Test, Measurement, and KPIs Validation Working Group

Whitepaper

Understanding the Numbers

Contextualization and Impact Factors of 5G Performance Results

Version 1.0

Date: 15-07-2021 Version: 1.0

DOI: 10.5281/zenodo.5094973

URL: https://doi.org/10.5281/zenodo.5094973

Table of Contents

Executive Summary

In the framework of the European 5G Public Private partnership, 5G end-to-end facilities have been developed and deployed across Europe by three infrastructure projects. These facilities have been further enhanced with on-site deployments and additional equipment in the course of seven advanced 5G validation projects that performed trials across multiple vertical industries. The full extent of the resulting experimentation environment provides all the required means to experimenters for validating applications from several vertical sectors in a fully operational 5G environment.

Already, in the aforementioned facilities, 5G performance is being assessed and vertical-specific trials are being performed. While the detailed list of the results from those activities is expected soon, fundamental questions have emerged from 5G stakeholders, namely:

a) are the results aligned with the 5G theoretical values (defined by ITU IMT-2020) and the target Key Performance Indicators agreed in the 5G PPP contractual agreement?

b) can we identify the factors that practically affect most the performance results?

c) do the results fulfil the application requirements or, from a broader perspective, do the results satisfy the expectations of the customers in the verticals?

In this whitepaper, we answer the first and second question, while the third question is to be answered when more results from the 5G end-to-end facilities projects become available.

Therefore, in the current white paper the effort is focused on clarifying the details behind the performance numbers and provide a series of interpretation guidelines that help the reader better understand the 5G domain. In addition, based on the analysis of performance results, the main impact factors that affect the results are identified, while a high-level explanation is provided that is clearly understandable by non-experts. The motivation is to create a first bridge between the telecommunication and the vertical domains and reach a common understanding in explaining what they can really expect from 5G.

1 Motivation

1.1 5G PPP TMV WG

The Test, Measurement, and KPIs Validation (TMV) Working Group was founded as part of the 5G PPP effort to promote commonalities across projects that have strong interest in Testing & Monitoring (T&M) methodologies needed to provide support to the vertical use cases in 5G Trial Networks. Such efforts include the development of test and measurement methods, test cases, procedures as well as the KPI formalization and validation to the greatest possible extent, to ensure a unique European vision on how the entire lifecycle of the 5G network, ranging from R&D to actual deployed environments, can be supported.

The Group is comprised by several Phase II and Phase III 5G PPP projects, and deals with the following research areas and technology domains:

- Testing KPI definition, KPI sources, collection procedures and analysis
- Testing frameworks (requirements, environment, scenarios, expectations, limitation) and tools
- Testing methodologies and procedures
- KPI validation methodologies
- Testing lifecycle (i.e., testing execution, monitoring, evaluation and reporting)
- Common information models for 5G T&M

Another important topic is the use of and contribution towards open-source projects such as OSM, OPNFV or ONAP and the identification of relevant exploitation and dissemination targets to promote the European vision on T&M towards a more global adoption.

1.2 Motivations

The facilities developed by the three ICT-17[1] projects are now further enhanced by on-site deployments of additional equipment in the ICT-19[2] projects. This is due to reaching out to vertical customers and establish "on-site" use cases where needed. One could think, for example, about a trial manufacturing line for Industry 4.0. The network needs to be deployed where the use case is located. While such facilities are to be considered at the forefront in the European telco industry, there are some common trends across the projects that need to be discussed to set the stage for understanding the result numbers coming out of the tests.

While a consistent number of Mobile Network Operators (MNOs) and Network Equipment Manufacturers (NEMs) are involved in ICT-17 and ICT-19 projects, most of the facilities can be considered experimental, in one way or another. There are facilities where research-based network elements are deployed, or others where NEMs are deploying and integrating their early

[1] Projects resulting from the H2020 5G PPP call https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-17-2018
[2] Projects resulting from the H2020 5G PPP call https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/ict-19-2019

products in complex multi-vendor environments endeavouring to meet requirements that exceed more traditional commercial deployments due to: 1) the expectation for flexible experimentation in these research environments; 2) the very high focus on providing very scattered multi-vendor solutions.

All the facilities involved, despite being constrained by project funding, do cover a wide set of 5G features and requirements, serving advanced 5G experimentations at a reduced scale. A very small number of base stations (generally one, sometimes a few) are installed, and a limited amount of cloud computing power is available. This last factor is worth mentioning since it impacts the scale of the core network deployment. One could argue that given the arguments that will follow, this might be the most significant, but it will be clear from the analysis of the results how an entry-level server deployment can have a severe impact on the network performance.

