE) coexistence. All these new technologies contribute to much faster data rates, lower latency, more capacity, better reliability and overall better network performance. [1]

2.1 New technology components introduced by 5G

2.1.1 New spectrum

The new spectrum offers substantially faster data rates than what older generation networks can provide, because 5G supports much higher frequency bands. 5G is designed to use any frequencies from 400 MHz to 90 GHz, but in practise it has two frequency ranges. The frequency range 1 (FR1) is 450 MHz - 6 GHz and frequency range 2 (FR2) is 24.25 GHz - 52.6 GHz. The FR1 is known as Sub-6 GHz and FR2 as millimeter wave (mmW). The lower bands are used for better signal coverage and surface penetration, and higher bands for faster data rates and better capacity. The mmW frequencies can reach data speeds up to 10 Gbps. [1] [2] UEs can communicate using mmW frequencies by deploying small cell nodes [3].

2.1.2 Massive MIMO

MIMO is a multiple antenna technique, which is often used in modern wireless network systems. The goal in MIMO is to improve network's reliability and capacity.

With 5G, the MIMO has become massive and beamforming is one of the most common approaches for massive MIMO. Beamforming is a signal processing technique, which uses multiple antennas to form a concentrated and directed beam. It can be used for both receiving and transmitting devices. Beamforming is particularly useful for 5G mmW, because multiple beams can be combined at the receiver resulting in better signal-to-noise ratio and increased signal's propagation capability. The beamforming antennas can also be focused on multiple directions to avoid signal being scattered or reflected from different surfaces. [4]

In massive MIMO, base stations are equipped with larger antenna arrays. A base station holds up to 16 antennas for each sector and can beamform the signal. This means that an antenna can simultaneously send and receive data with multiple end-user terminals and neighboring nodes can communicate simultaneously between each other. [3] The signals from every antenna will be processed together in a consistent way. The advanced antenna arrays also result in reduction of total radiation power, which means massive MIMO is more energy efficient. Because of higher frequencies poor propagation capabilities, mmW systems also benefit from massive MIMO, when there is a need for multi-user access in situations where base station is not visible. [5]

2.1.3 Network slicing

5G is aimed to provide various types of services with different kinds of requirements and devices. For example, ultra fast data speeds, ultra reliable connections and massive machine type of communications are all service demands set for 5G. The differences set diverse requirements for all the use cases, meaning one-size-fits-all-architecture fits very poorly for 5G. This will result in a need to slice the network for each service. For instance, one slice for classic UEs and another for IoT. [6]

When the user equipment (UE) enters a network, it will be given information about the available network slices. After this, the UE sends Network Slice Selection Assistance Information (NSSAI) to the 5G core to assist it to choose specific network slice. At the 5G core, Access Management Function (AMF) coordinates the slice related actions. AMF uses Network Slice Selection Function (NSSF) to gather information on which slices can be given for a specific UE. A single UE can be configured for up to eight slices. [1]

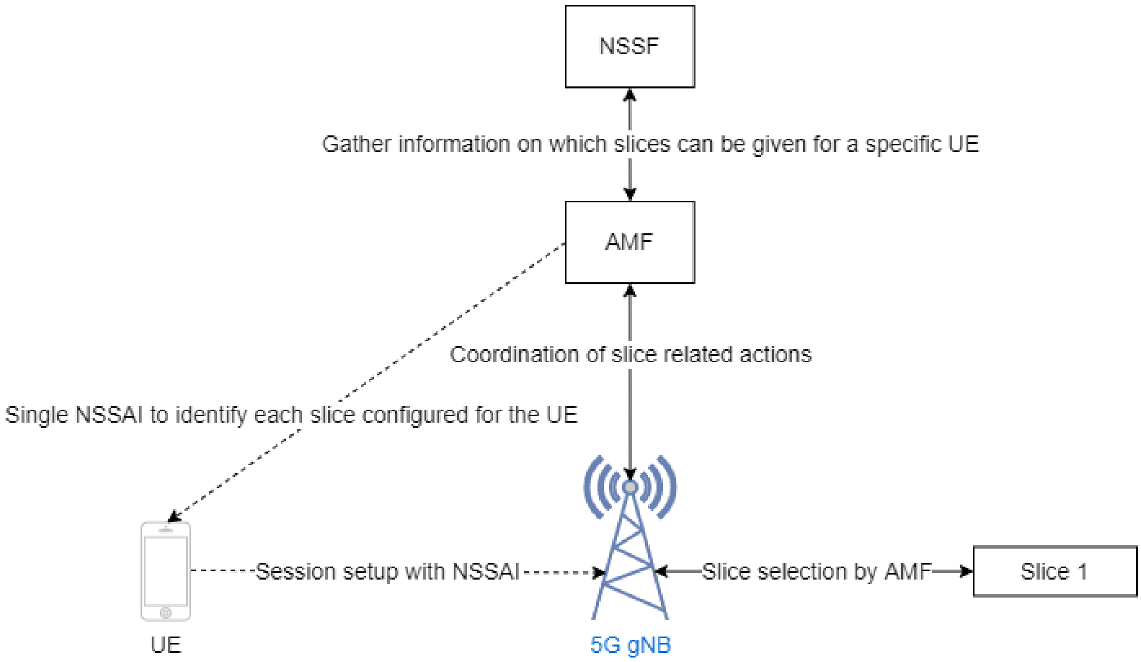

Figure 2.1: Process for setting up a network slice for 5G UE [1].

2.1.4 5G NSA and SA

3GPP approved the non-standalone (NSA) specifications of 5G in December 2017, and standalone (SA) specifications in June 2018 [2]. In both specifications 5G coexists with 4G LTE. Coexisting with 4G LTE makes spectrum farming and sharing possible [1].

The 5G NSA uses 4G Evolved Packet Core (EPC) with 5G NR cells. It means that 5G NSA is heavily dependent on 4G LTE architecture's control functions and add-on services. 5G NSA uses master-slave configuration, where 5G access node is the slave and 4G access node is the master. In short-term, adopting 5G NSA is a cost effective choice for network operators, because there is only a need to invest in 5G NR coverage and LTE network-wide upgrade. However, in the long run NSA network architecture needs to be upgraded to full 5G SA network. This process introduces many additional costs and can be higher than upgrading directly to SA without NSA first. [7]

The 5G SA is an independent 5G network, which provides end-to-end 5G experience using 5G air interface, New Radio, and 5G Core (5GC). Even though 5G SA is an independent network with its own architecture, it still operates with 4G LTE to offer continuity with the two networks. In early phases, adopting 5G SA is more expensive than NSA because there is a need for 5GC since day once. The core network takes about 20% of the network investment, but for 5GC it should be less costly, because of cloud and virtualization technologies. In a long run, it would be better for operators to upgrade straight to SA in order to avoid upgrading the LTE network. [7]

2.1.5 Cloud and edge computing

With 5G, there are many latency and scalability requirements. This leads to edge cloud computing at the edge of a network and close to UEs. The main difference between edge and the traditional cloud is that with edge cloud the servers are located locally instead of remote data centers. Edge computing enables some computationally intensive and high quality of service (QoS) demanding tasks. It is especially important for Ultra-Reliable Low-Latency Communications (URLLC). Edge computing is needed to receive, store, process and analyze enormous amounts of data. It is expected that these high QoS and low latency requiring applications can create over 30 exabytes of data monthly, so instead of using the limited capabilities of UEs, a local cloud system is used to compute and store this data. For applications with less strict latency requirements, tasks are only handled at the edge servers if the delay between UE and traditional cloud service is higher than the requirement. For applications without any latency requirements, tasks are always handled in the traditional cloud service. Edge computing has four key requirements, which need to be satisfied for successful deployment. [3]

1. Real-time interaction to verify low latency support and high QoS.
2. Local processing, which establishes all requests to be processed at the edge servers. Local processing leads to reduced traffic between cell and the core network and prevents bottlenecks.
3. High data rates establishes massive amounts of data sent to the edge clouds without accessing to the core network.
4. High availability of the edge cloud. [3]

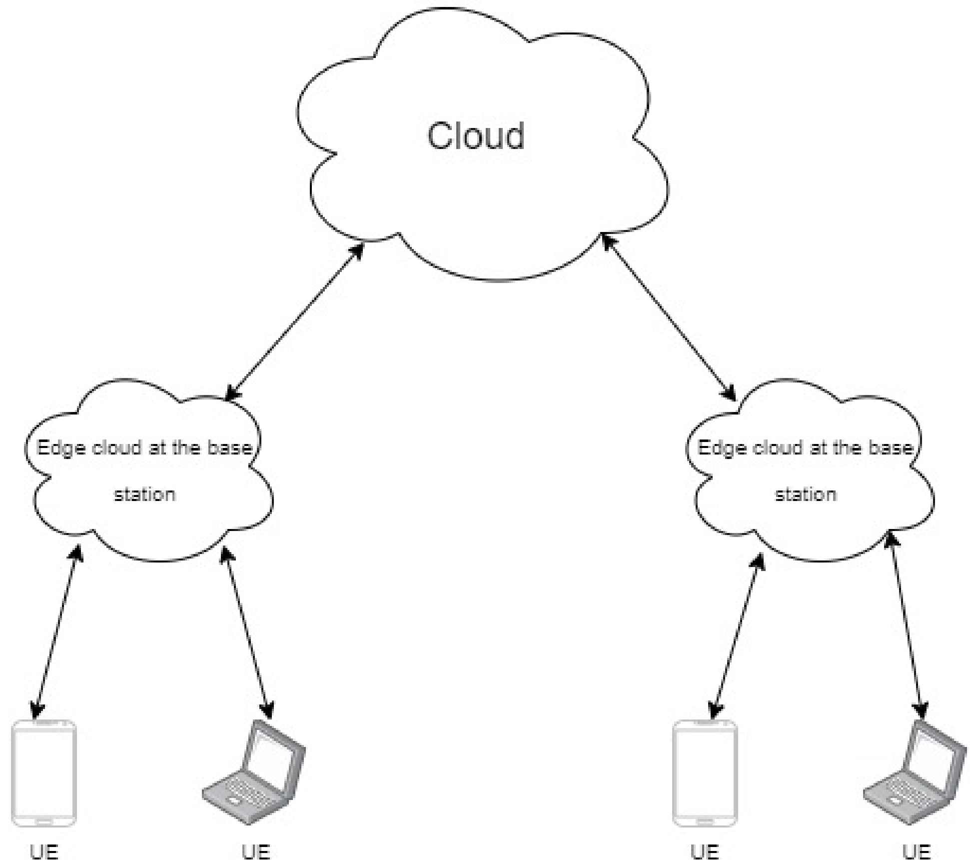

Figure 2.2: Data flow in 5G edge computing [3].

2.2 5G Layers and Protocol Stacks

5G protocol stack consists of three different layers with different functions. The radio protocol stack architecture was released by 3GPP in release 15, in specification 38.300 [8]. The User Plane Protocol Stack can be seen in figure 2.3 and Control Plane protocol Stack in figure 2.4

When a layer receives data from the upper layer, it is called SDU (Service Data Unit). The SDU is then encapsulated into PDU (Protocol Data Unit) for further interaction with the next layer. [8]

The Layer 1 is known as physical layer. For instance, it is responsible for signal

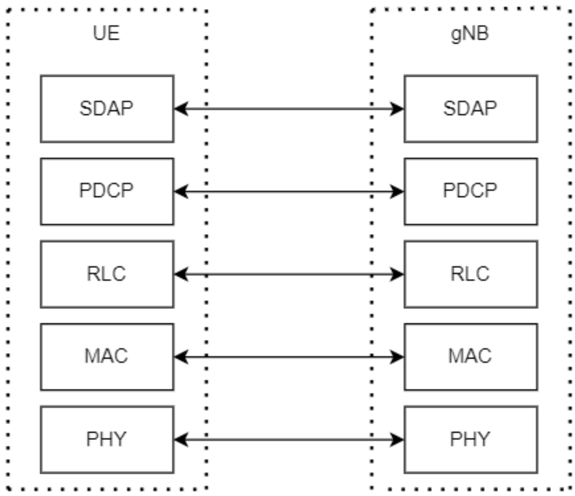

Figure 2.3: 5G User Plane Protocol Stack [8].

modulation, frame structure and downlink and uplink transmission schemes. Typical physical layer operations are link adaptation, power control, cell search and HARQ retransmissions. Physical sublayer offers data to upper layers through transport channels. [8]

The Layer 2 consists of MAC (Medium Access Control), RLC (Radio Link Control), PDCP (Packet Data Convergence Protocol) and SDAP (Service Data Adaptation Protocol) sublayers. [8]

Some typical functions for MAC sublayer are mapping logical and transport channels, mapping multiplexing or demultiplexing of MAC SDUs, scheduling information processing, HARQ error correction and UE priority handling. MAC sublayer offers data to upper layers through logical channels. [8]

RLC sublayer can operate in three transmission modes, which are Transparent Mode (TM), Unacknowledged Mode (UM) and Acknowledged Mode (AM). Typical functions for RLC sublayer are transferring PDUs to upper layers, reassembly and segment of SDUs and protocol error detection and correction. [8]

Typical functions for PDCP sublayer are data transferring, IP header compression and decompression, data ciphering and deciphering and operations to protect

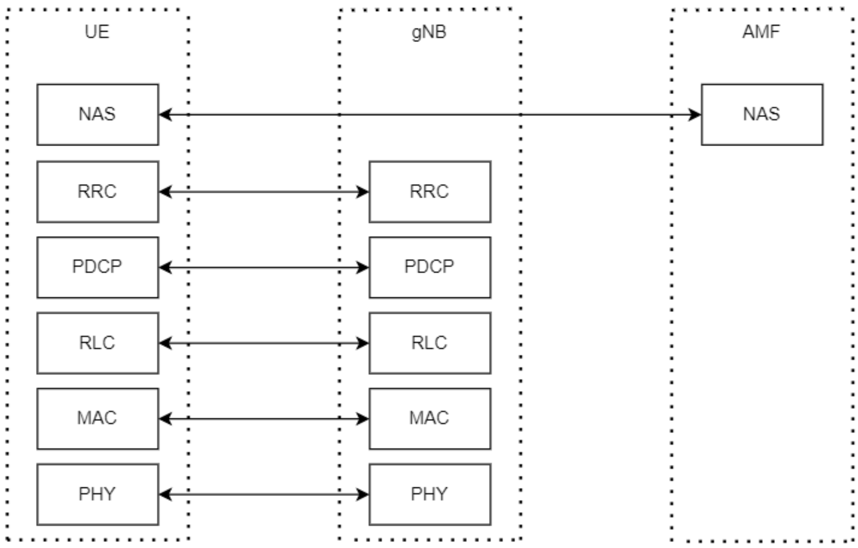

Figure 2.4: 5G Control Plane Protocol Stack [8].

and verify integrity. [8]

SDAP is a new kind of sublayer introduced by 5G, responsible for QoS flow handling for downlink and uplink packets [8]. It was introduced because 5G uses different kind of QoS handling compared to 4G LTE.

The Layer 3 is known as RRC (Radio Resource Control) layer. For instance, it is responsible for establishing, maintaining and releasing RRC connection between UE and radio access network, security functions, operating radio bearers, mobility functions such as handovers and UE context transfer, UE cell selections, QoS management, UE measurement reporting and radio link failure detections. [8]

2.3 Main types of 5G services

5G has three main types of services, which are Enhanced Mobile Broadband (EMBB), Massive Machine-Type Communications (MMTC), and Ultra-Reliable Low-Latency Communications (URLLC). [9]

EMBB provides enchanced data rates, and it can be seen as direct continuation from 4G services. It uses big data payloads and the peak data rates can be very high, but even at the edge of a cell EMBB should offer stable and average data rates. The goal with EMBB is to offer maximized data rates with fairly good packet error rate of 99.9%. [9]

MMTC provides support for large number of IoT devices connected to the 5G network. IoT devices are only occasionally active devices, which use small data payloads and low data rates. The basic principle with MMTC is that a huge number of devices with limited resources can be connected to a 5G base station, which are active only when they transmit data. Therefore, in MMTC the devices are considered only potentially active and there is a random subset of active ones. The packet error rate target for MMTC is 90%. [9]

URLLC provides support for the most mission-critical applications, with extensive reliability and latency requirements. URLLC will be a major concept in this thesis, and it will be discussed in-depth in chapter 3.

2.4 5G usage scenarios

5G establishes many new usage scenarios compared to older generation's mobile network technologies. This is because MMTC and URLLC are completely new types of services with many new possibilities, but also because EMBB establishes substantially higher data rates. [10] 3GPP has identified the following 5G deployment scenarios.

5G deployment scenarios identified by 3GPP in Release 14

1. EMBB indoor hotspot. Used for offices or shopping malls, with up to 10 users for each transmission reception point (TRxP) and 12 TRxPs for 120m x 50m area. Has inter-site (ISD) distance of 20 meters.
2. Dense urban area connectivity. Used to provide connectivity for areas such as city centres with high user densities and traffic load with also outdoor-to-indoor coverage. Has 10 users for each TRxP and ISD of 200 meters.

3. Larger and more continuous coverage in rural area. Has 10 users for each TRxP and ISD of 1700 or 5000 meters.
4. High speed trains, which provide connectivity for 100% of passengers. Cells are connected with Remote Radio Head (RRH) and are placed at around 1700 meters between each other, following the track. There are two TRxPs for each RRH site.
5. Extreme long distance coverage. Used with large areas with low amount of users. Uses isolated cell with 100 km range.
6. Urban coverage for MMTC. Uses larger cells and provides more continuous coverage for high density of IoT devices. Deployed with ISD of 1700 or 500 meters.
7. Highway scenario where vehicles with high speed are connected. Deployed with ISD of 1700 or 500 meters.
8. Urban grid to connect high density of vehicles. The grids use ISD of 500 meters, or 50-100 meters with Road-Side Units (RSUs) placed on intersections. RSU is a communicating device which connects vehicles to road infrastructure.
9. Connection services for commercial aircrafts. Serves both humans and machines. Cells need very large coverage area, with up to 100 km. The aircraft must have a carrier aggregation point to increase the data rates for users. [10] [11]

3 Ultra-Reliable Low-Latency Communications (URLLC)

The specification for Ultra-Reliable Low-Latency Communication (URLLC) started by 3GPP in release 15 and it is one of the main types of services provided by 5G. URLLC and MMTC make the biggest difference to 5G compared to previous generation's mobile network technologies. URLLC is designed to be used with the most mission-critical wireless communication applications, and the most important key performance indicators (KPIs) are latency, reliability and communication service availability. Other also valid performance indicators are jitter and synchronicity. URLLC enables support for wireless applications with extensive latency, reliability and availability requirements. These applications can be for example vehicle-to-vehicle communications for increased traffic safety and efficient driving. The strict requirements of URLLC make it a demanding service to implement and drives the development of advanced 5G techniques. [12]

3.1 Key requirements

3.1.1 Latency

3GPP has defined end-to-end latency as "the time that it takes to transfer a given piece of information from a source to a destination, measured at the communication interface, from the moment it is transmitted by the source to the moment it is successfully received at the destination". [13] The end-to-end latency is the sum of radio latency, transport latency and core latency, and it can be as low as 1 ms.

