Engineering Applications
https://publish.mersin.edu.tr/index.php/enap
e-ISSN 2979-9201

Practical QoS measurement and analyzes on a 5G non-standalone architecture

Olimpjon Shurdi *[*1], **Alban Rakipi** [1], **Arjola Biti** [2]

[1]Polytechnic University of Tirana, Department of Electronics and Telecommunications, Albania, oshurdi@fti.edu.al; arakipi@fti.edu.al
[2]Vodafone Albania, Albania, arjola.biti@vodafone.com

Cite this study: Shurdi, O., Rakipi, A., & Biti, A. (2023). Practical QoS measurement and analyzes on a 5G non-standalone architecture. Engineering Applications, 2 (2), 115-125

Keywords
5G
Quality of Service (QoS)
Throughput
Jitter
Latency

Research Article
Received:08.03.2023
Revised: 01.05.2023
Accepted: 14.05.2023
Published:26.05.2023

Abstract

Fifth Generation (5G) networks are becoming the standard in the global telecommunications industry and are becoming a permanent part of everyday life. As mobile network operators have commenced to publicize the implementations of their 5G networks, measurements are not frequently used to demonstrate the actual aspects of these networks' capabilities.This article presents an actual 5G architecture based on the 5G Option 3x reference model, together with findings from Quality of Service (QoS) testing. The outcomes are compared to our actual hands-on measurements of the 4G network in addition to being compared to the expectations for the 5G network. Based on examination of the 5G testbed results, 5G performed much better than 4G in all fundamental QoS, including up- and downlink throughput, latency, packet error rate and jitter. Additionally, practical measurement results on this non-standalone 5G architecture demonstrate that latency and jitter are not greatly impacted by the load on the cell or the core network provided traffic preferences are set up correctly. The study of QoS and a first performance assessment of 5G, along with the identification of application-level performance concerns, are the paper's main conclusions. These results underline the necessity of validating and testing generic 5G services and applications through fair benchmarking approaches.

1. Introduction

5G is the most recent mobile network being deployed globally. As was announced, a 5G network can provide high throughput, high reliability, low latency, increased capacity, availability and connectivity, and dynamic bandwidth allocation and in general better performance. In fact, it is expected that the 5G performance will span over the three extremes of bandwidth, latency, and capacity requirements, which enable enhanced Mobile Broadband (eMBB), Ultra Reliable and Low Latency Communication (URLLC), and massive Machine Type Communication (mMTC), respectively [1].

The 3rd Generation Partnership Project (3GPP) has regulated two primary 5G deployment modes in Release 15 (Rel-15), termed Non-Standalone (NSA) and Standalone (SA), to meet these requirements [2]. The main difference is that whereas 5G NSA uses the current 4G core network to handle the control plane, 5G SA uses its own 5G core and so operates independently of the 4G network (Figure 1). Both of these modes need a 5G New Radio (NR) Radio Access Network (RAN) made up of Next Generation Node Bs (gNBs), which are the 5G equivalent of 4G E-UTRAN Node Bs (eNBs).

The majority of Mobile Network Operators (MNOs) have chosen to use the NSA mode throughout the current stages of 5G implementation because it is an easier and less expensive option. However, 4G/5G co-dependence in the NSA presents a variety of setup, operational, and performance challenges that need more research and analysis. Actual 5G networks claims include 20Gbps data throughput, 1ms latency, 1 million devices=km^2, and 10 times less energy usage (than 4G UE), although their combined effects on the network are not yet completely understood.

Analysis of the performance of a 5G network and service can also be done using a variety of techniques and technologies. Although 5G networks can benefit from the same network and service validation systems now in use, considerable tool upgrades will be required for thorough validation and verification of 5G goals.

Recently, there has been significant research related to the 5G performance evaluation, however, these are not tested on live network, so it cannot be known which will fulfill the expectations. For examples, the authors in [3] had evaluated the performance of an NSA 5G architecture. They observed that if the traffic preferences are properly set, the load on the cell or the core network (CN) has no effect on latency and jitter. Also, the authors examined the performance of SA and NSA 5G new radio (NR) installations in terms of coverage, network capabilities, and deployment cost based on simulations in [4]. In more recent works, the effectiveness of NSA 5G networks has also been assessed; for examples, see [5] and [6].

In this paper, we present a testbed measurement study on 5G experimental mid-band NSA networks of Vodafone Albania. To the best of our knowledge, this work constitutes the first effort toward empirically studying deployment, coverage, and performance aspects of 5G NSA deployments in QoS terms, as well as providing insights regarding future 5G version updates.

The structure of the paper is the following: In section 2 are presented 5G non-standalone technology main elements functionality and the QoS main parameters. In section 3 studies the performance of 5G testbed network where our numerical results are presented. Results and discussions are presented in section 4 and conclusions in Section 5.

2. 5G Non-Standalone networks

This section describes the 5G NSA architecture, including its components, network functions and interfaces, as well as its mode of operation. Furthermore, we describe our NSA 5G testbed in relation to the reference architecture.

2.1. Background

The 3rd generation partnership project (3GPP) proposes 5G NSA architecture as part of Release 15 to enable mobile network vendors and operators in their transition from 4G to 5G [2]. This release defined a number of optional 5G network architectures for the NSA. Among the many candidates, however, one architecture stood out as the typical approach supported by mobile networks, in which the 4G core, i.e., the Evolved Packet Core (EPC), is interfaced with the next-generation node B (gNB) to enable 5G NR functionality. This architecture is referred to by 3GPP as Option 3 [2]. This configuration employs two cells: an evolved Node B (eNB) and a general node B (gNB). Dual cell configuration enables dual connectivity (DC), a feature carried over from 4G to 5G. This option allows user equipment (UE) to simultaneously connect to two cells, thereby enhancing throughput and mobility support. DC identifies the eNB as the master cell group (MCG) and the gNB as the secondary cell group (SCG). The connection between the two cell groups is then ensured via the X2 reference interface between eNB and gNB.