Continuing on the line of the core network, most of the facilities have deployed a Non-Stand-Alone (NSA) core. This means that performance-wise, it is not possible to leverage nor investigate the full power of 5G, given the fact we are still connecting to a 4G control and user plane. Stand-Alone (SA) deployments are still in early stages of deployment and utilization, since its full validation is mainly constrained by the lack of availability of UE's (especially of CPEs) in the market. Like we faced the issue in early 4G days, we need 5G terminals, preferably mmWave ones harnessing the large potential bandwidth offered by 5G.

Most of the vendors started first supporting the Verizon 5G specifications in April 2019 [1], de-facto delaying the introduction of the 3GPP ones. When that happened, of course NSA was the only supported functionality. Now, in the midst of Covid-19 pandemic, there is a sort of chicken and egg problem: SA UEs are delayed, and the few available ones might have compatibility issues. MNOs would like to introduce SA but they are facing lack of UEs. On the other hand, UEs will not be introduced until there are enough SA deployed to justify the release.

All the highlighted factors about the status of the facilities are somehow impacting the type of experimentation and testing that the ICT projects can effectively perform. One of the primary challenges that the projects faced, is that the range of architectures, technologies, standards and potential configurations for 5G systems was very broad, demanding deep learning curves and intense cross-discipline collaboration before actually putting the system in place. Transforming the high level KPI (latency, speed and reliability) into real system is much more than just slideware, it's real complex engineering. That led all the involved parties to focus on making the system work in a first place. This is not only valid for the 5G network performance, but for the vertical applications as well. Vertical customers and developers are facing a strong incognito as well; todefine properly their exact needs on a new technology. Service maturity nor infrastructure maturity was helping to go straight to the final configuration. Many paths, many features, many configurations exist brining the challenge to make the proper choices The type of test that got the primary focus was of course functional, addressing all the issues one by one, and making sure that all the components could interoperate.

As previously mentioned, UEs (especially Industrial CPEs) are rare goods. The ones that are available are flagship devices that have a high cost associated. This means all facilities are relying on a handful of devices for performing their tests. While this is an excellent starting point for a functional test, it unfortunately reduces the number of scenarios and KPIs that can be validated in a performance test. The verification of some of the KPIs might require, if done effectively end-to-end, the presence of extremely expensive, high-performance, dedicated test equipment, such as UE emulators, to generate the needed load to verify quickly the associated KPIs. Finally, the 2020-2021 pandemic played an important role in the ability for the different facilities to perform on site measurement campaigns.

After all these trends and limitations, it is reasonable to question what type of results we can actually extract from the facilities, and how these can help in the verification and validation process of the 5G performance KPIs. While the measurement results obtained by the facilities might seem limited, they start painting a picture about what 5G really is. A picture that might still be blurred, and whose colors might still be pale, but that start having a clear outline.

By looking at the numbers, one could start arguing why they don't reflect the shiny performance that have been declared in the past years. And why very similar facilities are generating very different performance numbers. As previously mentioned, the wide range of 5G configurations and options needed time to be understood and properly setup. Finally, system in actions will of course be resource constrained amongst all the services that needs to run concurrently.

This white paper targets at clarifying the details behind the performance numbers and at providing a series of interpretation guidelines that could help the reader better understanding the 5G domain. In the white paper there will be identified what are the main impact factors that affect the results and provide a high-level explanation that is clearly understandable by non-experts.

1.3 Traffic across the network: test vs application traffic

One of the most common misconceptions about generating results is the fact that any traffic passing through the network can be measured and used to extract any KPI. This fact is typically due to a more superficial understanding of the KPIs in question, and how different types of traffic affect them. Just to provide a concrete example of what is often heard, is that one could measure throughput out of vertical application traffic, as much as test traffic. This statement is superficially true, in the sense that throughput is measured, but **what throughput?** Before further diving into the example, let's start defining what **Testing** and **Monitoring** process are.

Testing (or Active Testing) provides a greater observability due to the active control over the type and intensity of traffic that is pushed through the network and through subsets of the network elements. This provides more degrees of freedoms in selecting what can be tested and measured (e.g., scalability or security resilience). **Monitoring** is instead a generally passive process (in comparison to the active testing) that is providing metrics from various components/layers of the 5G network.