User plane latency

The time it takes from a packet to reach from its source to destination is called user plane latency. This time is measured in milliseconds (ms) [14].

The URLLC requirement for user plane latency is 1 ms from server to the client or from client to the server [15], assuming an ideal scenario where both uplink and downlink have only a single user and small IP packets are exchanged [14].

Control plane latency

Control plane latency means the time it takes control plane to transfer from most battery efficient state to active data transferring state. The maximum allowed latency for URLLC control plane is 20 ms, but lower latencies are possible and encouraged. [14]

3.1.2 Reliability

Traditionally, reliability refers to the system's ability to transmit certain amount of data with a high probability of success, function properly and operate without interruptions in a given amount of time. A system which never fails is perfectly reliable and operates successfully 100% of the time. Such system is not possible to create, but with reliable components, system design without undiscovered errors and with failure masking where a single failure won't stop service completion the reliability can be increased significantly. A reliable system can detect and most importantly recover from failures and errors. It is also good to note that a reliable system doesn't have to run constantly to accept and complete requests. [16]

According to 3GPP definition, reliability is the "percentage value of the packets successfully delivered to a given system entity within the time constraint required by the targeted service out of all the packets transmitted" [13]. In 5G network packet transmission system, this can mean for example that one-way latency is < 10 ms for 99.99% of sent packets. For URLLC, the reliability is evaluated by its probability of successfully transmitting a layer 2 PDU (SDAP, PDCP, RLC and MAC) in a minimum time requirement and at a certain channel quality. [14] 3GPP has defined URLLC reliability in a following way: "A general URLLC reliability requirement for one transmission of a packet is $1 - 10^{-5}$ for 32 bytes with a user plane latency of 1 ms" [11]. This means that 99.999% of 32-byte packets should be received from source to destination in less than 1 ms latency.

3.1.3 Availability

Availability refers to the system's accessibility for the users. When a system answers users service requests in a timely manner, it is considered to be available. The difference between reliability and availability is that availability is instantaneous. It describes the time periods where the system is accessible for the user. [16]

System availability in its traditional definition doesn't describe system's readiness to successfully perform a given task, or accessibility over a network. This means that even if a system is considered to be available, it may still not be ready for new service requests. For example, that kind of situation may occur during fault recovery, when new service requests are blocked, even though the system is accessible. [16]

With URLLC, availability is discussed as communication service availability, which takes system's possible fault and recovery situations into account. In other words, when URLLC system is considered available, it is always expected to answer service requests. 3GPP has defined communication service availability as the time communication service meets agreed QoS, divided by the total amount of time service is delivered. For example, the communication service is considered unavailable if a message is not received within a specified maximum allowed latency and survival time. 3GPP has not defined an exact minimum availability requirement for URLLC, but there are URLLC applications with demand availability of 99.9999% and in theory availability can be even higher. [13]

3.2 Technology to achieve latency and reliability requirements

5G has introduced various new technologies which are used to improve directly latency and reliability, and indirectly availability. These technologies are driven by URLLC, hence can be called "URLLC toolbox". Most of the toolbox components are specified in 3GPP release 15, and they are divided into two categories, based on latency and reliability improvement. [12]

Some of the key technologies in the toolbox and furtherly discussed in this thesis are shortened TTI duration, Non-slot based scheduling, grant-free transmission, DAPS handover, micro-diversity, HARQ, interference mitigation and low rate MCS. [12] [17]

Table 3.1: Technology to establish URLLC [12] [17].

Low latency	Reliability
Shortened TTI duration	Micro-diversity
Non-slot based scheduling	HARQ
Grant-free transmission	Interference mitigation
DAPS handover	Low rate MCS

It should be kept in mind that this is not an exhaustive list of new technology introduced by URLLC, but more of an overview on what kind of operations can be done to reach URLLC standards.

Even though URLLC has many new technologies designed for latency reduction and reliability improvement, these two factors are not independent and correspond strictly to each other. [12]

3.2.1 Low latency

Shortened TTI duration

Transmission Time Interval (TTI) is used to measure transmission duration for a radio node. 5G uses flexible frame structure, which offers two ways to shorten the duration of TTI. First is reducing the amount of Orthogonal Frequency Division Multiplexing (OFDM) symbols for a TTI and second is configurable subcarrier spacing (SCS). These techniques allow data transmission to have lower latency, be more fault tolerant and also be more reliable. They offer a big advantage over 4G LTE latency, because latency directly correlates with TTI. [12]

OFDM is a multicarrier transmission scheme, where a single broadband stream is transmitted by splitting it into multiple narrower ones, called subcarriers. OFDM offers good resistance against signal interference and frequency-selective fading, because only a small number of subcarriers are affected by them. In OFDM, the subcarriers overlap, which results in better usage of the available spectrum. To reduce the overlapping subcarriers causing interchannel interference, the receiver behaves as a bank of demodulators, which integrates the signal over a symbol period, in which the raw data is received. The main advantages of OFDM are resistance against interference, enhanced capacity and low complexity. [18]

In 5G, the configurable subcarriers use different spacings for each channel, which are 15, 30, 60, 120 and 240 KHz. The higher the spacing, the more bits of data can be put in one frame. In 5G, the reduced TTI can be achieved by reducing the OFDM symbol duration by increasing subcarrier spacing or by reducing the amount of symbols per TTI. [12]

Non-slot based scheduling

Network packet scheduler manages all the packet transmission operations. It regulates how much data an application can receive and controls its overall traffic. Packet scheduler prioritizes traffic, decides in which order the network packets are transmitted and received and shapes the way packets are transmitted by smoothing packet transmission peaks over time. [19]

In 5G, frames are used to contain and carry network packets. A frame consists of 10 subframes, created in 1 ms duration between each other. Each subframe has 1, 2, 4, 8 or 16 slots, based on subcarrier spacing. Schedulers use slots as common units of transmission. Non-slot or mini-slot based scheduling means that the transmission of data can start at any OFDM symbol, and last only for as many symbols as are required for the communication. Mini-slots are the smallest possible scheduling units and they support extremely short transmission duration and processing time. 2, 4 or 7 OFDM symbols are some mini-slot lengths expected to be used for URLLC. [20] [12]

Grant-free transmission

In grant-free transmission, the UE transmits data without sending scheduling requests and without receiving resource allocations from network. All data sent by UE is grant-free, when it arrives to the network packet scheduler. This means that data arrives to scheduler in arrive-and-go fashion without any extra delays. [21]

Unscheduled packets may have some random collisions, but the grant-free transmission should still result in very reliable access satisfying URLLC standards. Because there are no scheduling delays for network packets, grant-free transmission initially lowers latency. However, when things go wrong there will be problems to meet URLLC latency standards. If a packet is transmitted unsuccessfully and needs a retransmission, the probability for the transmission to take over 1 ms is very high. [21]

There are some techniques to mitigate this, such as K repetitions. This technique sends the grant-free replicas of a same packet for each TTI for K amount of times and terminates unnecessary transmissions when one packet is successfully received. The terminating helps to reduce collisions by removing unnecessary retransmissions. [21]

DAPS handover

Handover is a change of radio connection, caused typically by mobility making coverage worse. The change of radio cell typically shortly increases latency, which causes problems to reach URLLC requirements. 5G uses hard handovers, in which the connection to source cell is dropped before attaching to the new cell. 3GPP has specified DAPS (Dual Active Protocol Stack) handover to mitigate this. DAPS is discussed in detail in section 5.5.2.

3.2.2 Improved reliability

Micro-diversity

Micro-diversity means that there are multiple antennas at the transmitting side, receiving side or both sides of a datalink, maximizing the signal diversity. For URLLC requirements, a data link should have at least two transmitting and two receiving antennas (2x2), but more is preferred. For reliable enough diversity order, 4x4 is recommended. [12]

A single-user single-stream is an ideal transmission system to meet URLLC reliability requirements. In this mode, multiple transmitting antennas at the transmitting device transmit multiple data streams to the same receiving device, resulting in increased reliability. [12]

HARQ

HARQ (Hybrid Automatic Repeat ReQuest) is a forward error and soft combining technique already introduced for 4G LTE, with the goal of enabling faster recovery from erroneous packets. Automatic Repeat ReQuest (ARQ) is used in a situation where data sender doesn't receive acknowledgement (ACK) signal from receiver. When no ACK is received, the packet is discarded and a new one retransmitted. In soft combining, erroneous packets are stored in a buffer in a hope that two or more erroneous packets can be combined together to receive the data. [22]

There are thee types of HARQ, based on the behavior when waiting for ACKs. First type of HARQ adds forward error correction bits to a packet before transmitting it. When the channel quality is good, some errors can be detected and corrected in the packet message without retransmissions. With bad channel quality, the errors may not be corrected and retransmissions will be made. In type 2 and 3 HARQ, retransmissions are made with different kinds of data, error correction bits and forward error correction bits. In type 2, redundancy is added to every packet's retransmission which the receiver needs to decode. In type 3, retransmissions are not made if the channel quality is not good enough. Also, similarly to type 2 the receiver is expected to decode the retransmitted data. [22]

When a strict reliability requirement is tied to low latency, retransmissions can bring challenges to reach the strictest URLLC requirements. That is why for URLLC a channel coding method to support efficient HARQ and error flooring optimization system should be selected. A good channel error correction candidate for URLLC can be for example Low-Density Parity-Check (LDPC) code system. [23]

Interference mitigation

SINR (Signal to Interference and Noise Ratio) measures the desired signal's theoretical performance to the actual signal's performance which suffers from background noise. SINR is determined by the ratio of these two, and it uses the unit of decibels (dB). The higher the dB, the better signal quality. When the ratio is over 0 dB, the signal is better than the noise level. An interfering signal causes lost packets and packet retransmissions which results in lower data rates and higher latency.

To increase SINR for URLLC, interference can be reduced with both network-based and UE based techniques. For example, by canceling interference from a neighboring base station. A basic principle is that by cancelling two strongest interferers the signal quality should be close to the desired level. [12]

Low rate MCS

Modulation and Coding Scheme (MCS) defines how many useful bits can be transmitted per resource element. UE sends CQI (Channel Quality Indicator) report to gNB, which allocates MSC to reach BLER target. Lower BLER is desirable for URLLC. MCS, BLER and Link Adaptation algorithm are discussed more in section 5.5.4.

3.3 URLLC deployment scenarios

URLLC has a lot of potential for applications with strict latency and reliability requirements. Some typical use cases for URLLC are industry automation applications, self-, or remote-driving vehicles (V2X), and smart grid applications. [24] [12]

3.3.1 Industry automation

The development of IoT technologies has driven change to industry producing processes. Many industrial communication networks have been built on wired networks, but with URLLC wireless networks have become a worthy option. There are three reasons for why wireless networks are much more worthy in harsh industrial environments. [24]

1. Costs. Wireless networks are cheaper because they require less materials, less installation and less maintenance.
2. Reliability. Even though wireless networks may suffer from signal fading and interference, physical cables will age and they can break quickly in harsh environments.
3. Suitability. Wireless networks are more practical in many industrial situations. For example with moving objects like robots, harsh environments like high temperatures or long distances. [24]

Smart industrial automation applications often use IoT devices or robots to collect environmental information to make automatic adjustments, adaptations or reactions based on the information they receive. These applications have the following requirements which must be fulfilled to make correct decisions consistently. [24]

1. Latency. These applications have very strict latency demands to minimize the time between an event in the factory environment and the automatic reaction to it.
2. Reliability. There can be some automatic decisions with very high value where a transmission error can result in fatal situations. The reliability demand should be adjusted to the latency requirement however, because stricter latency usually results in reduced reliability.

3. Throughput, for example to transmit images and videos with high resolution.

4. Signal interference and fading resistance. Industrial environments are mixed with many kinds of signals from many types of communication devices. There must be resistance to interference to have a reliable communication system.
5. Energy efficiency. It is especially important for some low power IoT devices, which are power supplied with battery.
6. Signal coverage, to make sure collected information can be efficiently transmitted and the decision making device reached. [24]

For these applications, latency, reliability, throughput, interference and fading resistance, energy efficiency and signal coverage are extremely important and without them correct decisions could not be made consistently. [24]

Table 3.2: Use cases for industrial automation with latency and reliability requirements [12] [25].

Scenario	Latency	Reliability	Data rate
Process automation for monitoring	60 ms	99.9%	1 Mbps
Process automation for remote control	60 ms	99.999%	1 Mbps - 100 Mbps
Discrete automation	10 ms	99.99%	10 Mbps
Discrete automation for motion control	1 ms	99.9999%	-

It is also good to note that URLLC is not always fully tenable solution for industrial control and manufacturing applications. There are some extreme use cases where reliability demand could be as high as $1 - 10^{-9}$ and latency demand below 0.01 ms, which is out of URLLC's capabilities to achieve. [24]

3.3.2 Self- and remote driving vehicles (V2X)

With the development of wireless communication and IoT systems, it is possible to achieve much more capable and efficient transportation. V2X is one of the most important applications for 5G URLLC. It has the goal of obtaining automated, accident-free and cooperative driving. V2X has the following requirements for the wireless communication network. [24]

1. Latency. The latency should be very low for autonomous vehicles communicating between each other, much lower than what 4G LTE can currently provide.
2. Reliability. Transmission errors, such as lost packets, may cause serious incidents with autonomous vehicles and therefore reliability of these systems should be very high.
3. Data rates. There are V2X applications which sense their environment through high-resolution images and videos, which require high data throughput.
4. Signal interference and fading resistance. Vehicles are exposed to heavy signal interference generated by other wireless communication systems, especially in an urban area. There may also be buildings or other large objects scattering or reflecting the signal.
5. Signal coverage. V2X transmissions should always be able to reach the closest base station, which can be hundreds of meters away.
6. Mobility support. V2X applications should be able to transmit and receive data in very high speeds, for example with high speed trains. [24]

Vehicle-to-Vehicle (V2V) is one of the key V2X applications. In V2V, vehicles change information between each other. For example, with V2V a front vehicle can inform vehicles behind about short stoppages and thus avoid a traffic jam. In Vehicle-to-Infrastructure (V2I), vehicles can communicate to roadside units, such as traffic lights, and high priority vehicles can be given access to go first. In Vehicle-to-Pedestrian (V2P) communication, vehicles can communicate to people who are at the side of the road. In Vehicle-To-Network (V2N) communication, vehicles can connect to a network, for example a traffic information system. [24] [26]

Each of these key V2X applications need to perform the following subsets of operations in order for the vehicle to be fully cooperative and automated: Cooperative awareness, Cooperative sensing, Cooperative maneuver, Vulnerable road user, Traffic efficiency and teleoperated driving. [24] [26]

1. Cooperative awareness (CA) for warning and environmental awareness operations.
2. Cooperative sensing (CS) to sense and exchange raw sensor data.
3. Vulnerable road user (VRU) has similar behavior to CA, but the data is transmitted to different kind of devices, typically to a smartphone.
4. Cooperative maneuvers (CM) include use cases such as collision avoidance and lane change. Actions like these require very low latency and high reliability.
5. Traffic efficiency (TE) supports V2N and V2I without hard delay or reliability requirements. Vehicles transmit their location to server every few seconds, with vehicle status and road information included.
6. Teleoperated driving (TD) is used for command and control actions for the remote controlled or autonomous vehicle, which collects data from cameras installed in the vehicle. [24] [26]

In table 3.3 the technical requirements for each operation is discussed.

Table 3.3: Requirements for different V2X operations and its overall infrastructure [24] [25] [26] [27].

Operation	Mode	Latency	Data rate	Reliability	Range
CA	V2V, V2I	100 ms - 1 sec	5-96 kbps	90 - 95%	$<$ 500m
CS	V2V, V2I	3 ms - 1 sec	5-25000 kbps	$>$95%	$<$ 200m
CM	V2V, V2I	$<$3 ms - 100 ms	10-5000 kbps	$>$99%	$<$ 500m
VRU	V2P	100 ms - 1 sec	5-10 kbps	95%	$<$ 200m
TE	V2N, V2I	$>$1 sec	10 - 2000 kbps	$>$90%	$>$ 500m
TD	V2N	5-20 ms	25000 kbps	$>$99%	$>$ 500m
Backhaul	-	30 ms	10 Mbps	99.999%	2 km

3.3.3 Smart grids

Traditional grids are used to transmit power from a few key generators to a large amount of customers. Smart grids are advanced energy delivery networks, which share information about electricity generation, transmission, distribution and consumption in an intelligent two-way manner. Smart grids use modern technology such as sensors and communication networks to transmit power more efficiently and more reliably. With the help of technology, smart grids can react and adopt to occurring events anywhere in the grid, for example if something breaks in the grid it is able to reroute the power and stay operating. Another example is demand profile shaping, where a smart grid lowers peak demand for electricity by adjusting its real-time price. This results in lower electricity demand and thus smoothed electricity demand peak. [28]

To support two-way electricity transferring, a smart grid must have low latency, high data rates and good reliability. A good example why these are required is demand profile shaping, where low latency and two-way communication is needed to provide accurate electricity demand data and price adjustments. Wireless networks are very suitable for smart grid applications, because they offer lower installation cost, faster deployment and better mobility and flexibility than wired networks. [24] [28]

Smart grids' main requirements have not yet been completely identified for all possible scenarios, but estimates have been made, such as some availability and latency requirements by European Telecommunications Standards Institute (ETSI) [24] [29]. Also, 3GPP has identified some electricity distribution reliability and latency requirements in release 15 [25].