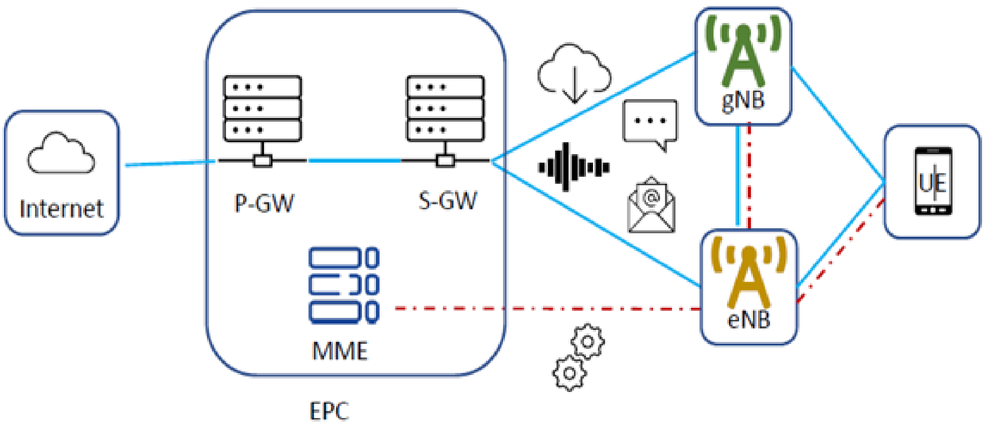

Figure 1. 5G NSA Option 3 architecture (source 3GPP TR 38.801)

Option 3 is broken down into three variants or subtle varieties. The primary distinction lies in the selection of the bearer for the user plane (UP), which consists of the tunnels connecting the user equipment (UE) to a packet data network (PDN) such as the Internet. Figure 1 depicts these configurations, which can be interpreted as follows: Option 3 uses the eNB as a split (Master Cell Group) MCG bearer.

This means that UP data is routed using the S1-U interface through the eNB to both the gNB (using the X2-U interface) and directly to the UE; (ii) Option 3a uses a SCG bearer, which means that UP data for 5G NR UEs is only routed through the gNB (using the S1-U interface); (iii) Option 3x uses a SCG split bearer, which means that UP data As previously stated, the S1 reference interface facilitates the connection between the NSA 5G core and the RAN. Figure 1 defines this interface as S1-C for the control plane (CP) and S1-U for the user plane (UP). 3GPP Release 14 introduces this separation and defines an updated EPC architecture with the control/user plane separation (CUPS) [2].

From the standpoint of the defined implementation alternatives for the 5G system, the non-Standalone option (addition of gNBs under the management of EPC – 4G core) is one of the most viable solutions for the rapid market adoption of 5G. In addition, in the European Union, frequencies in the Frequency Range 1 (FR1: frequencies up to 6 GHz) have already been assigned to 5G systems and are being used in 5G trials and deployments. In this paper, the possible KPIs for different bandwidths inside the FR1 of a 5G NSA system are analyzed to quantify the data rates that end users can expect from early 5G installations.

2.2. 5G testbed proposed and design

The testbed we employed is a defined architecture called "Option 3X," which stands for "migration path to Non-Standalone Next Generation Radio with LTE aided mode coupled to Evolved Packet Core" (Figure 1). The EPC is responsible for AAA functions, service mobility management, and it ensures connectivity with the IP network. Another key part of the 5G architecture is gNB, which is responsible for the radio attachment between UEs and the Core network's interfaces in the case of 5G media. For the basic understanding, Figure 4 on [2] depicts the generic 5G architecture, where the main elements are also defined.

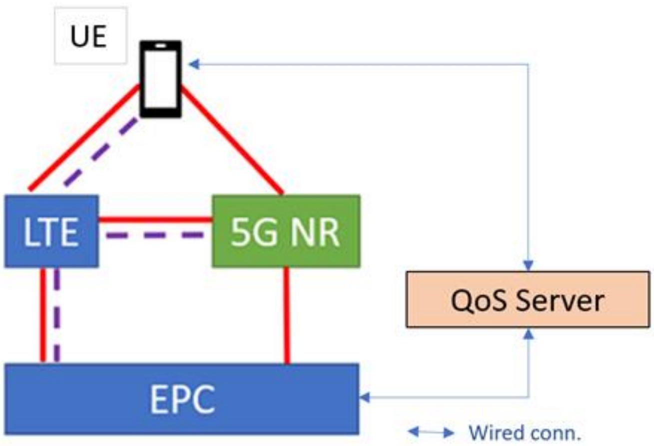

Figure 2. The architecture of 5G Option 3X testbed

In our testbed, we use a dedicated server, where we have installed different tools like iPerf3, Speed test by Ookla, and SolarWinds. The 5G NR is simulated using Amarisoft Callbox Advanced, which is a self-contained 5G base station deployed in an indoor environment. Allowing us to provide an overview of the most critical network experience KPIs that MNOs will consistently measure in the assessment of their performances.