Let's go now back to our previous example. If Testing is performed on the network, it can be loaded as much as needed. It is therefore possible to get parameters such as *Peak Throughput* but also *Average User Throughput* by emulating enough users with average traffic characteristics. Let's now imagine that the network is simply monitored, and it is loaded only by a vertical application traffic. If the application requires a throughput that is remarkably lower than what the network can carry, what is the monitored throughput telling us? It is only possible to infer that the network can support the application, but nothing about the performance characteristics of the network itself.

Generalizing the previous example, the KPIs that can be measured via testing are substantially different than through monitoring alone. An example of these differences can be seen in two similar documents coming from NGMN [2] and 3GPP [3]. The former is focused on testing aspects, while the latter provides an overview of KPIs to be measured during normal network operations.

Both testing and monitoring processes are vital across the whole lifecycle of a network, but they need to be carefully planned and applied at the right point of the lifecycle itself. This is the first principle that needs to be considered when analysing the results that will follow in this white paper.

2 Technology head-to-head: theoretical comparison between 4G and 5G

In the whitepaper we try to present the main impact factors of 5G after the analysis of the 5G results from the 5G PPP projects. We start this analysis by presenting the theoretical results from both 4G and 5G as presented from the Independent Evaluation Group of 5GIA and other related studies.

The KPIs considered to compare 4G & 5G are bandwidth, peak data rate for uplink, peak data rate for downlink, user experience data rate for uplink, user experienced data rate for downlink, user plane latency, control plane latency, reliability, and area traffic capacity.

Bandwidth is defined as the maximum aggregated system bandwidth in Hz, including frequency guard bands. The maximum supported bandwidth may be composed of either a single or multiple radio frequency (RF) carriers.

Peak data rate is the maximum achievable data rate under ideal conditions (in bit/s), which is the received data bits assuming error-free conditions assignable to a single mobile station, when all assignable radio resources for the corresponding link direction are utilized (i.e. excluding radio resources that are used for physical layer synchronization, reference signals or pilots, guard bands and guard times).

User experienced data rate is the 5% point of the cumulative distribution function (CDF) of the user throughput. User throughput (during active time) is defined as the number of correctly received bits, i.e., the number of bits contained in the service data units (SDUs) delivered to Layer 3, over a certain period of time.

The user plane (UP) latency is defined as the delay necessary to transmit data between the gNB and the UE. It consists of the transmission ($\tau 1$), HARQ request ($\tau 2$) and retransmission ($\tau 3$) between both entities.

The control plane (CP) latency in 5G NR refers to the UE transition time required from inactive to connected state.

Reliability relates to the capability of transmitting a given amount of traffic within a predetermined time duration with high success probability. Reliability is the success probability of transmitting a layer 2/3 packet within a required maximum time, which is the time it takes to deliver a small data packet from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point of the radio interface at a certain channel quality.

Area traffic capacity is the total traffic throughput served per geographic area (in Mbit/s/m2). The throughput is the number of correctly received bits, i.e., the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time.

The summary of 4G and 5G theoretical performance on a set of basic KPIs is illustrated in Table 1. The summary is generated based on the results of the final evaluation report from the Independent Evaluation Group of 5GIA ([4], [5]), IMT-2020 and IMT-Advanced. Table 1 illustrates for each KPI, the theoretical performance for both 4G and 5G, the threshold set by the IMT-2020 and some basic parameter values for the scenarios (both 4G and 5G) in which this performance result was observed.

All the evaluation studies assume two frequency ranges: FR1 (sub-6 GHz) which spans between 450 MHz and 6000 MHz and; FR2 (mmWave) which covers the range between 24.25 GHz and 52.6 GHz. As illustrated in the table, regarding bandwidth, 4G can aggregate 32 Component Carriers (CCs) of 20 MHz offering up to 0.64 GHz bandwidth, while 5G can reach up to 6.4

GHz bandwidth by aggregating 16 CCs of 400 MHz both in FR1 and FR2. This aggregation allows 5G to transmit in FR1 at a peak data rate of 78 Gbit/s on downlink and 40 Gbit/s on uplink, compared to the 27 Gbit/s and 13 Gbit/s respectively theoretically achieved by 4G. In addition, the use of FR2 in 5G (which is not supported in 4G) can further increase peak data rates to 174 Gbit/s and 95 Gbit/s on downlink and uplink. It can be observed that all theoretical values of peak data rate are far above the IMT-2020 target of 20 Gbit/s.