Table 3.4: Smart grid latency, availability and data rate requirements identified by ETSI [29].

Scenario	Latency	Availability	Data rate
Protection	5-10 ms	99.99%	64 kbps
Monitoring Class A	100 ms	99.99%	64 kbps
Monitoring Class B	100 ms	99.99%	64 kbps
Monitoring Class C	1 s	99.99%	64 kbps
Voice	500-100 ms	99%-99.99%	8-32 kbps
CCTV Monitoring	500ms-1s	99%-99.99%	64-256 kbps
Management	500 ms	99.99%	64-512 kbps

Table 3.5: Smart grid latency, reliability and data rate requirements identified by 3GPP [25].

Scenario	Latency	Reliability	Data rate
Medium voltage electricity distribution	40 ms	99.9%	10 Mbps
High voltage electricity distribution	5 ms	99.999%	10 Mbps

In table 3.4, protection means all of the following: protecting devices at the end of a two-way communication line, current protection on transmission lines, blocking signals and equipment protection from medium voltage. [29]

Monitoring is divided into three different subclasses. Monitoring class A means monitoring and controlling transmitting devices at substations. Monitoring class B uses primary distribution substation, and Monitoring class C secondary distribution substation, thus much higher latency requirement. [29]

Voice allows messaging between the control centre and remote sites during fault recovery type of operations, and for this 99.99% reliability with 100 ms latency is required. When the remote sites or their staff are not in a fixed location, only 99% reliability with 500 ms latency is required. [29]

CCTV (Closed Circuit Television) monitoring allows control room to watch live stream video and view images of a remote site and interact with employees in the location. [29]

Management establishes remote site employees to access real time operational data to support work tasks. It also establishes control for telecommunication infrastructure from a central position. [29]

As seen in table 3.5, automation in medium and high voltage electricity distribution is a very critical operation, so it has strict reliability and latency requirements. Also, high voltage electricity distribution system has very high availability requirement at 99.9999%. [25]

4 Network testing

The goal for this chapter is to give an explanation on what is network performance testing, what is a field test environment, what are the basic procedures in network testing, how to present test results and what kind of conclusions can be made. Then, the test setup used for this thesis' measurements will be introduced and relevant concepts related to the testing will be explained.

4.1 Field verification as a concept

5G has digitalized many industrial applications. The digital transformation of these applications has created many new requirements for network's data rate, latency and reliability. The main goal with field verification is to measure the KPIs (Key Performance Indicators) in an environment which is close to a real life use case. The measurement results should then be analysed to give some recommendations related to the real life use case. These recommendations can be for example parameter configurations. Another motive for network testing is to give guidance for further development. [30]

The field test environment should have UEs such as phones or modems, base station with NR cells (gNB) and a server. The latency between UE and base station is called air interface latency, and the time between base station and server is called transmission latency.

Each base station in the field test environment has their own properties, which should be taken into account when testing. For example, typical base station parameters consist of carrier frequency, duplex mode, output power, antenna configurations (e.g. transmitting and receiving antenna ports), channel bandwidth, subcarrier spacing and schedule request cycle time. [30]

Field testing requires a lot of configuring. Test conditions refer to the parameter configurations designed for a specific test case. For example, a latency test configuration may consist of selecting proactive scheduling, schedule request periodicity, schedule time interval, PUCCH mode and ping packet size.

4.1.1 Reliability and latency testing

Reliability and latency are some of the main factors contributing to network's capacity and performance, and are especially important for URLLC.

For reliability testing, the test characteristics should represent potential URLLC use cases. 5QI (5G Quality of Service Identifier) is an indicator representing QoS for different 5G services. It includes desired packet latency budget, minimum packet error loss rate, priority level, maximum data burst volume and data rate averages for many example services. For URLLC testing, a desired 5QI level should be selected, and then it should be measured whether the desired reliability and latency level can be achieved with current test setup. [31]

A CCDF (Complementary Cumulative Distribution Function) is an excellent way to represent reliability under a given one-way latency. It visualizes both head and tail of the latency distribution and reveals possible packet loss. From a CCDF, it can be directly seen how different latency levels affects reliability. [30]

4.2 Field verification environment

The measurements conducted for this thesis were made at Nokia field test environment, located in Espoo, Finland. The field test environment supports many different kinds of network deployments, such as 5G TDD cmW, mMIMO, 5G FDD, 5G TDD mmW and LTE.

4.2.1 Setup for URLLC testing

URLLC tests are conducted on TDD cmW 5G NR cells. The testing consists of one-way latency and reliability measurements between time synchronized commercial UE and server. The UE is a Linux PC using Fedora 33 as its operating system, and it's connected to an Askey 5G SA modem, which is located in a field verification test van. This modem will be a good candidate for URLLC testing, because it's compatible with 5G SA and an internal network analytic tool, called Nemo Outdoor 5G NR Drive Test Solution. The server is located in Nokia premises, and it's also running on Fedora 33. The setup is visualized on figure 4.1.

The Iperf test results are saved automatically to Cassandra database, which is then utilized for further analysis and plotting with Python libraries.

4.3 Test tools and other relevant concepts

4.3.1 Iperf

Iperf is a traffic generator tool to measure network's performance. It supports protocols such as TCP, UDP and SCTP with many different measurement parameters. Iperf creates a data stream between client and server, and it can be used to measure network's behavior on various parameters. [32]

For this thesis, Iperf is used to create a UDP packet stream between time synchro-

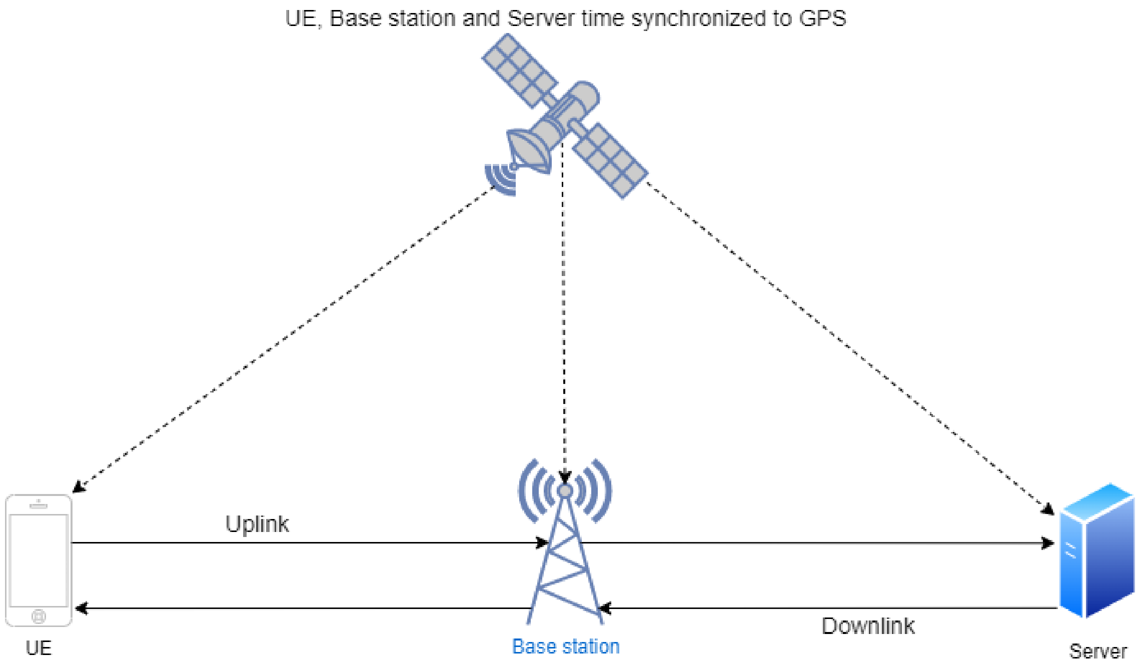

Figure 4.1: URLLC test setup.

nized client and server. It is a good tool of choice, because it supports URLLC type of packet burst generation. Iperf is configured to save data in three second intervals in the Cassandra database. The saved data is then used to create the CCDF plot representing reliability, and the average latency interval plot representing average latency after each 3 second time interval.

4.3.2 Wireshark

Wireshark is an open source network traffic analyzer, which is used to capture, save and analyze data flowing through a network. Wireshark runs on Windows, Linux and on many other operating systems. It offers support for a wide range of different kinds of network interfaces for live capture. Different kinds of internet protocols are also widely supported for live capture and packet analysis. For the analysis of the capture files there are wide range of filters, tools and plugins available. There is also a command line based version of Wireshark, called Tshark, which is useful for example for monitoring network traffic on servers with no GUI. [33] All these factors make Wireshark an excellent tool for packet analysis for this thesis.

For this thesis, Wireshark was installed on the client and Tshark was installed on the server, for packet level analysis. Then, packets were uniquely identified using an Iperf dissector plugin and their transmission between server and the client were analysed for both uplink and downlink. The goal was to identify which packets never arrived. Wireshark was also used to measure each packet's latency from server to client or vice versa. This was done by comparing each packet's arrival times and distinguishing the difference.

The basic procedure with Wireshark was to start the capture on client and server, launch the Iperf UDP stream, collect and save the packet capture (pcap) files after the stream ended, filter them, create CSVs and use Python libraries for further analysis.

4.3.3 UDP

User Datagram Protocol (UDP) is an internet communication protocol, used for low latency and packet loss tolerating connections. Compared to TCP, it provides faster and less resource demanding connection. UDP doesn't use any kind of handshake procedures to establish a connection, meaning it is a connectionless protocol. UDP's messages are called datagrams and they are transmitted in a best-effort way. This means that UDP protocol does not verify in any way that the transmitted packets will be received. The UDP packets can also arrive in wrong order or they may appear as duplicates.

UDP wraps a datagram in a UDP header, which has four different kind of fields. First field is source port, which is simply the port of the transmitting device. Second field is the destination port number where the datagram is transmitted. The port can be anything from 0 to 65535. Third field is the UDP length. It consists of the length in bytes of the UDP header and the UDP data being transmitted. Fourth field is checksum. Checksum is used to error check the UDP packet header and its data. [34]

UDP has three types of applications that fits for it the best. First types are the applications that can tolerate packet loss, but require very low latency. For example Voice over IP (VoIP) is that type of application. Second types are the applications which use simple request and reply transactions. This means a host sends a request, but there is no any certainty whether there will be a reply. For example, DHCP and DNS are these kind of protocols running on top of UDP. Third types are the applications that can handle reliability on their own, thus dont need TCP to handle reliability. For example, Trivial File Transfer Protocol (TFTP) is this type of protocol, operating on top of UDP. It has its own error detection and recovery mechanisms and therefore it takes care of its own reliability. [35]

4.3.4 Packet loss

As discussed in section 3.1.2, reliability is the percentage value of successfully delivered packets, out of all delivered packets in a given time constraint. A lost packet means that a transmitted network packet never reached its destination. Lost packets have big impact to reliability, especially from URLLC point of view, because they will result in certain reliability levels to become unreachable under any latency requirement.

For end-user, packet loss impacts the presentation quality of an application and it can occur as laggy service or loss of connection. Especially real-time packet processing application's quality suffers from packet loss. For example sound, voice and picture quality of a video will be heavily reduced by packet loss. The impact is also affected by which protocol is used. With TCP, a lost packet will be retransmitted and packet loss causes mainly slowness to the network. With UDP, lost packets are not re-transmitted and there will be wider effect from packet loss. [36]

There are several factors which can cause packet loss. A network link can have more traffic than it can store, and become congested. A congested link has an overflown buffer, which means it has lost the packets it had prepared to be transmitted. A network device, such as router, may as well fail. When a device crashes while it had some network packets stored in its buffer, the packets will be lost. Also network reachability status can change, for example the routing protocol. This can cause packets to have unreachable destination, and thus become lost. [36]

There are different kind of methods to prevent packet loss, depending on what is causing it. If packet loss is caused by network congestion, increasing bandwidth, reducing packet retransmissions and reducing packet burst size are all possible methods to eliminate or reduce packet loss. If packet loss is caused by a failing network device, updating its firmware may help to speed up the traffic and free some bandwidth and also fix some bugs. Updating old hardware is also a possible solution, especially if the device is crashing for no apparent reason. Some less viable solutions to prevent packet loss are using wired connections and reducing interfering signals, which are obviously not possible in all scenarios. This thesis includes a packet loss analysis study in chapter 6.

4.3.5 Uplink resource scheduling

In the test environment for this thesis, it is expected that uplink's reliability and latency are worse than in downlink. It might have some indirect reasons such as UE's weaker transmission power, but the main reason for that is uplink resource scheduling.

Generally, in resource scheduling the network tries to fulfill the demands that applications have for network resources. These demands can be for example bandwidth and buffer space. [37]

Buffers are used as a network traffic management mechanism. They serve as memory blocks, to reduce packet loss when routers cannot forward packets due to a large amount of traffic. Buffer size can be measured for example with maximum number of stored packets, amount of stored data in bytes or milliseconds. [38]

Fulfilling all the applications network resource demands is very rarely possible, which will result in some applications receiving fewer resources than others. The key task for resource scheduling is to identify which resources should be given to which applications. [37]

In uplink resource scheduling, the base station's packet scheduler allocates resources for UEs transmitting uplink data. A UE sends a scheduling request to a base station, asking for air-interface resources for a new transmission. The sequence of actions in scheduling request transmission is the following: First, data that needs to be transmitted queues in the UE buffer. This waiting data triggers the UE to send a schedule request to the base station. Base station receives the scheduling request and decides a correct logical data link or links for the UE. The base station then allocates resources to the PUCCH (Physical Uplink Control Channel). After this, the UE sends a buffer status report with some of the buffered data. The report includes information on how much base station can expect data from the UE and adjust resources accordingly. If the buffer is emptied in the first transmission, there is no need to send buffer status report. [39]

In proactive scheduling, UE can transmit data without sending schedule requests to the base station's packet scheduler and waiting for an allocation. This is established by base station giving PUCCH resources before any requests. The purpose of proactive scheduling is latency reduction, because UE doesn't have to send a scheduling request and wait for an allocation. Proactive scheduling is usually valid for cells which are not heavily loaded. [39]

In the test configuration used for this thesis, the base station packet scheduler gives proactive UL grants for the UE automatically every 4 ms. This means that the UE does not have to send scheduling requests and can directly transmit the data it has in its buffer. The latency correlates on how much transmitted data the UE had buffered after each 4 ms slot, and when is the next possible slot to send data in the TDD radio frame.

5 Reliability and one-way latency analysis

In this chapter, ways to measure and visualize 5G SA network's performance on URLLC requirements are presented. The measurements are done in a field test environment with access to Nokia 5G SA base station, used specifically for network capacity and performance testing. The 5G cell uses n77 frequency band, 100 MHz channel bandwidth, 30 KHz subcarrier spacing, 36.9 dBm maximum output power and TDD (Time Division Duplex) to adjust uplink and downlink resources. The radio uses 4x4 MIMO for downlink and 2x2 MIMO for uplink.

The measurements are done in 5G SA network, in which the downlink data rate can reach up to 1 Gbps in most optimal testing conditions. In the near future, when the basic URLLC features are implemented, the network will be sliced to create a dedicated URLLC slice.

The tests are 5 minute maximum throughput tests using UDP packet stream between server and client, where both are synchronized to a common clock. The packet size for each test is 255B, with 199 sent packets per second (bursts). There are about 59 700 packets in total for each 5 minute test run.

In 5G quality of service identifier (5QI), a packet stream with this type of packet size and packets per second has a value 85, which is considered as very delay critical guaranteed bit rate. Example services for this type of packet stream is high voltage electricity distribution system automation and V2X messages. [31]

The result data will be gathered from Iperf with three second time intervals, and from Wireshark capturing packet data from both server and the client. The results are then discussed, analysed and plotted using Python libraries.

There is a known issue with Iperf marking the first packet of each packet stream incorrectly as lost. This will not have a big impact to the tests, basically it means that when there are no lost packets the reliability never reaches the full 100% under any latency. Instead, the reliability increases way past 1-10^{-4} (99.99%) and close to the bottom of the Y-axis, which represents 100% reliability. When there are some actual lost packets, the plot will grow a tail which never reaches full 100% under any latency. The Iperf issue can cause the tail to appear in marginally lower reliability level, depending on how many lost packets there are in total. It is verified that this is an Iperf issue and not an actual lost packet with Wireshark packet capture analysis. The first packet has never been identified as lost with it and in total there is always exactly one lost packet less than what Iperf shows.

For this thesis, four different test scenarios were identified which imitate the scenarios where URLLC applications will be used. There are stationary and driving tests, with two testing scenarios for each one. Stationary tests are done with good coverage and bad coverage from the 5G cell. Driving tests are done by moving from good to bad coverage while connected to the same 5G cell, and in handover in which the 5G cell is changed during the 5 minute measurement. All the tests are done in both downlink and uplink.

5.1 Good coverage

The good coverage tests were conducted with a clear line of sight to the 5G cell, at about 60 meter distance. There was no interference and no other network traffic than the Iperf UDP packet stream. The RSRP (Reference Signal Received Power) for this test was about -75 dbm, which can be considered as an excellent value.