The 5G NR is simulated using Amarisoft Callbox Advanced, which is a self-contained 5G base station deployed in an indoor environment. Next, we will discuss one of the most important components, the eNB/gNB or base station in general. When choosing a base station, it is important to focus on both the software and the hardware, as together they describe the overall capabilities of a base station. The hardware used for base stations in testbeds consists of computers that directly contribute to the overall processing and storage capabilities, and SDR cards. Features that need to be considered for SDR cards include the number of downlink and uplink antenna ports supported, the frequency bands supported, maximum bandwidth allowed and maximum radiated power. Among other solutions we opt for the Amarisoft Callbox Advanced. The main specifications of Callbox are summarized in Table 1. The Callbox supports up to 2 SDR cards. Each SDR card has a maximum power of up to 5 dBm, provides 4x4 MIMO and can operate in the sub-6 GHz band with a maximum bandwidth of 100 MHz. The AMARISOFT Callbox comes with antennas designed for indoor use and has a limited range, which is acceptable for the power level emitted from the Callbox SDRs of 5dBm. A key consideration when selecting software for our testbed is support for 5G protocol stacks. In addition, support for 3GPP standard features such as MIMO, carrier aggregation, dual connectivity should also be considered.

Table 1. Main characteristics of the AMARISOFT Callbox

Features	Amarisoft Callbox
3GPP release	16
Frequency bands	All FDD and TDD bands in sub-6GHz
Bandwidth	up to 100MHz
MIMO	up to 4x4 MIMO
Modulation schemes	up to 256 QAM
Configurations	NSA: 1 5G NR 100MHz 4x4 + 1 LTE 20MHz 4x4

The main configuration parameters adopted in our simulation tests are presented in Table 2. We use Iperf software to measure the KPIs by transmitting Transmission Control Protocol (TCP) packets between the base station and the UE.

Table 2. Environmental parameter setup

Variable Value	Variable Value
NR bandwidth, band 50MHz, n78	NR bandwidth, band 50MHz, n78
4G bandwidth, band 20MHz, B3	4G bandwidth, band 20MHz, B3
DL and UL modulation QAM 256 and QAM 64	DL and UL modulation QAM 256 and QAM 64
4G and 5G frequency mode FDD	4G and 5G frequency mode FDD
Distance to UE 4 m	Distance to UE 4 m
Number of slots (DL test) DL slots=7 and UL slots=2	Number of slots (DL test) DL slots=7 and UL slots=2
Number of slots (UL test) DL slots=6 and UL slots=3	Number of slots (UL test) DL slots=6 and UL slots=3
Number of antennas 2 _ 2	Number of antennas 2 _ 2
Base station height 2m	Base station height 2 m
Samsung Galaxy S23	Samsung Galaxy S23

2.3. QoS parameters and expectations

In the 5G networks, Quality of Service (QoS) model is based on the QoS Flows. Each QoS flow has a unique identifier called QoS Flow Identifier (QoI). There are always two types of bearer flows GBR (Guaranteed bit rate) QoS flows and NGBR (Non-Guaranteed bit rate) QoS flows. GBR bearers are used for real-time services such as rich voice and video services which occur in real-time. A GBR bearer has a minimum amount of bandwidth that is always reserved by the network. Non-GBR bearers do not have specific bandwidth allocation e.g., file downloads, email, internet access, etc. There are different parameters to evaluate QoS in 5G networks, in this paper we will present some of the most important that end users and applications are affected.

- ***Throughput:*** 5G throughput performance is one of the most important indicators of user experience. Increasing bandwidth, utilizing various coding techniques, and improving modulation all result in higher data transmission rates. Even 100 MHz of bandwidth can be employed in the downlink direction, and 1.5 Gbps is the maximum theoretical capacity. Naturally, the quantity of receiver and transceiver antennas can also be used to tune the transmission rate.
- ***Latency:*** One of the main promises of 5G is the dramatic reduction of latency from the levels experienced with 4G to approximately 1 ms. In particular, latency can be maintained at a low level even when the network is experiencing significant traffic loads.
- ***Packet Error Rate (PER):*** It defines an upper bound for the rate of PDUs (e.g., IP packets) that have been processed by the sender of a link layer protocol (e.g., RLC in RAN of a 3GPP access) but that are not successfully delivered by the corresponding receiver to the upper layer (e.g., PDCP in RAN of a 3GPP access).
- ***Maximum Packet Loss Rate:*** The Maximum Packet Loss Rate (UL, DL) indicates the maximum rate for lost packets of the QoS Flow that can be tolerated in the uplink and downlink direction.
- ***Maximum Packet Loss Rate:*** The Maximum Packet Loss Rate (UL, DL) indicates the maximum rate for lost packets of the QoS Flow that can be tolerated in the uplink and downlink direction.
- ***Jitter:*** the variation in time delay between when a signal is transmitted and when it's received over the network connection, measuring the variability in ping. This is often caused by network congestion, poor hardware performance and not implementing packet prioritization.