Table 1: 4G and 5G theoretical performance

KPI	4G	5G	IMT-2020 threshold	4G scenario	5G scenario
Bandwidth (GHz)	0.64	6.4	1	FR1	Both in FR1 and FR2
Peak data rate (downlink) (Gbit/s)	27.82 (FR1)	78.05 (FR1), 173.57 (FR2)	20	FR1: FDD, FR1, 32X20 MHz bandwidth, 1024QAM, MIMO	FR1: FDD, 16X100 MHz bandwidth, MIMO FR2: TDD, 16X400MHz bandwidth, MIMO
Peak data rate (uplink) (Gbit/s)	13.28 (FR1)	39.99 (FR1), 94.57 (FR2)	20	FR1: FDD, FR1, 32X20 MHz bandwidth, 256QAM, MIMO	FR1: FDD, 16X100 MHz bandwidth, MIMO FR2: TDD, 16X400 MHz bandwidth, MIMO
User experienced data rate (downlink) (Mbit/s)	10	111.45 (FR1), 104.71 (FR1+FR2)	100	IMT-Advanced	FR1: TDD, 16CCs, Dense Urban – eMBB FR1+FR2: Macro + micro layer at 4GHz
User experienced data rate (uplink) (Mbit/s)	5	59.2 (FR1), 64.4 (FR1+FR2)	50	IMT-Advanced	FR1: FDD, 16CCs, Dense Urban – eMBB FR1+FR2: TDD on 30 GHz, supplementary uplink (SUL) band on 4 GHz
User plane latency (ms)	0.69	0.24	1	FDD, initial transmission error probability=0, PRACH length = 2 OFDM Symbols	FDD, initial transmission error probability=0, PRACH length = 2 OFDM Symbols
Control plane latency (ms)	16	11.6	20	TDD	TDD, PRACH length = 2 OFDM Symbols
Reliability	99.9%	99.999995%	99.999%	IMT-Advanced	NLOS, Urban Macro
Area traffic capacity (Mbit/s/m2)	0.1	12.19 (FR1), 17.43 (FR2)	10	IMT-Advanced	FR1: 160 MHz bandwidth, 3 sectors per site, Indoor Hotspot – eMBB FR2: 400 MHz bandwidth, 3 sectors per site, Indoor Hotspot – eMBB

The user experience data rate presents the actual throughput made available to each user. In 5G, a value of 111 Mbit/s can be achieved on downlink in FR1 in dense urban environment when 16 CCs are utilised. In FR2, additional micro cells in FR1 are required in order 5G can accomplish values above the 100 Mbit/s IMT-2020 threshold. In the uplink, the performance of user experienced data rate is 59 Mbit/s in FR1 scenarios and 64.4 in combined FR1 and FR2 scenarios.

The studies on user plane latency illustrates performance values of below 1ms target both for the 5G and 4G assuming a set of favourable configurations e.g. zero initial transmission error probability and PRACH length of 2 OFDM Symbols. The theoretical studies present values of 0.24 ms and 0.69 for 5G 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 TDD and small PRACH length.

Regarding reliability, the IMT-2020 target is set to five-nines. The theoretical studies demonstrated values above seven-nines for the 5G in urban macro scenarios with NLOS transmissions. Finally, as far as traffic capacity is concerned, 5G achieves higher performance values compared to the IMT-2020 target of 10 $Mbit/s/m^2$, reaching to 12 $Mbit/s/m^2$ in FR1 scenarios and 17 $Mbit/s/m^2$ in FR2 scenarios.

3 Results analysis and performance impact factors

The reflection of the theoretical values (presented in the previous section) to achievable ones over real 5G deployments is a challenging yet a prerequisite task for revealing performance impact factors and deployment dependencies. To this end, 5G experimentation results from 5GPPP ICT-17 projects have been collected and studied, providing us with an initial set of insights.

The performance impact factors can be classified into three main categories, namely:

- *The deployment and configuration aspects*, referring to the technology that is selected for each domain of the service provisioning chain (Core, Transport, and Radio domains) and, also, to set-up configurations, such as the selection of bandwidth size and antenna layers.
- *The scenario under which an experiment/measurement is performed*, referring to run-time factors that affect the performance values, such as the mobility of end devices, the channel conditions, and the traffic type used.
- *The testing/experimentation procedure,* referring to the methodology used for collecting measurements and calculating the KPI values, such as experiment repetitions applied, outlier values identification, and mean values calculation process, i.e., factors that affect the reliability of the measurements.