5.1.1 Reliability

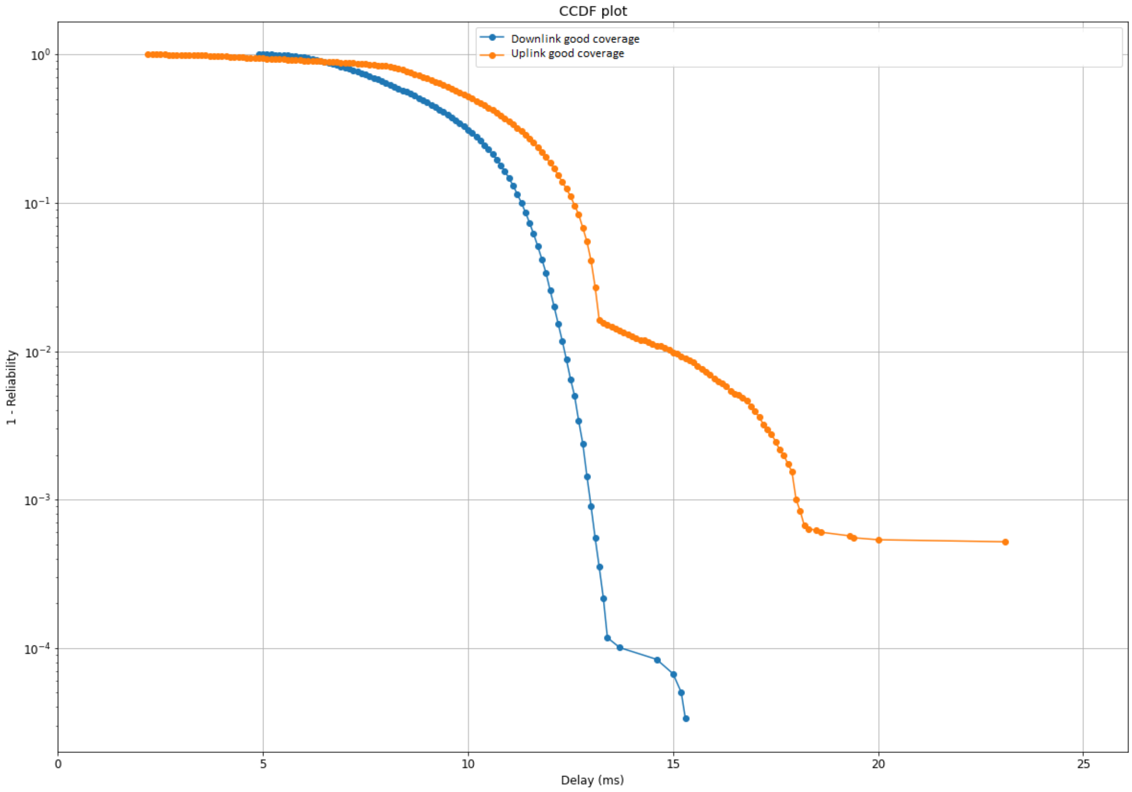

Figure 5.1: Reliability in good coverage for downlink (blue) and uplink (orange)

Figure 5.1 is a logarithmic CCDF, generated from Iperf result data. Y-axis represents reliability and X-axis latency between server and UE in milliseconds. It is used to visualize the 5G system's reliability, which was explained in section 3.1.2. The plot reveals under which latency certain reliability levels were achieved. Y-axis represents 1-Reliability and for example 1-10^{-4} is equal to 99.99% reliability level, 1-10^{-3} is equal to 99.9% reliability level and 1-10^{-2} is equal to 99% reliability level.

For downlink, it can be seen that 99.99% (1-10^{-4}) of the packets are received when one-way latency is less than about 14 ms. For 15 ms, the reliability is closing in on 99.999%, which has potential for some URLLC applications discussed in the chapter 3. There are also other indicators that fairly good reliability levels are met, such as steep curve where majority of the packets can be seen to be at the lower end of the latency. Also, the tail is very short which means there aren't outlier packets with high latency affecting the reliability in the end.

For uplink, it can be seen that 99.9% of the packets were received when the latency was less than about 18 ms. Unlike with downlink, at the end of the curve reliability stops improving when latency increases and it will be only slightly better than 99.9%, even at over 20 ms latency. The fact that reliability never reaches better percentages, even when the latency increases, refers to some lost packets.

It is expected however, that uplink has weaker performance than downlink. The main reason for that is uplink resource scheduling, which was explained in section 4.3.5. Generally, uplink has allocations less frequently than downlink, which causes higher latency to it. The higher latency from uplink resource scheduling causes the reliability to distribute much more unevenly than in downlink. Another reason for increased uplink latency are packet retransmissions, which will be discussed in sections 5.5.4 and 5.5.5.

5.1.2 Latency

In figure 5.2, each blue dot represents an individual packet's latency from the server to the client. It is constructed by capturing Wireshark logs from both server and the client and calculating each packet's arrival time difference. Iperf gives each packet a unique sequence number, and the X-axis represents each of the almost 60 000 packets in the 5 minute test. Y-axis represent each packet's latency in milliseconds.

From the plot it can be seen that most of the packets have latency between 6 - 12 ms and there are very few outliers. The latency pattern is looking promising for future URLLC deployments.

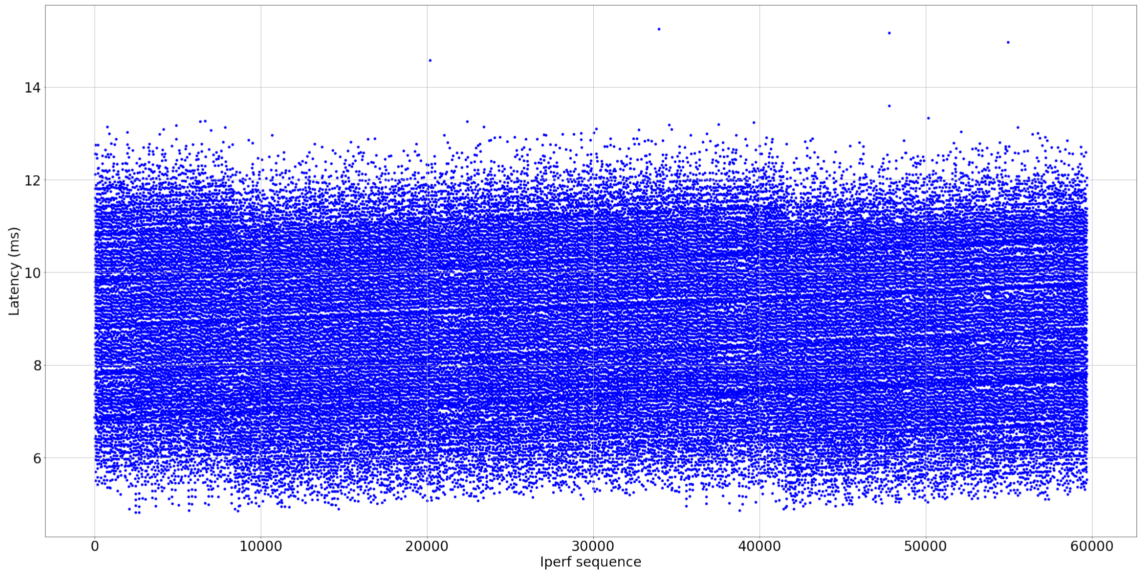

Figure 5.2: Good coverage downlink one-way latency on packet level.

In figure 5.3, uplink latency is plotted in a similar way as downlink above. It is good to note that compared to downlink, Y-axis has extended scale, because there are much more outliers. There are packets consistently appearing in two segments: Around 3 - 7 ms and around 7 - 13 ms, where majority of the packets are. Another big difference to downlink is that there are much more outliers, between 13 - 20 ms. These variating packet latencies can be explained by how much the UE had buffered packets after each 4 ms proactive grant.

The 300 second test was split into 3 second time intervals, and the average latency for each time interval can be seen in figure 5.4, for both downlink and uplink. As figures 5.2 and 5.3 showed, downlink has slightly better latency throughout the test run.

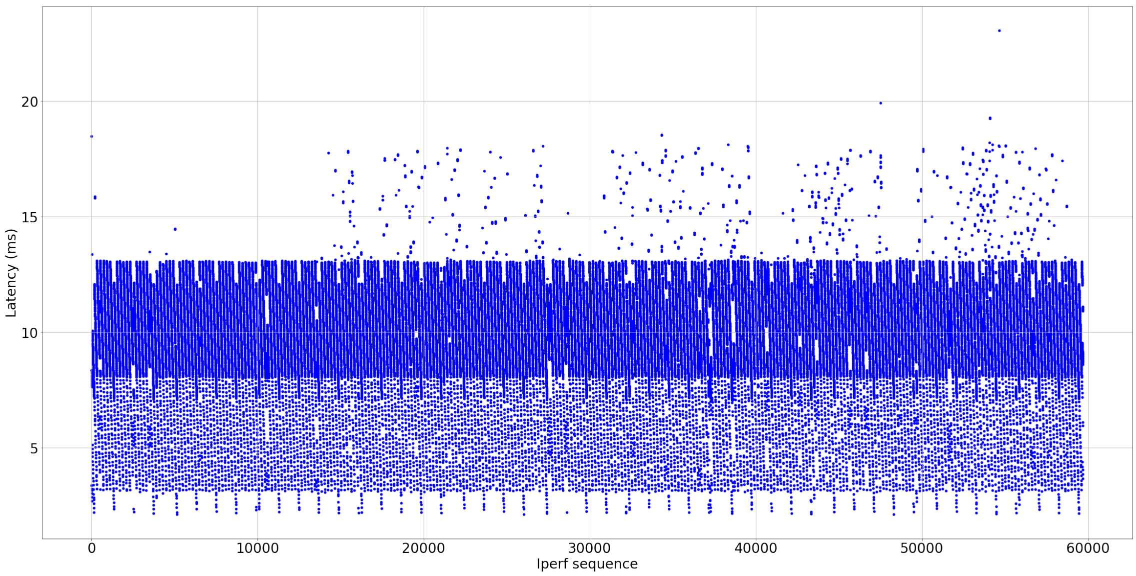

Figure 5.3: Good coverage uplink one-way latency on packet level.

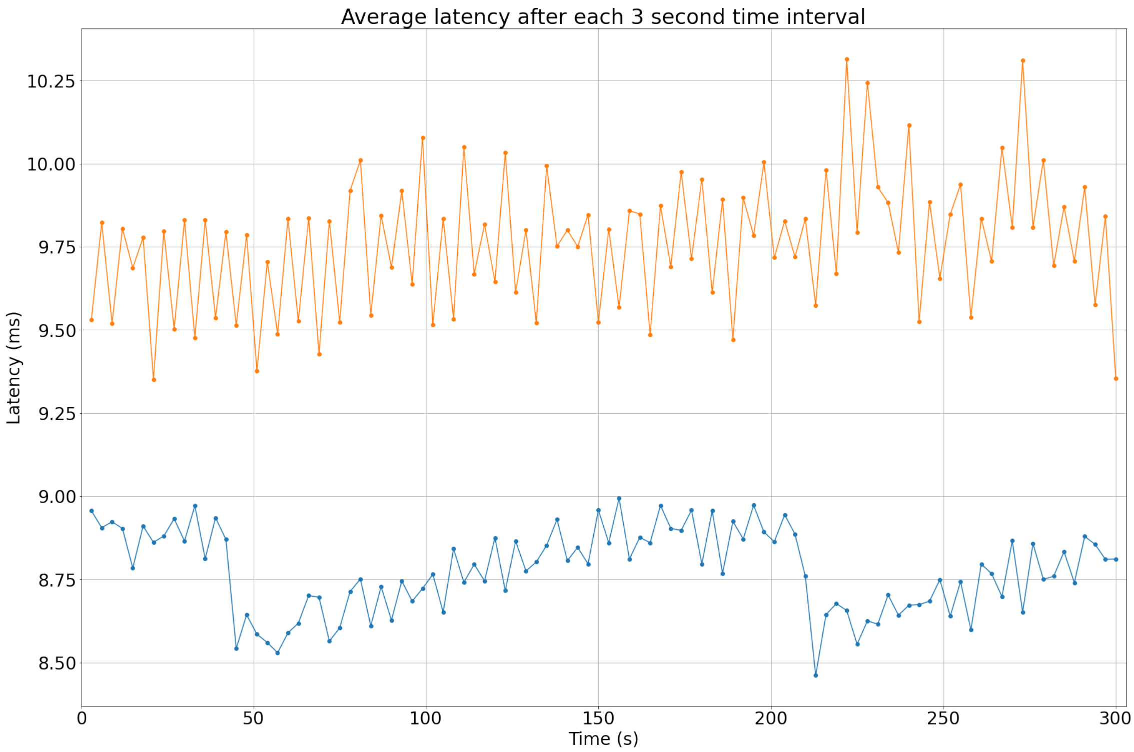

Figure 5.4: Good coverage each time interval average latency for downlink (blue) and uplink (orange).

5.2 Bad coverage

The bad coverage tests were conducted about 330 meters away from the cell, with a large building in-between. The RSRP was about -113 dbm, which can be considered bad and therefore suitable for this test. Just like with good coverage, there was no interference and no other network traffic than the Iperf UDP packet stream.

5.2.1 Reliability

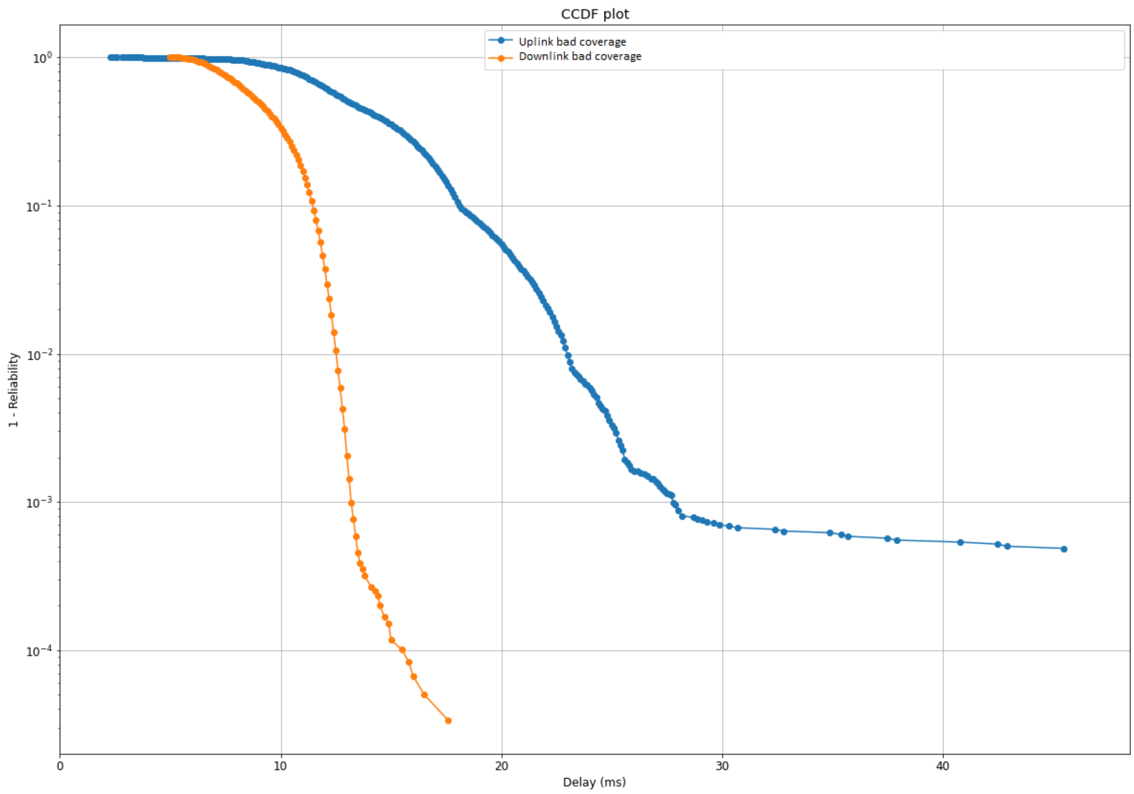

Figure 5.5: Reliability in bad coverage for downlink (orange) and uplink (blue).

Figure 5.5 is a logarithmic CCDF plot similar to the one which was used for good coverage earlier as well. In downlink, 99.99% of packets are received when latency is less than about 16 ms. When the latency rises above 16 ms the reliability is closing in on 99.999%, which has potential for some URLLC use cases. It can be seen that downlink is not much affected by the bad coverage. Most of the downlink packets are at the lower end of the latency axis and the curve is very steep with very short tail, which are very similar characteristics to good coverage results.

For uplink, the bad coverage had much bigger effect. There is a constant stream of packets all the way until about 28 ms, where the reliability is about 99.9%. There are some packets even around 50 ms with the same overall reliability, which causes the tail to be very long. The plot is very gradual, which means there is much more packet loss than in downlink and the packets' latencies have more scattering. Since at 28 ms latency the reliability is 99.9%, URLLC requirements are very far away.

5.2.2 Latency

Figure 5.6 represents each packet's latency between client and the server and it is constructed in a similar way as in good coverage section. From the plot it can be seen that most of the latencies are between 6 - 12 ms, with some outliers above 14 ms.

The latencies are good for bad coverage, and there is hardly any difference to good coverage latencies in figure 5.2. The overall pattern is very similar to good coverage, which means that coverage did not have much effect to the packets' latencies. These latency results look encouraging from URLLC point of view.

For uplink, the results can be seen in figure 5.7. The latency for uplink is clearly worse than downlink, with majority of the packets settling in around 10 - 18 ms. There is much more scattering even below 10 ms and well above 18 ms. There are packets appearing regularly all the way up until about 25 ms, and some outlier packets even with over 40 ms. Compared to uplink in good coverage in figure 5.3, it is clear that there is much more scattering, meaning there is much more variance in packet's latencies. In the good coverage most of the packets have latency between 3 - 13 ms, which is significantly better than the bad coverage's 10 - 18 ms. It is

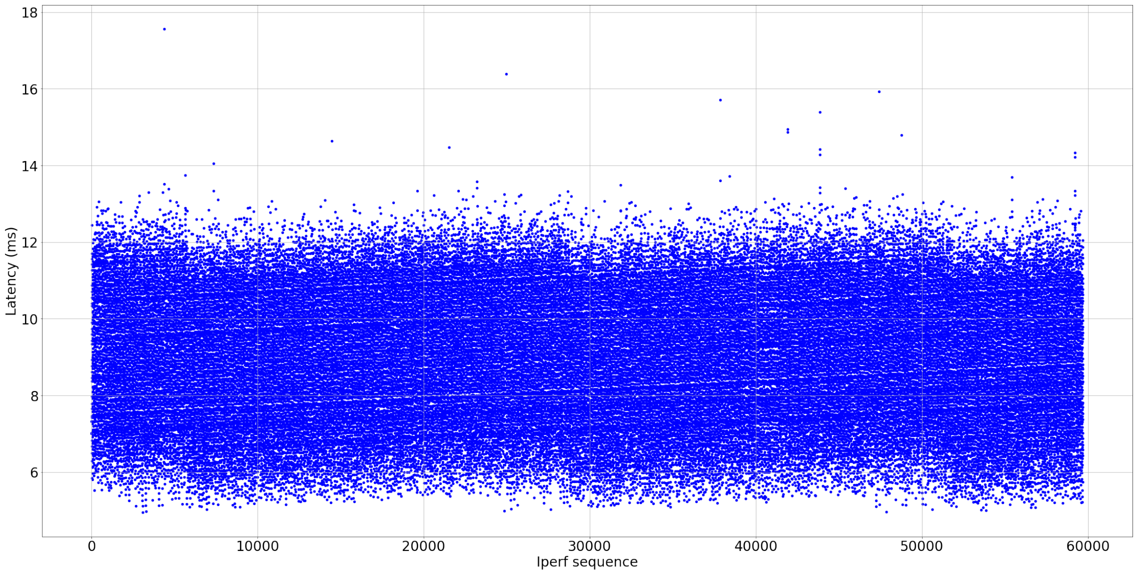

Figure 5.6: Bad coverage downlink one-way latency on packet level.

obvious that coverage had an impact to the one-way latency in uplink.