There are further, more subtle measurements that can be conducted which can reflect further aspects of user experience. Some examples are the following:

- ***WEB Browsing tests***: use case overall session time (DNS time + IP service access time + Transfer time)

- ***File Download/Upload tests***: HTTP download/upload reflecting Customer Experience with file 10 MB for DL and 5 MB for UL in connection with the evolution of networks and the spread of 5G service.
- ***Network Capability Download/Upload tests:*** HTTP FDTT multisocket network capability download/upload with files of 500MB files.
- ***ICMP PING tests:*** The ICMP PING test case consists of 6 packets of 40Byte-sized. PING messages sent subsequently without any intermediate pause. Ping evaluation will be performed only for the last 5 pings in the test sequence.
- ***Payload PING tests:*** The Payload PING follows the same methodology as the regular ICMP PING use case (described in the section above) but using a size of 800 Byte.
- Sending ICMP packets is a common method to examine network latency, but the output and features are very poor for detailed analysis. Furthermore, latency can increase for many network applications in case of the high traffic load; for such cases, ICMP is not able to provide proper insights, we will investigate into interactivity measurements, traffic patterns, that reflect real-life user behavior.
- ***New TWAMP based Data tests:*** This new testcase will perform an emulation of eGaming real time traffic pattern compressed in 10s, using TWAMP protocol approach: "eGaming real time" 10s traffic pattern against Interactivity Server.

The UE executes the identical test cycle in the following sequence: non-Secure Socket Layer (non-SSL) focus for fixed filed DL live webpages + 10MB DL/UL + 7s DL/UL + Interactivity eGaming +Speed Test. At the beginning of each cycle the ICMP test pings are executed. For each of the tests we calculated the statistical KPI, P10, median and P90 reflecting different user experience for each of the categories.

3. Measurement and results

Numerous standardization efforts are concentrated on defining criteria for evaluating the performance and quality of 5G networks. Existing approaches are typically tied to a limited perspective of the entire system. This is warranted by the overall complexity of the 5G network and the decomposition of the 5G architecture. The standard approach to quality assessment focuses on the requirements and characteristics of 5G services as defined by ITU-R [7-8]), and 3GPP [9-10]. Next Generation Mobile Networks Alliance (NGMN) has published its KPI and 5G requirements recommendations [11]. The 3GPP specification [12] has thus far defined several general 5G KPIs for network slicing parameters. One of the most focused areas is the Integrity KPIs: end-to-end latency of the 5G network, upstream/downstream throughput for Network Slice Instance (NSI) and at N3 interface, RAN-UE throughput.

In our setup, the 5G network is operating with real live 4G network creating the perspective NSA network. Server is connected directly to EU and EPC; we are using several tools for timing synchronization and valuating the QoS mentions parameters. Here are some of the tools:

- The iPerf tool: A popular application for data transmission speed measurements is iPerf – and its updated version, iPerf3 [13]. It can be used for both TCP and UDP transmission, even to create dynamic traffic load for a system. Furthermore, iPerf output can be compared easily even for significantly diversified systems.
- ICMP-related measurements: Sending ICMP packets is a common method to examine network latency. Such applications can be found on almost any x86-based system, but the output and features are extremely poor for detailed analysis. Furthermore, latency can increase for many network applications in case the traffic load increases; for such cases ICMP is not able to provide proper insights. A small improvement can be achieved with special settings, but it is still not a fully adequate solution.
- Application-specific measurements: During the transmission of exceptionally large files, latency and transmission speed can be analyzed at the same time. Although this method might also reveal information on retransmitted packets, their application-specific nature is a significant disadvantage.
- Tools for user measurements: Before a service is made available, it is important to assess how the network operates under challenging traffic circumstances. For this purpose, targeted measurement tools can be used such as IXIA solutions [14] or Cisco TRex [15]. As TRex provides free access to most of the features we require, we employed it for our analysis; however, any other system can be used to replicate the measurements with the correct settings.

Four different scenarios have been defined, simulating the four main parameters explained above, how are related to the bandwidth part, and other network values affected.

3.1. Throughput versus BW

A 5G NR channel of 50 MHz is nominally adopted, with 5G NR numerology 1 and SubCarrier Spacing (SCS) of 30 kHz. In this case, the maximum number of Physical Resource Blocks (PRB) that can be used is 133, as per 3GPP

TS 38.101-3 [2]. Moreover, the 5G carrier is currently configured to use two antennas. Hence, we can enable a maximum of two layers to increase the throughput, taking advantage of the Multiple Input Multiple Output (MIMO) diversity using close loop spatial multiplexing. We also use a single beam and a maximum modulation of 256-Quadrature Amplitude Modulation (256-QAM). As per 5G NR Modulation and Coding Scheme (MCS) mapping in TS 38.214, 256-QAM modulation is enabled with the maximum MCS of 27, 8 bits encoded per OFDM symbol, and a coding rate of 948/1024. For the experimentation process presented here, bandwidths of 50, 80, 90 and 100 MHz were considered in band N78 (i.e., 3.5 GHz). DL and UL throughput measurements were conducted in 5G experimentation platforms described, using UDP and TCP traffic through the Iperf3 tool. The results for e single UE are depicted in Figure 3. In this setup, the L1 payload spectral efficiency results in 7.4063 bits per symbol. It is noted that the channel conditions were almost ideal, with UEs and gNB in fairly close distance.

As can be noted in Figure 3, the outcomes of the experimentation measurements are below the theoretical values, but extremely near to them. Considering that the measurements belong to transport layer (TCP and UDP have been employed) it is acceptable to claim that the theoretical values are validated.

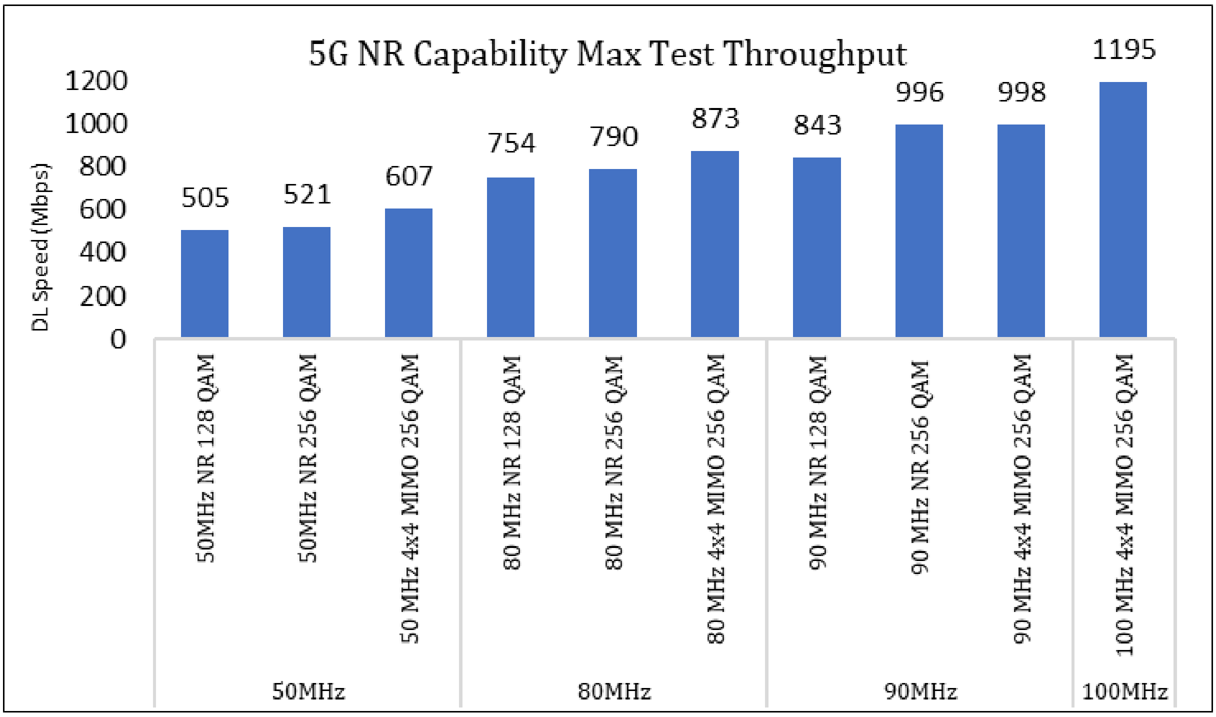

Figure 3. Throughput versus bandwidth for single UE

With the above setup, the maximum theoretical DL throughput is of approximately 505 Mbps, derived from multiplying the number of layers (2), PRBs (133), OFDM data symbols (11), and bits per symbol (8 bits), with the coding rate (0.926), the number of slots in DL used for data transmission (7), the patterns per frame (2), and the number of frames in one second (100).

Figure 6 shows the case when in the simulated system we have 5 users carrying out a light traffic, as can be seen compared to Figure 4 we have a slight decrease in this due to the occupation of the band by other users.

3.2. Latency of single UE in case of light traffic

We analyzed the effect of Device or Network Under Test (DoNUT) on 5G latency deviation using 100 Byte and 1000 Byte packets. In four different test scenarios, the time between sending packets ranged from 10 to 1000 milliseconds. After completing these tests, we measured latency with continuous background traffic whose volume was 50 percent of the network's capacity: 800 Mbps. Figure 5 depicts the one-way latency results when background traffic consumed 70% of the link's capacity and when there was no background traffic.

In the observed simulation, in the case of a single user connected to the gNB base station, the delays are negligible even when the network load is significant. this can be explained for several reasons: firstly, in the simulated experiment, the transmission distance is quite small, so the signal-to-noise ratio is very satisfactory, secondly, we have a high bandwidth available, and finally, the impact of load delays that the network is negligible.