The average latencies for each time interval for downlink and uplink can be seen in figure 5.8. When comparing the latencies to good coverage in figure 5.4, it is noticeable that for downlink the latencies are just slightly bit better in good coverage than in bad coverage. This verifies the earlier observation that coverage does not have a big impact for latency in this test scenario for downlink.

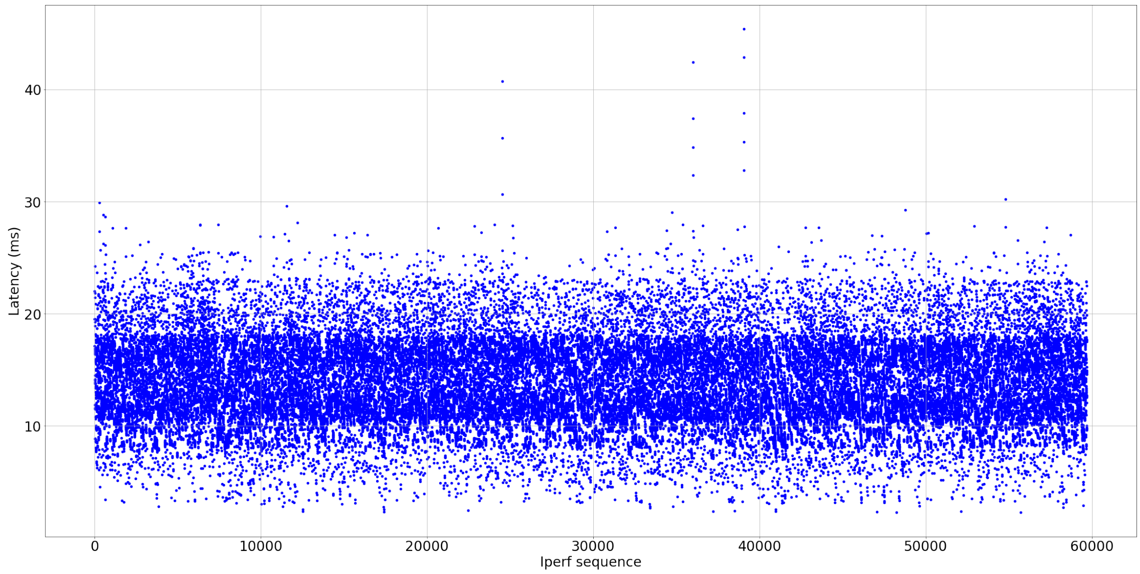

Figure 5.7: Bad coverage uplink one-way latency on packet level.

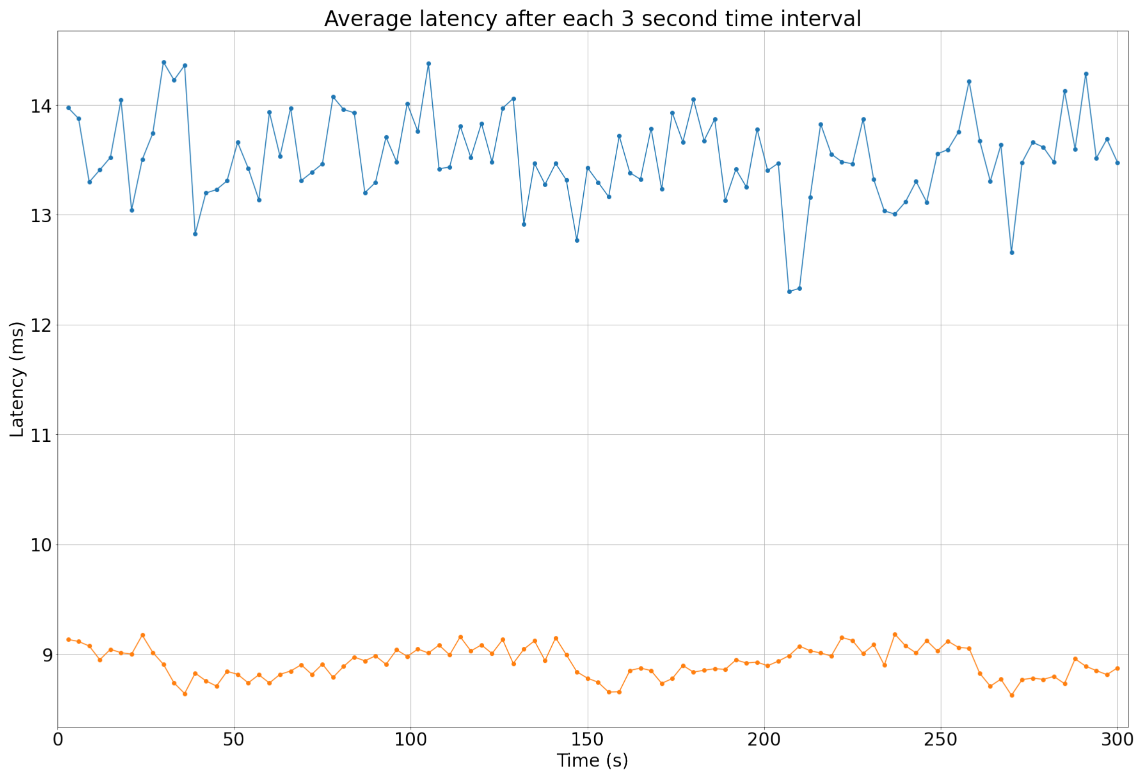

Figure 5.8: Bad coverage each time interval average latency for downlink (orange) and uplink (blue).

5.3 Good-to-bad coverage

Good-to-bad coverage tests are driving tests, in which the test starts at the location where good coverage measurements were made (-75 dbm) and ends where the bad coverage measurements were made (-113 dbm). The cell stayed the same for the whole test. The driving distance was about 450 meters, and the test was the same 5 minute Iperf UDP stream test as before. The goal in this test is to see how the cell reacts to increasingly worsening coverage. If the worsening coverage and mobility causes a lot of scattering to the latencies, it will impair the reliability. Different V2X operations are real life use cases similar to this test.

5.3.1 Reliability

Figure 5.9 is the CCDF plot for good-to-bad coverage reliability. For downlink, the reliability looks what is expected. Up until about 13 ms, the reliability is 99.9%, which is equal to good coverage in figure 5.1. The difference happens at 99.99% reliability, where latency is 18 ms for good-to-bad, and 14 ms in good coverage. At bad coverage in figure 5.5, downlink had 99.99% reliability under 16 ms latency, which means bad coverage achieved 99.99% reliability with slightly better latency than good-to-bad coverage. This could be explained with car driving around in the bad coverage area, which can result the RSRP to go briefly worser than the -113 dbm, which was constant in the bad coverage test. Also, little variance is always expected between test runs.

For uplink, the reliability is much worse than downlink, which was expected from the good coverage and bad coverage tests. At the beginning of the plot, the good-to-bad results do not look anything like good coverage uplink results. It can be seen by looking at the latency at 99.9% reliability, which was 18 ms for good coverage and about 33 ms for good-to-bad coverage. In fact, there are no similarities to good coverage uplink reliability at all. On the other hand, when comparing the run to

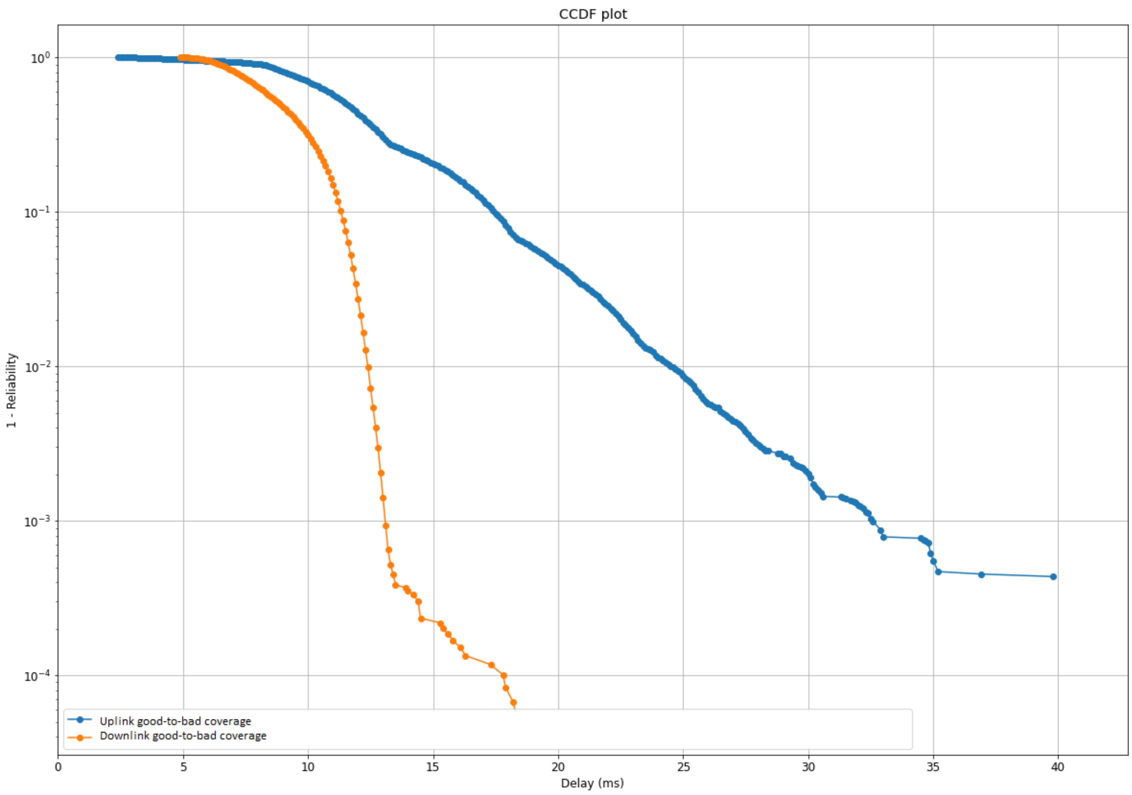

Figure 5.9: Reliability in good-to-bad coverage for downlink (orange) and uplink (blue).

bad coverage uplink, their reliability looks very similar.

The plot reveals that uplink had a lot of scattering between packet latencies with substantial number of packets having high latency, thus making reliability suffer and growing such a huge tail in the plot. Also, the reliability didn't improve much from 99.9% even with higher latencies, which refers to packet loss. The only difference between this test and the earlier ones was driving, which means mobility will cause a lot of scattering for latencies in uplink.

5.3.2 Latency

Figure 5.10 plots the good-to-bad one-way latency in downlink, similarly as in good and bad coverage earlier. The plot looks very similar to good and bad coverage results, verifying the earlier observation that coverage had very little effect on downlink one-way latency. This plot also verifies that driving did not have an impact on the latencies for downlink.

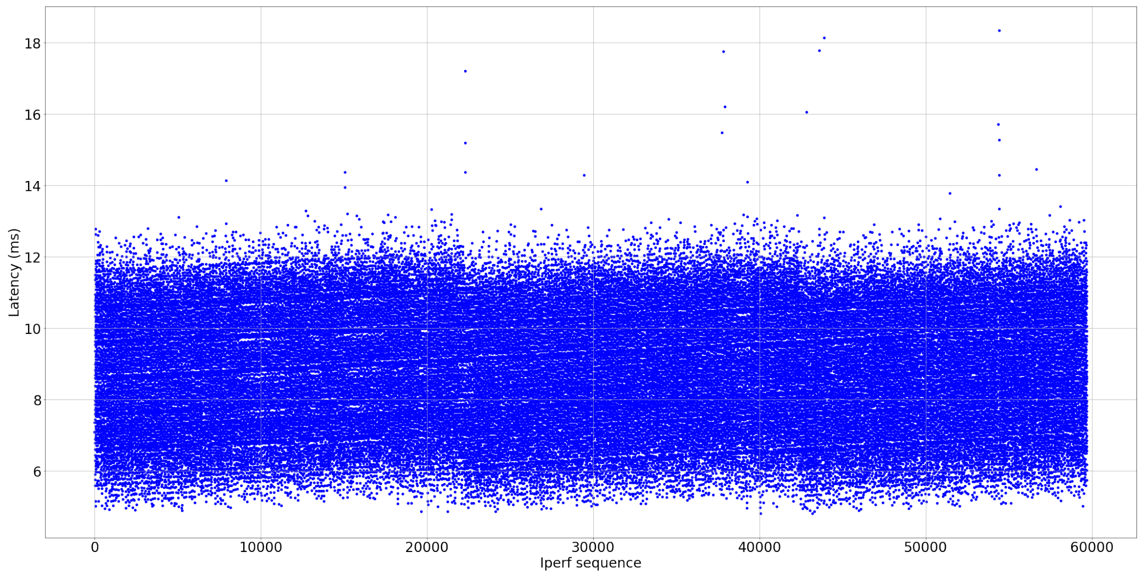

Figure 5.10: Good-to-bad coverage downlink one-way latency on packet level.

Figure 5.11 plots the good-to-bad coverage's one-way latency in uplink. It visualizes well on how the driving proceeded and how the latency increases during the 5 minutes. For the first 30 000 packets which is half of the total packets, the results look quite similar to good coverage results in figure 5.3. There are some gaps in the lower end of the latency and somewhat more packets above the 13 ms which were very rare in good coverage. However, these are expected and explained by driving and the increasingly worsening coverage. The latencies are surprisingly good, because driving towards worsening coverage started immediately but the latencies stay close to good coverage results until midway.

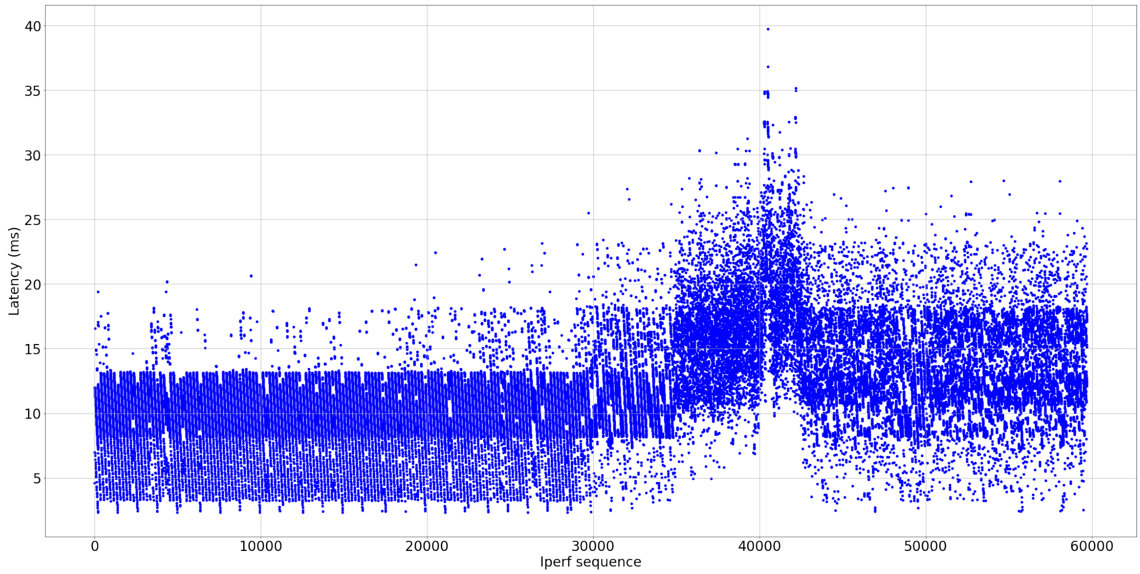

Figure 5.11: Good-to-bad coverage uplink one-way latency on packet level.

After halfway, the latency increases abundantly and goes shortly even above bad coverage latencies. This scattering explains the uplink's impaired reliability seen in the reliability plot in figure 5.9. As explained earlier, it is likely that the coverage has been momentarily even worse than in stationary bad coverage test. There was also a big number of packet retransmissions increasing latency, which will be discussed in section 5.5.5 and can be seen in figure 5.19. In the end, the latencies settle in very similarly as in bad coverage test, around about 10 - 18 ms with a lot of scattering.

The average latencies for each time interval for both downlink and uplink can be seen in figure 5.12. For downlink, at around 9 ms average latency interval it has very much potential for URLLC implementation. For uplink, the average latency rises above 20 ms for one interval, which is very far away from URLLC requirements and many V2X applications.