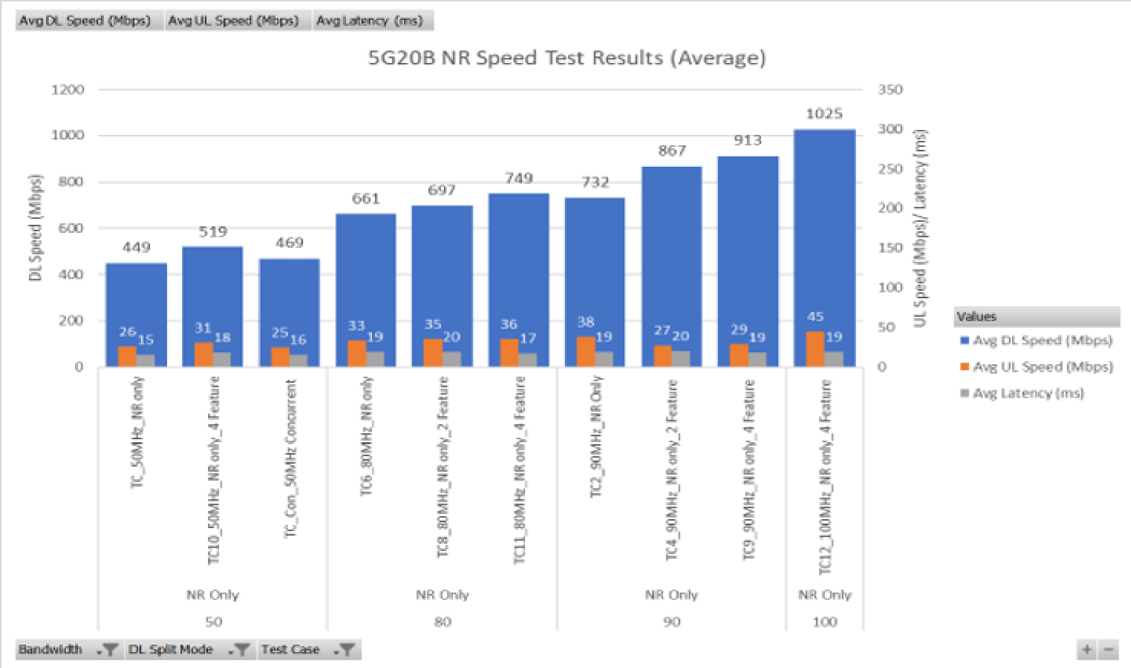

Figure 4. Throughput versus bandwidth for single 5UEs

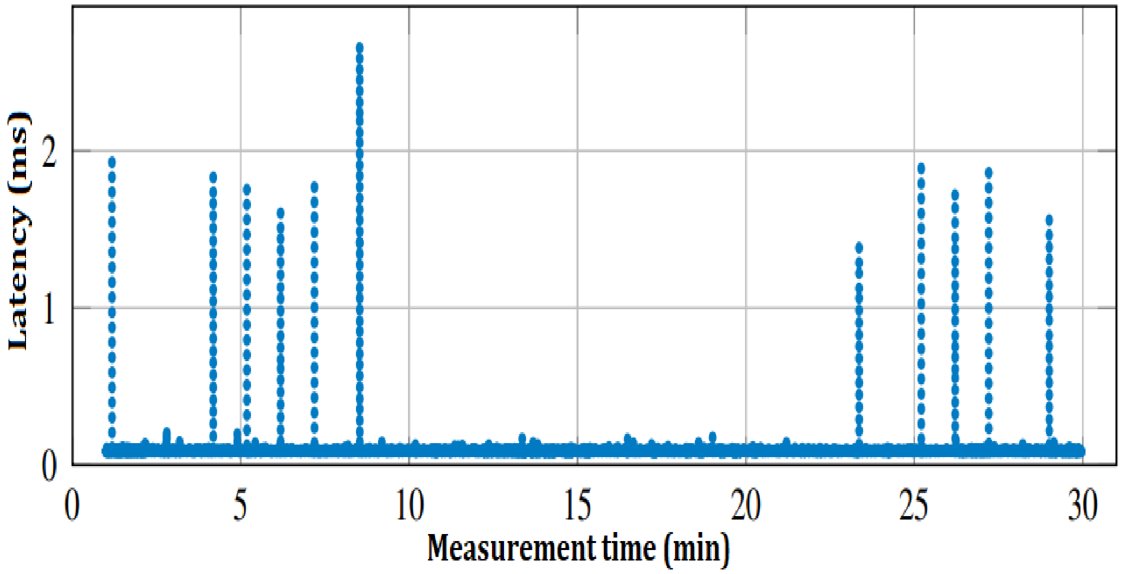

Figure 5. Latency in 30 min with light traffic

3.3. Packet error rate (PER) versus signal to noice rate (SNR)

The Packet Error Rate (PER) defines an upper bound for the rate of PDUs (e.g., IP packets) that have been processed by the sender of a link layer protocol (e.g., RLC in RAN of a 3GPP access) but that are not successfully delivered by the corresponding receiver to the upper layer (e.g., PDCP in RAN of a 3GPP access). Figure 6 depicts the average PER versus the signal to noise rate in simulated channel of light of sight, which helps to understand the context in which the measurements were taken. As the SNR decreases (the noise intensity becomes larger), the PER value decreases exponentially, the receiver will require for every packet error retransmission. Among others, in the non-line of sight (NLoS) scenario, low Signal to Interference & Noise Ratio (SINR) values directly map to low Modulation and Coding Scheme (MCS) index and, thus, the resulting low throughput. Given the poor signal level in NLoS, the scheduler reduces the amount of transmitted data to increase the redundancy and maintain the error rate at a reasonable level. In fact, the Bit Error Rates (BER) observed in the NLoS scenario is lower than 10%, which is a typical target value when the propagation conditions are not favorable enough, so that the Hybrid Automatic Repeat Request (HARQ) mechanism can operate optimally.

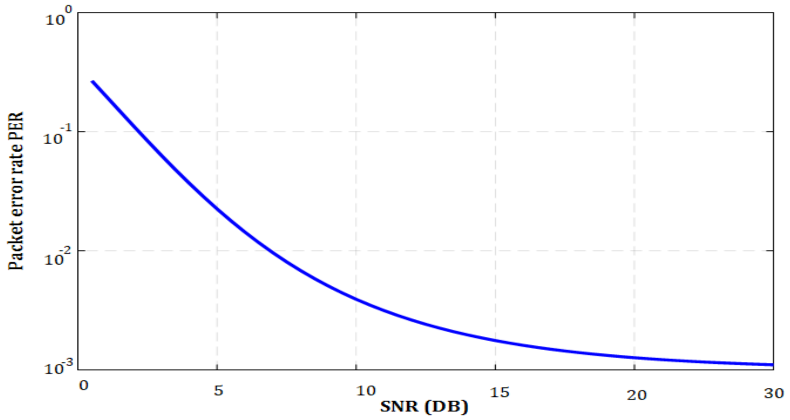

Figure 6. Packet error rate PER vs SNR

Another factor that affects the KPI parameters is the transmission channel. Due to increased path loss, the SINR decreases as the user's distance from the gNB increases, resulting in lower bit rates. In addition, during unfavorable channel conditions (multipath fading declines), the bit rates are reduced further due to a further decrease in the SINR. This will increase the transmission latency of packets and consequently the scheduling latency of consecutive packets.

3.4. Delay in case of heavy traffic

In this simulation we have examined the cases of the dependence of the delay time on the bandwidth, only the case of NR was examined to see also the dependence of the traffic load against this parameter. In the first case, when we only have the 50MHz band at our disposal, we can see an average delay of 15ms as the bandwidth increases, we expected the latency to decrease, but due to the high traffic load that we simulate using the download of some for large files, we have a small increase of 17-19ms.

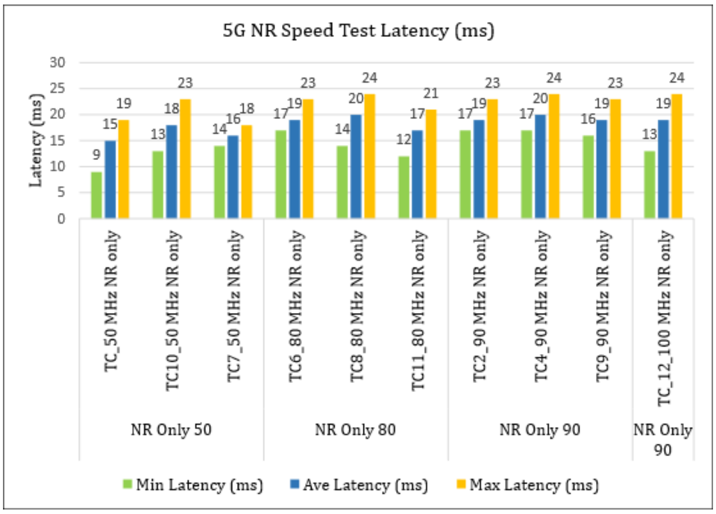

Figure 7. Latency vs bandwidth with heavy traffic

Figure 7 shows the values for NR. 50-90 bands are configured, an increase in the middle of the band is noticed.

3.5. Channel modeling and challenges

In this paper we considered only the cases of the ideal communication channel, but to generalize the developments of 5G technology towards new Tera Herz communication bands, the behavior in real environments remains to be observed and analyzed. Realistic channel modeling is an additional important aspect of 5G KPI testing. Several studies have discovered that the existing 3GPP channel models have some applicability to higher frequency bands up to 100 GHz. As expected, the measurements indicate that shorter wavelengths increase the sensitivity of the propagation models to the scale of the environment, as well as the frequency dependence of the path loss and the occurrence of blockage. In addition, the penetration loss is highly dependent on the material and increases as the frequency of operation increases. The shadow fading and angular spread parameters are greater, and the line of sight (LOS) and non-line of sight (NLOS) boundary depends not only on antenna heights but also on the local environment. This has simplified some initial proposals, but the primary drawback remains, namely, how to model a signal that is split into multiple carriers and MIMO paths, which can extend from vastly different frequency bands. It has been demonstrated that FR1 (Sub-6GHz) and FR2 (mm-wave) bands can be combined successfully, resulting in user throughputs of more than one gigabit per second and new channel modelling challenges.

3.6. Using dual connectivity to archive high throughput

In our testbed we simulate UE connectivity to gNB, but 5G NSA has the opportunity to use the solution EN-DC (E-UTRAN NR Dual Connectivity) or also called Architecture Option 3x, where the control signaling of 5G Radio is done by the LTE CORE using and 4G node. This NSA solution enables us to utilize the Dual Connectivity Uplink aggregation feature, which enables the usage of the LTE leg for user data and the aggregation of both traffic bearers to achieve a high uplink throughput. This innovation has no effect on LTE/NR nodes and improves reordering buffer processing by introducing a timer that makes the reordering process more robust and efficient.

The peak Downlink (DL) throughput was evaluated for LTE, NR and NR+LTE UL traffic aggregation, obtaining the following values as shown in Figure 8: The utilization of the NR+LTE traffic aggregation function has a significant impact on the peak throughput, as the addition of both traffic bearers is efficient.

Notice the following graph, which depicts the peak Downlink throughput for LTE, NR, and NR+LTE DL traffic aggregation scenarios. The increase from 519 Mbit/s to 780 Mbit/s equals 261 Mbit/s and is a 33% increase in peak downlink throughput.

However, if we take the data from the Uplink tests, the improvement from 33Mbps to 49Mbps = 16Mbps and that represents and improvement of 148% of uplink throughput peak.

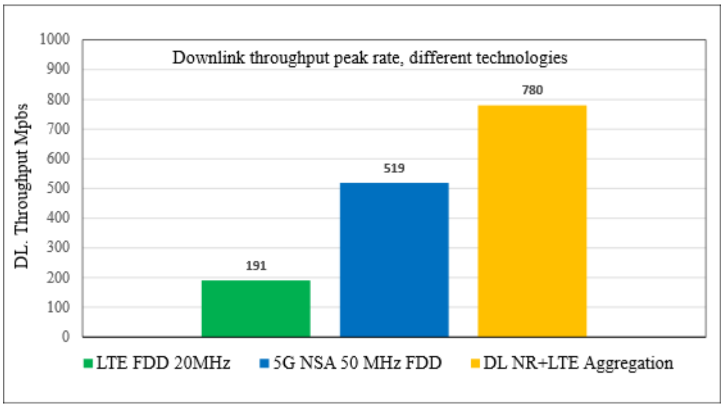

Figure 8. Downlink throughput peak rate

4. Discussion

The main configuration parameters adopted in our simulations present e supposed near future 5G network. We measure the primary KPIs by exchanging user datagram protocol (UDP) packets between the base station and the UE using the Iperf software.