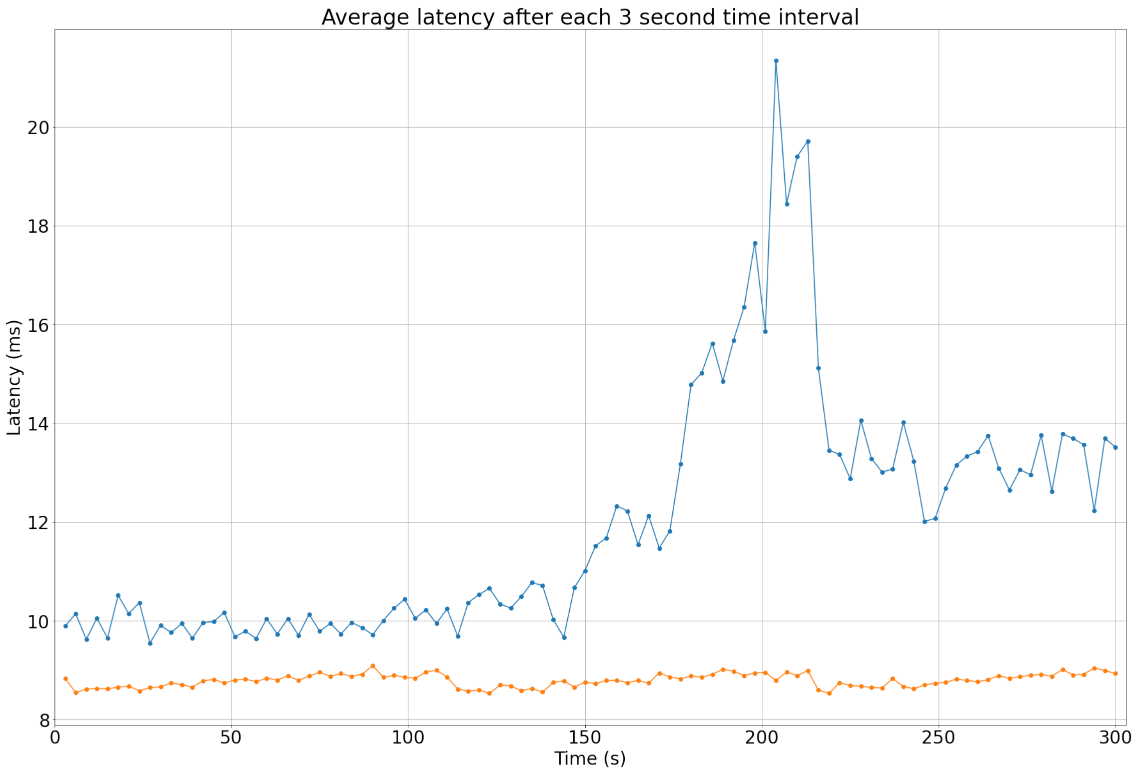

Figure 5.12: Good-to-bad coverage each time interval average latency for downlink (orange) and uplink (blue).

5.4 Handover

Handover is a controlled and instantaneous change of radio connection between UE and the network, with the target of having no lost data nor interruptions. Handover is triggered by UE measurement data to the network, when it indicates that better coverage and channel quality can be achieved with other cell. In this test scenario and in 5G overall, handovers are hard, which means connection exists to only one cell at a time and packets are forwarded from source cell to the target cell. Only after resources are prepared in the target cell the network commands UE to move to the new 5G cell.

Handover tests are driving tests, in which the whole test is done in good coverage area. In handover test the transmitting or receiving 5G cell is changed by driving from one cell's coverage area to another. The Iperf UDP stream is running during the cell change and the point is to find out how the cell change affects reliability and latency. The handover done for this test is intra-frequency handover, which means the frequency band stayed same after changing the cell.

The test lasted for 5 minutes with total driving distance of about 450 meters. The cell change happened after about 150 meters of driving. As section 5.3 showed, driving had some impact to reliability and latencies, but it is expected that the results look somewhat similar to good coverage in section 5.1. The goal is to see is there a spike in packet latencies, does the cell change impact reliability and is there packet loss caused by the cell change.

Basically, any URLLC application where the change of 5G cell can happen due to change of coverage area or exceeding cell capacity are real life use cases similar to this test.

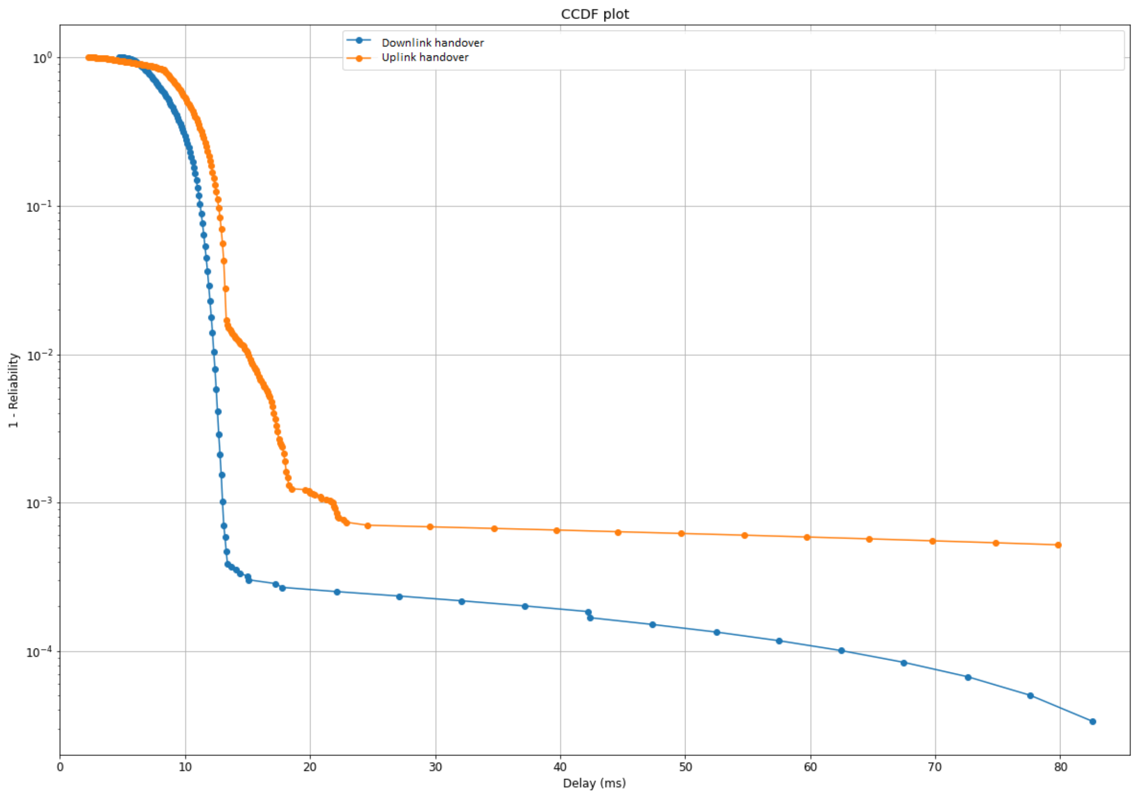

Figure 5.13: Handover reliability for downlink (blue) and uplink (orange).

5.4.1 Reliability

Figure 5.13 is the CCDF plot for handover reliability, similar as the ones used in earlier sections. For downlink, the reliability looks very good and similar to good coverage at figure 5.1 until about 15 ms, where reliability sits between 99.9% and 99.99%. At 15 ms, the reliability is just slightly worse than in the original good coverage test, and this difference is obviously explained by driving. After 15 ms, the plot grows a huge tail, with latency rising over 80 ms. This is explained by the handover, when the cell changes and the packets to be transmitted need to reroute to another cell. The latency is caused by the network preparing the 5G cell before commanding the UE to move to the new target cell to receive data.

For uplink, the reliability also looks fairly similar to good coverage in figure 5.1.

The 99.9% reliability is achieved at about 22 ms latency, which means driving had a minor impact to the reliability. After about 22 ms, the latency rises all the way up to 80 ms, growing a huge tail in the plot, just like with downlink. The handover had very similar impact to both downlink and uplink.

Because the uplink reliability reached just a bit over 99.9% at 80 ms and never improved after that, there was some packet loss. Given the fact that every uplink test done earlier also ended up exactly with a bit over 99.9% reliability with varying latencies, it is unlikely that handover caused the packet loss.

5.4.2 Latency

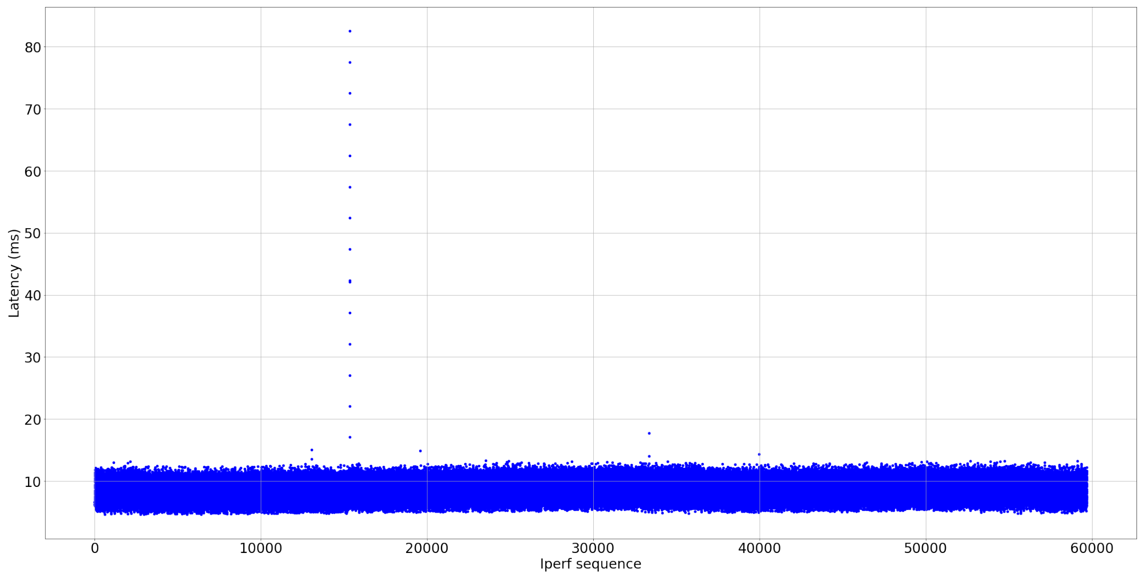

Figure 5.14: Handover downlink one-way latency on packet level.

Figure 5.14 plots the handover one-way latency for downlink in similar way as in earlier sections. It is very clear that the handover happened in at about packet number 15 000, which caused few packet's latencies to increase massively. These latency increases happened when the network prepared the new target cell to transmit data and reroute the packets.

The handover latencies caused the plot to have very big scale in the Y-axis, but it still can be seen that the latencies are very similar to downlink good coverage latencies in figure 5.2, as expected.

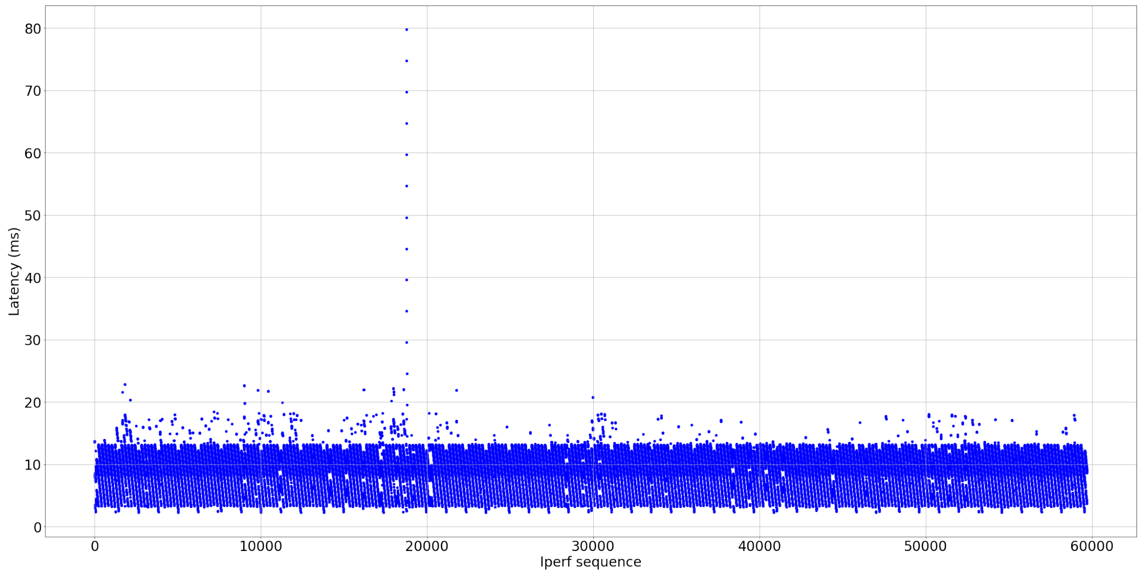

Figure 5.15: Handover uplink one-way latency on packet level.

Figure 5.15 plots the handover one-way latency for uplink. It looks very much as expected. Just like with downlink, it has very large scale due to the handover latencies. Otherwise, it looks very similar to the good coverage uplink in figure 5.3, with just a bit more scattering over 15 ms.

The average handover latencies for each time interval can be seen in figure 5.16. The spike caused by the handover is clearly visible in both runs. Compared to good coverage average latencies in figure 5.4, these results are just slightly higher, which is explained by driving.

For URLLC, the spike to latency is hard to avoid. This will cause problems to reach the strictest requirements, because in real-life mobility scenarios handovers are occurring events and constantly happening spikes in latency make it impossible to reach 99.999% reliability levels in URLLC latency standards. This issue will be discussed further in section 5.5.2.

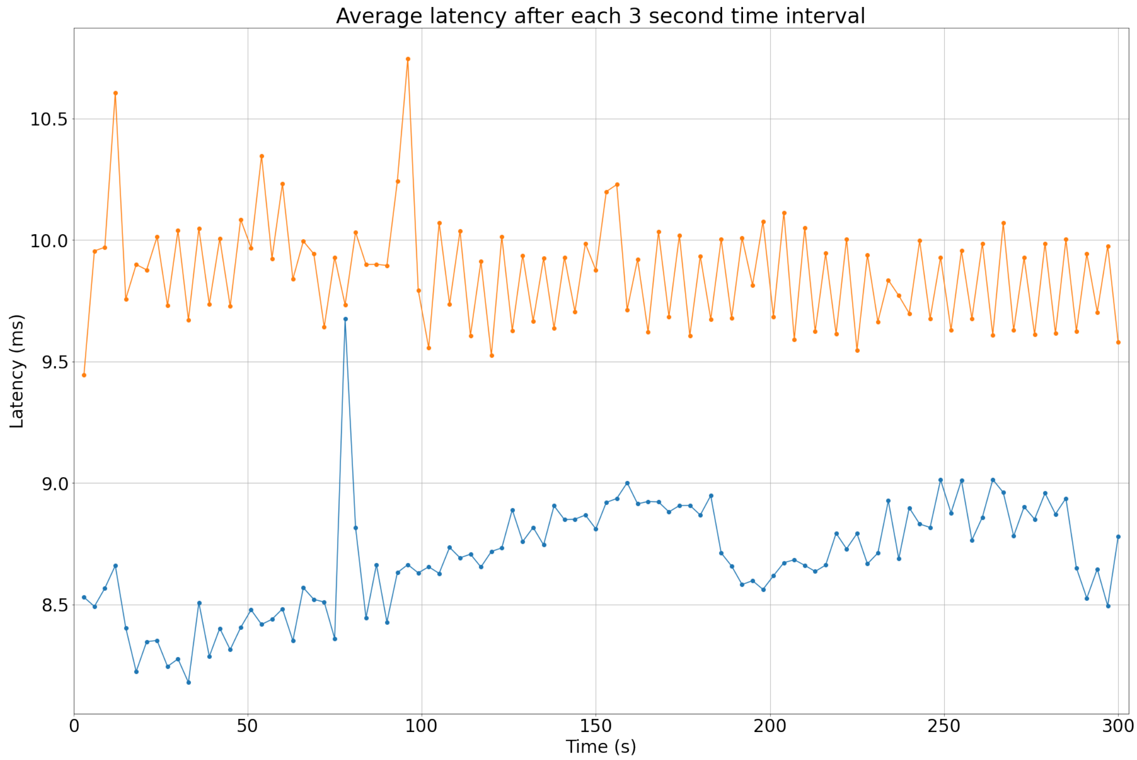

Figure 5.16: Handover each time interval average latency for downlink (blue) and uplink (orange).

5.5 Results analysis

After running all four tests for both downlink and uplink, it is clear that downlink's reliability performed much better, as expected. The tests were run in ideal conditions, were executed successfully and as planned. The results indicate the achieved reliability and one-way latency levels for each test, and will be useful in future for analysing what kind of impact different URLLC features have had after implementation. In this section, the results of all tests will be analysed.

5.5.1 Achieved reliability levels

Table 5.1: Achieved reliability levels for downlink. Latencies are rounded to the closest integer.

	99.99%	99.9%	99%
Good coverage	14 ms	13 ms	12 ms
Bad coverage	16 ms	13 ms	12 ms
Good-to-bad coverage	18 ms	13 ms	12 ms
Handover	63 ms	13 ms	12 ms

Table 5.2: Achieved reliability levels for uplink. Latencies are rounded to the closest integer.

	99.99%	99.9%	99%
Good coverage	Never reached	18 ms	15 ms
Bad coverage	Never reached	28 ms	23 ms
Good-to-bad coverage	Never reached	33 ms	24 ms
Handover	Never reached	22 ms	16 ms

Tables 5.1 and 5.2 are made from the data from the reliability plots. They indicate under which latency a certain reliability level was achieved. The latencies

are rounded to the closest integer. Uplink has status "Never reached" under 99.99% reliability, simply because of packet loss it never reached such reliability level under any latency.

For downlink, there are zero lost packets in all of the test runs and besides handover, the 99.99% reliability is achieved under 20 ms latency in every run. The results look encouraging from URLLC point of view. Undoubtedly, the tight URLLC requirements will cause problems which are not even realized yet nor discussed in this thesis. Nevertheless, it can be expected that with shortened TTI duration, Non-slot based scheduling and DAPS handover reliability and latency can be improved significantly from this level. Therefore, for downlink, achieving URLLC standards is realism, at least in the test conditions conducted for this thesis.

The URLLC requirements do not specify what is for downlink and what is for uplink, which means they are same for both. Given that for uplink 99.9% reliability is achieved between 18 ms - 33 ms latency, uplink is quite far away from URLLC standards. The main reason why uplink performs much worse than downlink is uplink resource scheduling. Therefore, with uplink, implementing grant-free transmission is crucial to reach the URLLC requirements. Obviously, other techniques in the URLLC toolbox should be implemented as well, but grant-free transmission is a technique directly designed to resolve the increased latency issue from uplink resource scheduling. Uplink never reaches 99.99% or higher reliability under any latency because of packet loss. Therefore, for URLLC deployment the cause for uplink packet loss is analysed in chapter 6.

5.5.2 DAPS handover

In the handover test, 99.99% reliability was achieved under 63 ms latency for downlink and uplink had similar results as well. The peak in latency is hard to avoid, and it would repeat constantly in a real life V2X scenario impairing reliability effectively for URLLC.