Following this, we examine the performance of an NSA 5G network with the same 5G RAN configuration. In Figure 3, for example, the UL data rate is plotted against the NR cell bandwidth for both network designs. Figure 3 displays the preliminary findings that the feasible UL rate grows monotonically with the NR bandwidth. This is logical, as greater bandwidth correlates to a greater maximum data transfer rate.

In Figure 4, the DL data rate is plotted against the NR cell bandwidth for both SA and NSA 5G networks. And according to Figure 4, the feasible DL rate for both SA and NSA 5G networks increases with the NR bandwidth. In addition, the impact of utilizing a 4G core in a self-contained node is minimal, as the downlink rate of the NSA 5G network could be superior to that of the SA 5G network. It is important to remember that, for reasons of fairness, Option 3a is used in the NSA mode, in which UP data is transmitted exclusively over the gNB.

In Figure 6 we display the network delay between the UE and the 5G CN versus the NR cell bandwidth. Figure 6 demonstrates that the network latency lowers marginally when the NR bandwidth is raised, such as when the NR bandwidth is increased from 50MHz to 80MHz. Nevertheless, this reduction is less apparent for the NSA's 5G network. Moreover, it is demonstrated that the perceived network latency for NSA 5G networks is lower than the previous 4G network.

In Figure 7 and 8, we shown the effect of bandwidth in latency and in throughput in different technologies, carrier aggregation has a good impact on both parameters, and this is good opportunity for operators to reach the desirable KPI values.

5. Conclusion

This paper aims to provide some comprehensive measurements and validate real-world 5G Quality of Service capabilities. As expected, 5G data latency is significantly lower than it is for 4G -although for heavy traffic require further tune. Moreover, in the tested, real-world 5G network testbed, latency and jitter statistics do not dramatically rise with traffic load. On the throughput parameters it is observed that the theoretical limit is reached in the basic band of 100MHz. It is necessary to research more in cases of network loading with diverse types of traffic, as in these cases the tested data showed an increase of the latency.

The studies on user plane latency illustrate performance values of below 1ms target both for the 5G and 4G assuming a set of favorable configurations e.g., zero initial transmission error probability and PRACH length of 2 OFDM Symbols. We tried it with the presence values of 0.24ms and 0.69ms for 5G NSA and 4G, respectively. Similar conclusions can be drawn for control plane latency with values down to 11.6 and 16 for 5G and 4G, assuming the ideal channel conditions.

For future work, we want to investigate the impact of hosting different 5G scheduling types on the same system especially regarding potential QoS crosstalk and potential mitigation strategies.

Acknowledgement

This study was partly presented at the 6th Advanced Engineering Days [16].

Funding

This research received no external funding.

Author contributions

Olimpjon Shurdi: Conceptualization, Methodology, Investigation, Writing-Reviewing and Editing, **Alban Rakipi:** Writing-original draft preparation, operation, and testing. **Arjola Biti:** Investigation, operation, and testing.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Series, M. (2015). IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU, 2083(0).
2. 3GPP. (2018). Summary of Rel-15 Work Items, TR 21.915 version 0.5.0 Release 15, 3GPP, Version 15, 2018-12.
3. Soós, G., Ficzere, D., Varga, P., & Szalay, Z. (2020, April). Practical 5G KPI measurement results on a non-standalone architecture. In Noms 2020-2020 IEEE/IFIP network operations and management symposium (pp. 1-5). IEEE.
4. Liu, G., Huang, Y., Chen, Z., Liu, L., Wang, Q., & Li, N. (2020). 5G deployment: Standalone vs. non-standalone from the operator perspective. IEEE Communications Magazine, 58(11), 83-89.
5. Heimann, K., Gorczak, P., Bektas, C., Girke, F., & Wietfeld, C. (2019, April). Software–defined end–to–end evaluation platform for quality of service in non–standalone 5G systems. In 2019 IEEE International Systems Conference (SysCon) (pp. 1-8). IEEE.
6. Giordani, M., Polese, M., Roy, A., Castor, D., & Zorzi, M. (2019). Standalone and non-standalone beam management for 3GPP NR at mmWaves. IEEE Communications Magazine, 57(4), 123-129.
7. IMT (2015). IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond", ITU-R M.2083-0
8. Series, M. (2017). Minimum requirements related to technical performance for IMT-2020 radio interface (s). Report, 2410-0.
9. 3GPP. (2018). Service requirements for next generation new services and markets. 3GPP, TS 22.261, ver. 16.6.0, Dec. 2018.
10. 3GPP. (2018). Study on scenarios and requirements for next generation access technologies. 3GPP, TR 38.913, ver. 15.0.0, Jul. 2018.
11. Alliance, N. G. M. N. (2016). Recommendations for NGMN KPIs and Requirements for 5G. Techreport: Berkshire, UK.

12. 3GPP. (2018). Management and orchestration; 5G end to end Key Performance Indicators (KPI). 3GPP, TS 28.554, ver. 15.1.0, Dec. 2018.
13. iPerf, (2022). iPerf - The ultimate speed test tool for TCP, UDP and SCTP.
14. Ixia, (2022). Ixia validate the performance of your networking solutions. https://www.ixiacom.com/solutions/network-test-solutions
15. Cisco (2022). TRex - Realistic traffic generator. https://trextgn. cisco.com
16. Shurdi, O., Rakipi, A., & Biti, A. (2023). Practical QoS measurement and analyzes on a 5G non-standalone architecture. Advanced Engineering Days (AED), 6, 148-151.