In release 16, 3GPP has specified a technique for URLLC to mitigate handover interruptions. It is called Dual Active Protocol Stack (DAPS) handover. In DAPS handover, the source gNB connection is maintained after receiving of RRC (Radio Resource Control) reconfiguration message (handover command) for handover. The source gNB connection will not be discarded until there is a successful random access procedure to the target gNB. Because connection to the source gNB is maintained while target gNB is initiated, handover interruptions can be reduced to avoid spike in latency and therefore to meet URLLC standards. [40]

For downlink, this means that during handover execution the UE receives data from both source and target gNBs. The source gNB connection is not released until the target gNB gives a direct release command. Similarly in uplink, the UE continues data transmission to the source gNB until random access procedure, after which the UE switches data transmission to the target gNB. Before handover execution, the UE keeps transmitting layer 1 CSI, HARQ and layer 2 RLC feedback and information about HARQ and RLC retransmissions to the source gNB. [40]

If a DAPS HO fails, the UE continues the connection with source gNB and reports DAPS handover failure from the source gNB without triggering a new RRC reconfiguration message [40]. The DAPS flow is visualized in figure 5.17.

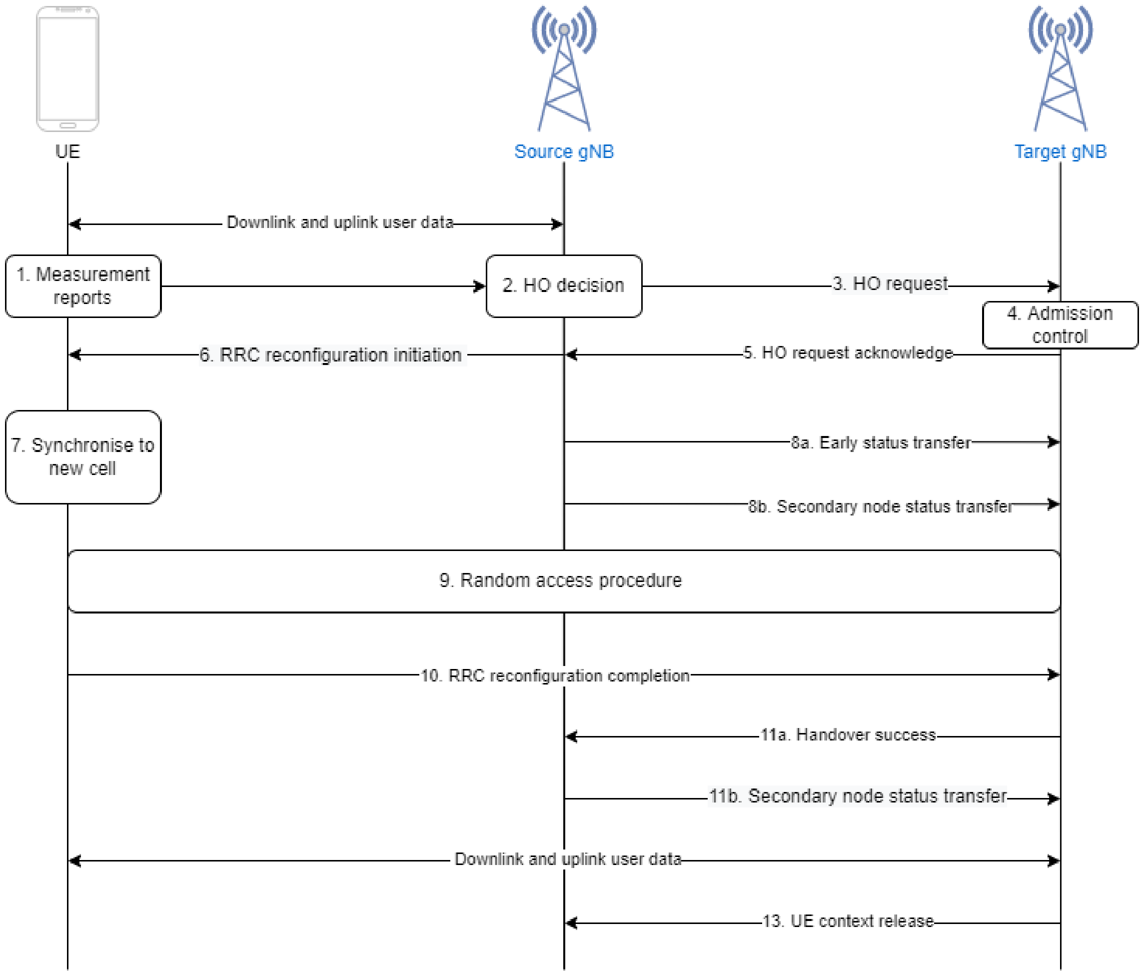

Figure 5.17: Dual Active Protocol Stack handover scenario [40].

5.5.3 One-way latency performance

The average latency, out of all packets, for downlink and uplink for each test is listed in table 5.3. It should not be mistaken with the average latency intervals in earlier sections, which indicates average latency after each 3 second time interval. In this case the average results do not directly tell much, because URLLC's tightest requirements of over 99.999% reliability can be unreachable with just a few outlier packets with high latency. However, these results are a good indicator on how coverage, driving, uplink resource scheduling and packet retransmissions affect to latency and they also directly show how much downlink and uplink latency differ from each other.

Table 5.3: Average one-way latency for each test

	Good coverage	Bad coverage	Good-to-bad coverage	Handover
Downlink	8.75 ms	8.90 ms	8.77 ms	8.67 ms
Uplink	9.76 ms	13.53 ms	12.00 ms	9.86 ms

5.5.4 Link adaptation algorithm and MCS

According to table 5.3, downlink has practically the same latency for each test and it is not affected by coverage or mobility. This is because of link adaptation algorithm, which allocates MCS (Modulation and Coding Scheme). MCS correlates on radio signal quality, by defining how many useful bits can be transmitted for a given time interval in which signal and its subcarrier is modulated (spectral efficiency). This time interval is called resource element, and better quality means higher MCS which results in more useful bits to be transmitted in the resource element. Bad signal means lower MCS, which results in less useful data to be transmitted within a resource element. To maintain block error rate (BLER) not raise above the configured 10% value in constantly changing radio channel quality, gNB dynamically allocates

MCS with link adaptation algorithm. After this, the allocated MCS is signalled to the UE. [41] [42]

In the scenario used for this thesis, for downlink the base station can use all available resource blocks to transmit the small 255B packets. This is because with 100 MHz channel bandwidth and 30 KHz SCS used in the tests, for downlink the minimum possible RBG (Resource Block Group) size is 16 PRBs (Physical Resource Block) [39]. When the base station uses 16 PRBs to send the small 255B packets, it can use low MCS with high amount of forward error coding in the packet transmissions, which will result in fewer packet retransmissions. For uplink, there will be less PRBs available than downlink, which results in higher amount of retransmissions. The target BLER value for link adaptation algorithm was 10%, and for downlink the actual value was zero, which can be seen for good-to-bad coverage test in figure 5.18. The uplink retransmissions, which are about 10%, can be seen for good-to-bad test coverage test in figure 5.19.

For URLLC, it is desirable to support lower BLER target. The lower BLER can be achieved by reducing spectral efficiency in modulation and coding scheme. Lower MCS results in worse data rates for application, but reduces retransmissions and therefore latency. [23] In the upcoming URLLC features, link adaptation will be adjusted according to 5G quality class (5QI) parameters. This means that in URLLC different applications can be mapped to have different BLER target values, so there is no one universal BLER target which is used for example with EMBB. This results in reliability optimized BLER values. There are also upcoming functionalities to adjust link adaptation automatically.

5.5.5 Packet retransmissions

Uplink has more variance in latencies than downlink, and the main reason for that are packet retransmissions and resource scheduling. The results do correlate on coverage conditions and mobility. These latencies have much potential to decrease, when for example grant-free transmission will be implemented. It can be also seen that coverage had bigger impact to average latency than mobility.

To illustrate how retransmissions correlate with latency, the good-to-bad coverage test's retransmissions for downlink and uplink were plotted using network analytics software Nemo Analyzer. Good-to-bad coverage test was chosen, because it has the most variating coverage conditions. The good-to-bad BLER for downlink, which is zero, can be seen in figure 5.18. Uplink retransmissions, which are around 10%, can be seen in figure 5.19. Downlink BLER and uplink retransmissions are practically a same thing in this context and can be considered comparable to illustrate changing one-way latency values in different coverage conditions. When compared to average latency intervals in figure 5.12, the correlation can be seen. When downlink has latency around 9 ms, it has practically zero BLER. Uplink has retransmissions even up to 50% at the edge of the cell. According to Nemo, there are some packets retransmitted twice, but not more than that. In 5G protocol stack, packet retransmissions can be done at MAC and RLC layer. At MAC layer the retransmissions are HARQ retransmissions which were explained in chapter 3. In the good-to-bad coverage test, all retransmissions happened at the MAC layer which means they are HARQ. The HARQ retransmissions were configured to have cap at 5, which means if there would have been more retransmissions they would have been retransmitted at the RLC layer.

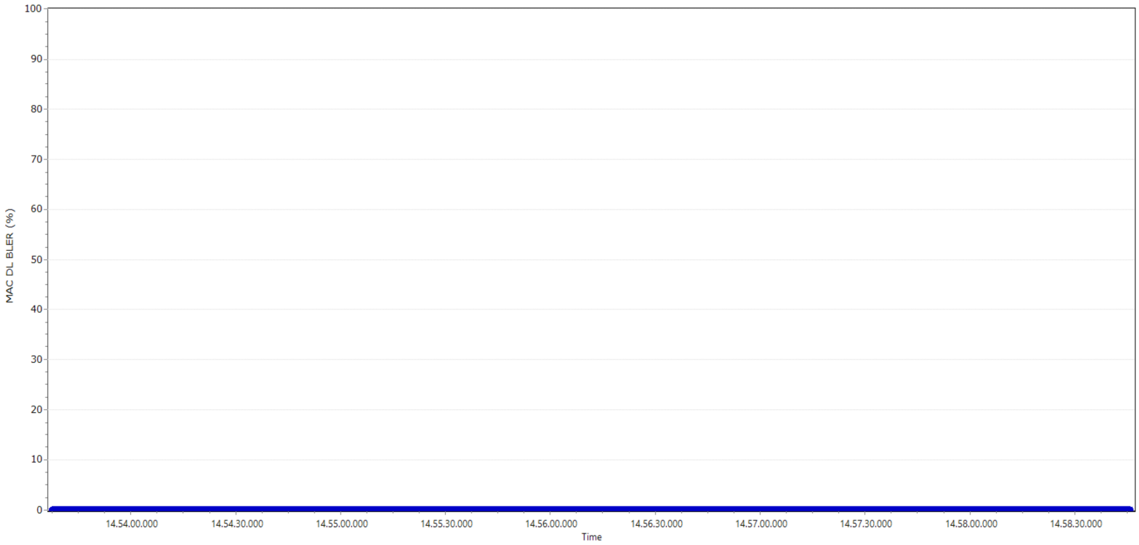

Figure 5.18: Good-to-bad downlink BLER.

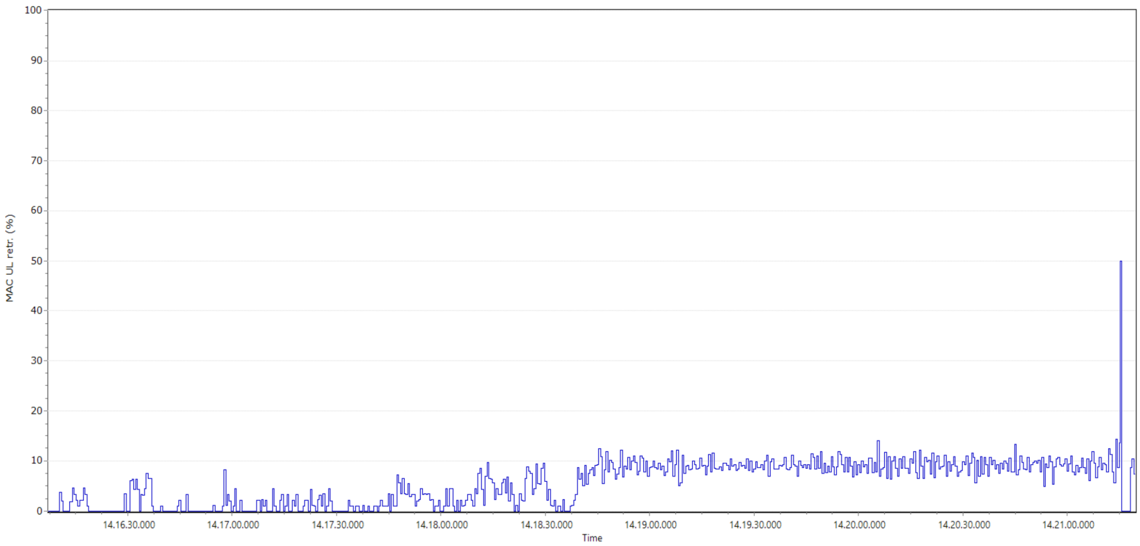

Figure 5.19: Good-to-bad uplink retransmissions.

6 Packet loss analysis

For URLLC requirements, packet loss has huge impact to reliability and should be kept to absolute minimum. This is because reliability is defined as successfully transmitted packets in a given time constraint. When a packet is lost, certain reliability levels will become unreachable. This chapter covers a case study where packet loss from earlier tests is documented and analysed. Then, an attempt is made to identify when, where and why the packet loss happened.

As discussed in section 4.3.2, to detect lost packets Wireshark was installed on cliend side and Tshark on server side. After capturing packets from both endpoints, all packets were identified using an Iperf dissector plugin, which gives each packet a unique sequence number in the Iperf data stream. When all packets were uniquely identified, a simple Python script was used to detect which packets never arrived.

In the last chapter, it was noticed that all packet loss happened in uplink transmission. In table 6.1, the total amount of lost packets for each 5 minute test is listed. From this table, it can be seen that about same number of packets were lost in each test, so coverage or mobility did not have much impact to the packet loss.

Based on that, some 5 minute tests were ran again on good coverage. In first test the packet burst sizes were set to 100 and 300 for their own 5 minute tests. In second test the packets were captured from client and server, but also from the base station this time. In third test, a different UE was tried to see if the askey 5G SA modem had an impact to the packet loss. These tests were included to have

Table 6.1: Total amount of lost packets in each test's uplink transmission.

Test	Total amount of lost packets (out of 59 700)
Good coverage	29
Bad coverage	27
Good-to-bad coverage	24
Handover	29

a broader view on the packet's path, packets' behavior, and to narrow down the factors which can cause packet loss.

6.1 Changing the packet burst size

In the first attempt, the 5 minute good coverage uplink test was ran again, in identical conditions compared to section 5.1, but this time with 100, 200 and 300 packets per second (pps). In these tests, 100 pps resulted in about 30 000 total packets, 200 pps in about 60 000 total packets and 300 pps in about 90 000 total packets. This was done to find out how much impact increased traffic has to packet loss, and does the UE get congested with higher pps amounts. It is expected that 100 pps has slightly less packet loss and 300 pps has slightly more packet loss than the amounts listed in table 6.1, in which the pps was 199. However, if 100 pps results in very minimal or zero packet loss and 300 pps results in a major increase with packet loss, it means that packet loss is caused by network congestion and the congested hardware should be identified.

Table 6.2 lists the total amount of lost packets with a given burst size. In total, two runs were made for each burst size. The packet loss increases accordingly when the traffic increases, which means the packet loss is scaling with the traffic size so network congestion is unlikely to be the root cause for lost packets. These results show that the packet loss is what was expected for each burst size.

Table 6.2: Total amount of lost packets with different burst sizes, in 5 minute good coverage uplink test.

Packets per second	Lost packets (Run 1)	Lost packets (Run 2)
100	9	9
200	22	27
300	34	43

6.2 Base station monitoring

The 5 minute uplink good coverage test was ran again exactly same way as in section 5.1, but with this time the packet traffic was also monitored in the base station. This was done to narrow down the packets' path, and to find out do the packets get lost between the UE and base station, base station and the server, or both.

The pcap files were collected from UE, base station, and server. With Wireshark analysis, it can be seen that 27 packets were lost, so the plan is to first compare the collected packets between UE and base station. In total, all 27 lost packets were identified between UE and the base station, meaning 100% of the packet loss happened between those. The lost packet's IDs, sequence numbers, arrival time at the UE, epoch times and basic info can be seen in figure 6.1

The fact that all packet loss happened between UE and base station, means it is possible that UE drops packets somewhere at the beginning of the transmission. For example, the problem can be at the hardware with chipset or at the ethernet port. It is also possible that the BTS drops packets when receiving them from the UE. Based on that, a Layer 2 BTS RND team was asked to track the packets in PDCP (Packet Data Convergence Protocol), RLC (Radio Link Control) and MAC layers. There was no packet loss identified, so it is very likely that UE is the cause for packet loss.

	Total lost packets	Iperf3 sequence	Time	Arrival Time	Epoch Time	Info
1241	1	1242	6.232214	Feb 7, 2022 15:28:37.050036826 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
2489	2	2490	12.503570	Feb 7, 2022 15:28:43.321393068 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
3718	3	3719	18.679448	Feb 7, 2022 15:28:49.497271083 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
9889	4	9890	49.689508	Feb 7, 2022 15:29:20.507330926 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
13601	5	13602	68.342767	Feb 7, 2022 15:29:39.160590032 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
14847	6	14848	74.604074	Feb 7, 2022 15:29:45.421897357 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
16079	7	16080	80.795027	Feb 7, 2022 15:29:51.612850491 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
22469	8	22470	112.905572	Feb 7, 2022 15:30:23.723395251 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
27211	9	27212	136.734728	Feb 7, 2022 15:30:47.552551170 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
28446	10	28447	142.940757	Feb 7, 2022 15:30:53.758580637 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
30929	11	30930	155.418146	Feb 7, 2022 15:31:06.235968806 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
33417	12	33418	167.920655	Feb 7, 2022 15:31:18.738478591 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
35905	13	35906	180.423169	Feb 7, 2022 15:31:31.240991701 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
36112	14	36113	181.463354	Feb 7, 2022 15:31:32.281177126 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
38385	15	38386	192.885481	Feb 7, 2022 15:31:43.703304223 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
39631	16	39632	199.146787	Feb 7, 2022 15:31:49.964610431 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
40865	17	40866	205.347792	Feb 7, 2022 15:31:56.165614642 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
43541	18	43542	218.795028	Feb 7, 2022 15:32:09.612851478 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
44574	19	44575	223.985984	Feb 7, 2022 15:32:14.803806677 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
45819	20	45820	230.242265	Feb 7, 2022 15:32:21.060088077 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
47053	21	47054	236.443271	Feb 7, 2022 15:32:27.261093738 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
49530	22	49531	248.890516	Feb 7, 2022 15:32:39.708339408 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
50762	23	50763	255.081449	Feb 7, 2022 15:32:45.899272232 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
51992	24	51993	261.262364	Feb 7, 2022 15:32:52.080187238 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
55709	25	55710	279.940757	Feb 7, 2022 15:33:10.758580331 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
56951	26	56952	286.181965	Feb 7, 2022 15:33:16.999788591 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227
58192	27	58193	292.418128	Feb 7, 2022 15:33:23.235950885 FLE Standard Time	1.644241e+09	5001 > 34326 Len=227

Figure 6.1: The packets identified to be lost between UE and BTS, after running the 5 minute uplink good coverage test again. Arrival time is at the UE.

6.3 Trying different UE

Because all packet loss happened when UE transmitted data to the base station, the root cause for packet loss may be the Askey 5G SA modem. Based on this, the UE was changed to a commercial Huawei P40 5G SA phone. With a commercial phone the data goes through USB interface to the Linux PC causing some delays and reduced throughput, but it is still viable option for packet loss analysis. The 5 minute good coverage uplink tests were ran again with burst sizes of 100, 200 and 300, with two runs for each burst size.

After all tests were ran two times for each burst size, there was zero lost packets detected. This means that the packet loss was possibly caused by the Askey 5G SA modem. This is a surprising finding, because the Askey modem's firmware has been updated recently, there are no known sources of interference and overall it should be an excellent device for this type of testing.

Obviously, Huawei P40 is a very limited UE for this type of testing, so a new UE compatible for URLLC should be chosen and its performance should be verified on the field. A Nokia FastMile 5G Gateway was chosen as next candidate UE and the good coverage 5 minute uplink test was run with it, with 199 packets per second. This time, there was about 20 - 30 lost packets for each test run. Because two different 5G modems were steadily losing packets, but a commercial phone using USB interface to transmit data did not lose any packets, it is possible that the ethernet port between the Linux PC and Askey modem can be the issue.

The Linux PC was changed to another Linux computer, to see if the ethernet port or some other hardware component in the PC had an impact to the packet loss. The same 5 minute uplink test with 199 packets per second was ran again, and the number of lost packets was almost identical as with the old Linux PC.

Because the change of Linux PC did not resolve the issue, the next thing to try was to update the original Askey UE's firmware. The UE was using firmware released in October 2021 and the latest available firmware for the model was released in March 2022, so there was an upgrade available. After downloading and installing the firmware, the 5 minute uplink reliability test was ran again twice. Again, there was about 30 lost packets for both runs which means the firmware was not the issue.

6.4 Changing radio conditions

So far, all the tests were ran on same radio using the n77 frequency band. This radio is identified as a good candidate for future URLLC deployment and is now used specifically for capacity and performance testing. It was earlier observed that the issue is likely to be at the UE end, but after all the previous attempts, it is worth to check if a different radio with other frequency band impacts to packet loss. The 5 minute good coverage uplink test was ran again on a different radio, with n78 frequency band in a completely different site. This time, there was about 30 lost packets for every run which means that the original radio with the n77 frequency band did not cause the packet loss.

6.5 Packet loss conclusion

Because increased traffic was not the root cause for packet loss, and the packet loss was identified to happen between UE and the BTS, it was expected that the reason for packet loss is in the UE.

The good coverage uplink test was done using three different modems. Huawei P40 5G SA phone, Askey 5G SA modem, and Nokia FastMile 5G Gateway. The Huawei phone did not drop a single packet on multiple runs, which lead to the expectation that the issue is just some hardware problem with the Askey modem or with the Linux PC. Because the Huawei phone is not a suitable modem for URLLC type of testing, the Nokia FastMile 5G Gateway was selected as the next modem candidate. The Nokia device however, lost just as many packets as the Askey modem. This verified the expectation that the problem was in the Linux PC hardware, most likely in the ethernet port. After changing the PC to another Linux computer and running the test with Askey modem, there was still same amount of lost packets. The next action was to update the Askey modem's firmware, which also did not resolve the packet loss issue. The final effort was to change radio and frequency band, to verify there are no issues with the radio conditions. Again, the amount of lost packets stayed the same.

The Askey 5G SA modem and Nokia FastMile 5G gateway use the same Qualcomm SDX55 (Snapdragon X55 5G Modem-RF System) chipset[1]. Because none of the efforts to fix the packet loss resolved the issue and the Huawei P40 phone with different chipset did not lose any packets at all, the results indicate towards the chipset. The SDX55 is the first 5G SA supporting chipset released by Qualcomm. The next generation chipset SDX65[2] has been introduced already, and it is expected to be available in 2022.

Figure 6.2 is a flowchart visualizing everything that was done to find reasons for packet loss.

[1] https://www.qualcomm.com/products/technology/modems/snapdragon-x55-5g-modem

[2] https://www.qualcomm.com/products/technology/modems/snapdragon-x65-5g-modem-rf-system

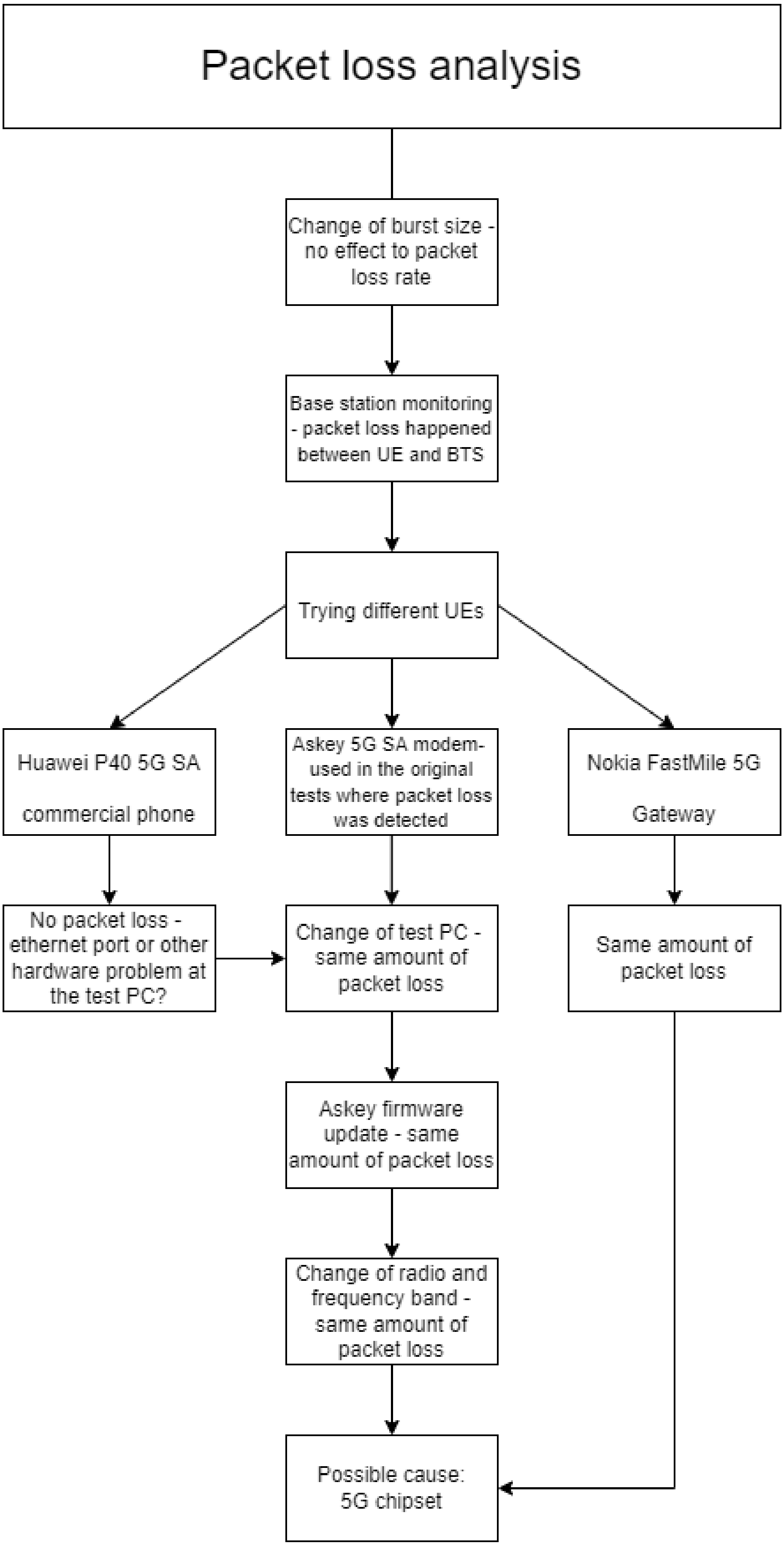

Figure 6.2: Packet loss analysis for the chapter 5 uplink transmission lost packets.

7 Conclusion

The motivation for this thesis was to find ways to make measurements on how a 5G SA network performs on key URLLC performance indicators, analyse and visualize these measurements, find reasons for certain network behavior and estimate what kind of impact different URLLC features will have when implemented. Another motivation was to find a way to detect packet loss and find what causes it, because packet loss impairs reliability significantly and should be eliminated before URLLC deployment. The results of this work can be used to see what kind of impact different URLLC features have had.

The test cases used in this thesis were chosen to mitigate V2X, which is one of the most relevant and highly specified URLLC use case. The reliability and one-way latency for each test was measured and analysed.

For downlink, the results looked encouraging and it is expected that with some of the key URLLC features, which are shortened TTI duration and non-slot based scheduling, reliability and latency can be improved significantly from current level, towards URLLC standards. Coverage or mobility did not have a meaningful impact to downlink one-way latency. This was because in this thesis' test setup there was no need for packet retransmissions, which would have caused more latency. This is is promising especially for mobility use cases such as V2X. Undoubtedly, the tight URLLC requirements will cause problems which are not even realized yet nor discussed in this thesis. Also, the final results always depend on various static and dynamic network conditions. Nevertheless, achieving URLLC standards is realism for downlink transmission, at least in the test conditions conducted for this thesis. Uplink performed much weaker than downlink, because of uplink resource scheduling and packet retransmissions. Because URLLC has same requirements for both uplink and downlink, grant-free transmission is very important feature for uplink transmission to reach URLLC standards. Handover was problematic for both downlink and uplink, because of the brief but massive increase in latency. To mitigate this problem, the proposed solution is DAPS handover, which establishes UE to not detach from source 5G cell before attaching to the new target cell.

Packet loss has a huge impact to reliability, because certain reliability levels become unreachable under any latency if a packet was lost. This is especially important for the URLLC applications' reliability requirements, because only a few lost packets can make the difference. It was noticed that all packet loss happened in uplink transmission and after consistently eliminating different factors potentially causing packet loss, the results indicate towards the 5G SA chipset, which was used in all of the reliability tests.

There will be some essential future work to continue from this thesis. Firstly, gNB could be configured to give proactive grants for UE even in shorter time intervals than 4 ms which was used for this thesis. This could help to see how much uplink latency can be reduced and balance uplink and downlink before actually implementing grant-free transmission. Secondly, more URLLC use cases could be mitigated. Especially with bigger packet size than 255B, it could be tested how well the base station can use all available resource blocks. For example, discrete automation would be a good test scenario which uses 1354B packets. Also lower channel bandwidth could be used to see how the base station would behave with fewer available resource blocks. In future, after the basic URLLC features will be implemented, these same tests should be ran again to see each feature's impact to reliability and latency. Also, in future the URLLC tests should be done using a 5G SA modem with next generation chipset and it should be verified that there are no lost packets.

References

[1] H. Holma, A. Toskala, and T. Nakamura, *5G technology: 3GPP new radio*. John Wiley & Sons, 2020.

[2] X. Lin, J. Li, R. Baldemair, *et al.*, "5g new radio: Unveiling the essentials of the next generation wireless access technology", *IEEE Communications Standards Magazine*, vol. 3, no. 3, pp. 30–37, 2019.

[3] N. Hassan, K.-L. A. Yau, and C. Wu, "Edge computing in 5g: A review", *IEEE Access*, vol. 7, pp. 127 276–127 289, 2019.

[4] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "Mimo for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?", *IEEE Communications Magazine*, vol. 52, no. 12, pp. 110–121, 2014.

[5] E. G. Larsson and L. Van der Perre, "Massive mimo for 5g", 2017.

[6] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges", *IEEE communications magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[7] S. Teral, "5g best choice architecture", *IHS Markit*, 2019.

[8] 3GPP TS 38.300, "5G; NR; Overall description; Stage-2", version 15.8.0 Release 15.

[9] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view", *Ieee Access*, vol. 6, pp. 55 765–55 779, 2018.

[10] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5g usage scenarios and traffic models", *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 905–929, 2020.

[11] 3GPP TR 38.913 "5G; Study on Scenarios and Requirements for Next Generation Access Technologies" version 14.2.0 Release 14.

[12] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, "5g urllc: Design challenges and system concepts", in *2018 15th international symposium on wireless communication systems (ISWCS)*, IEEE, 2018, pp. 1–6.

[13] 3GPP TS 22.261 "Service requirements for the 5G system; Stage 1" V18.6.0 Release 18.

[14] R. S. o. I. International Telecommunication Union ITU-R, "Minimum requirements related to technical performance for imt-2020 radio interface(s)", International Telecommunication Union, 2017, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2410-2017-PDF-E.pdf. (Accessed: 23.12.2021).

[15] M. A. Uusitalo, Z. Li, S. R. Khosravirad, *et al.*, "Reliability and low latency for cellular wireless",

[16] A. Heddaya, A. Helal, *et al.*, "Reliability, availability, dependability and performability: A user-centered view", *Boston, MA, USA, Tech. Rep*, 1997.

[17] R. Ratasuk, "Ultra reliable low latency communication for 5g new radio", Nokia Bell Labs, 2018. [Online]. Available: `https://futurenetworks.ieee.org/images/files/pdf/FirstResponder/Rapeepat-Ratasuk-Nokia.pdf`.

[18] R. Prasad, *OFDM for wireless communications systems*. Artech House, 2004.

[19] Microsoft, *Packet scheduler.* [Online]. Available: `https://docs.microsoft.com/en-us/previous-versions/windows/desktop/qos/packet-scheduler`, (Accessed: 5.1.2022).

[20] 3GPP TR 38.912 "Study on New Radio (NR) access technology" Version 17.0.0 Release 17.

[21] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Grant-free transmissions for ultra-reliable and low latency uplink communications", in *Proc. Eur. Conf. Netw. Commun.(EUCNC)*, 2018, pp. 1–6.

[22] N. Katyal, *Hybrid automatic repeat request (harq) in lte fdd*, 2018.

[23] G. Americas, "New services applications with 5g ultra-reliable low latency communications", Technical Report, 5G Americas, Tech. Rep., 2018.

[24] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, "High-reliability and low-latency wireless communication for internet of things: Challenges, fundamentals, and enabling technologies", *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 7946–7970, 2019.

[25] 3GPP TS 22.261, "5G; Service requirements for next generation new services and markets", Version 15.7.0 Release 15.

[26] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Use cases, requirements, and design considerations for 5g v2x", *arXiv preprint arXiv:1712.01754*, 2017.

[27] 3GPP TR 22.186 "Enhancement of 3GPP support for V2X scenarios; Stage 1", Version 15.4.0, Release 15.

[28] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—the new and improved power grid: A survey", *IEEE communications surveys & tutorials*, vol. 14, no. 4, pp. 944–980, 2011.

[29] ETSI TR 103 401 V1.1.1 (2016-11), Smart Grid Systems and Other Radio Systems suitable for Utility Operations, and their long-term spectrum requirements.

[30] K. Zeng, W. Deng, R. Wang, *et al.*, "5g network performance evaluation and deployment recommendation under factory environment", in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, IEEE, 2021, pp. 1370–1375.

[31] 3GPP TS 23.501, "System architecture for the 5G System(5GS); Stage 2", Version 16.13.0, Release 16, Table 5.7.4-1: Standardized 5QI to QoS characteristics mapping.

[32] *Iperf - the ultimate speed test tool for tcp, udp and sctp*. [Online]. Available: `https://iperf.fr/`, (Accessed: 9.2.2022).

[33] *About wireshark*. [Online]. Available: `https://www.wireshark.org/`, (Accessed: 8.2.2022).

[34] J. Postel, "User datagram protocol", Tech. Rep.

[35] *Tcp and udp*, Cisco Networking Academy. [Online]. Available: `http://cisco.num.edu.mn/CCNA_R&S1/course/module7/7.2.4.2/7.2.4.2.html`, (Accessed: 22.2.2022).

[36] X. Xiao, "Chapter 2 - what is qos?", in *Technical, Commercial and Regulatory Challenges of QoS*, ser. The Morgan Kaufmann Series in Networking, X. Xiao, Ed., Boston: Morgan Kaufmann, 2008, pp. 11–35. DOI: `https://doi.org/10.1016/B978-0-12-373693-2.00002-1`.

[37] D. Clark, *Computer networks: A systems approach*. [Online]. Available: `https://book.systemsapproach.org/congestion/issues.html`, (Accessed: 2.2.2022).

[38] L. Sequeira, J. Fernández-Navajas, J. Saldana, J. R. Gállego, and M. Canales, "Describing the access network by means of router buffer modelling: A new methodology", *The Scientific World Journal*, vol. 2014, 2014.

[39] C. Johnson, *5G New Radio in Bullets*. Independently Published, 2019, ISBN: 9781077484351.

[40] 3GPP TS 38.300, "NR and NG-RAN Overall description; Stage-2", Version 16.2.0 Release 16.

[41] *5g nr modulation and coding scheme – modulation and code rate*. [Online]. Available: `https://www.techplayon.com/5g-nr-modulation-and-coding-scheme-modulation-and-code-rate/`, (Accessed: 1.4.2022).

[42] 3GPP TS 38.214, "5G NR Physical Layer procedures for data", Version 16.9.0, Release 16.