Regarding the last category, 5GPPP TMV WG had published the basic concepts on experimentation methodology for 5G KPIs [6], guaranteeing the reliability of the measurements conducted. In addition, 5G PPP infrastructure projects (ICT-17) contribute to a recent ETSI Technical report (TR 103 761 [8]) related to Core Network and Interoperability Testing (INT) Methodologies for E2E Testing & Validation of Vertical Applications over 5G & Beyond networks.

The 5G experimentation results, used for this analysis, consider well-controlled scenarios, e.g., indoor or in campus scenarios, with small number of static devices, under good channel conditions (Line of Sight) - LOS. Thus, the impact of factors of the second category is minimized, allowing us to extract insights related to 5G deployment and configuration aspects.

The focus is on two deployment factors, namely *the transport network characteristics*, and *the network core deployment type*, as well as on five configuration aspects, namely *the bandwidth size*, the *MIMO layers* in RAN, the *UL/DL intensive patterns*, the *scheduling approach*, and the *target coverage.*

3.1 Impact of transport network characteristics

The theoretical values depicted in Table 1 have been extracted analytically based on the 5G NR structure (PHY structure). Thus, they represent the best performance values expected at the MAC layer, when any other domain of the service provisioning chain is omitted. In this subsection, the impact of the transport network domain (connection type with and without functional split) on those values is examined, focusing mainly on the data rate and latency KPIs when UDP and TCP traffic is used.

3.1.1 Satellite backhauling

The impact of using a Satellite link in the transport part of the end-to-end service provisioning chain has been studied [7]. Tests were conducted, using the iPerf3 utility, to measure throughput on the forward and return satellite links, and, also to determine UDP jitter and ICMP Round Trip Time (RTT). iPerf3 employs a client/server model, in which the client sends data to the server for a specified duration, after which both server and client report the resultant performance metrics (i.e. throughput for TCP, and throughput & jitter for UDP). On the downstream path, the iPerf3 server was deployed on a SAT MEC server, and the iPerf3 client on a SAT GW FN; for the upstream path the opposite was true. The satellite network validation test topology is illustrated in Figure 1.

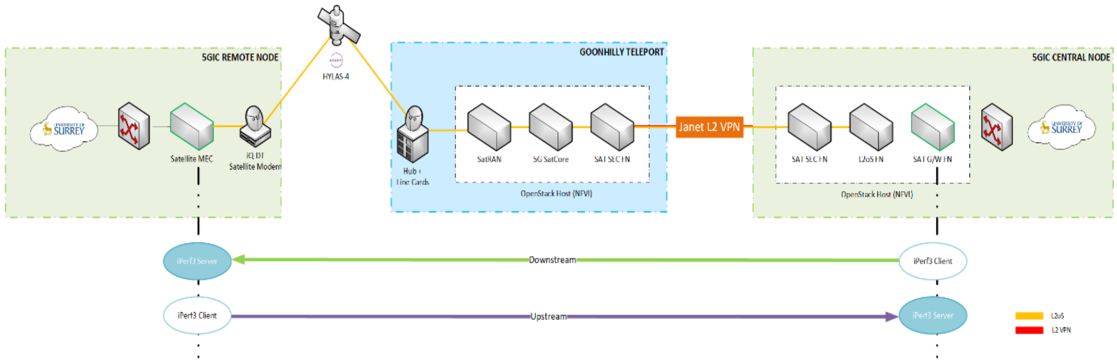

Figure 1: Satellite network validation test topology

Satellite capacity for the testbed was provided, via a Geostationary Orbit (GEO) satellite. The main characteristics relevant to the performance of the forward and return satellite carriers are depicted in Table 2. It is worth noting that the bandwidth allocated for the return carrier is quite conservative (5 MHz); this is intentional, since the focus is primarily on downstream performance. However, one side effect of limited upstream bandwidth is poor performance, particularly for tests involving TCP (which is further exacerbated by the lack of a Performance Enhancing Proxy (PEP) at the satellite network edge). A set of various tests conducted, with main results those depicted in Figure 2. The key outcome is the fact that, for the used carrier attributes, a satellite link adds about 600ms delay (measured in RTT) and supports on average a rate of 25 Mbps.

Table 2: HYLAS-4 Carrier Attributes

Carrier Direction	Occupied Bandwidth (MHz)	Symbol Rate (Msps)
Forward (Downstream)	24	20
Return (Upstream)	5	4.1667

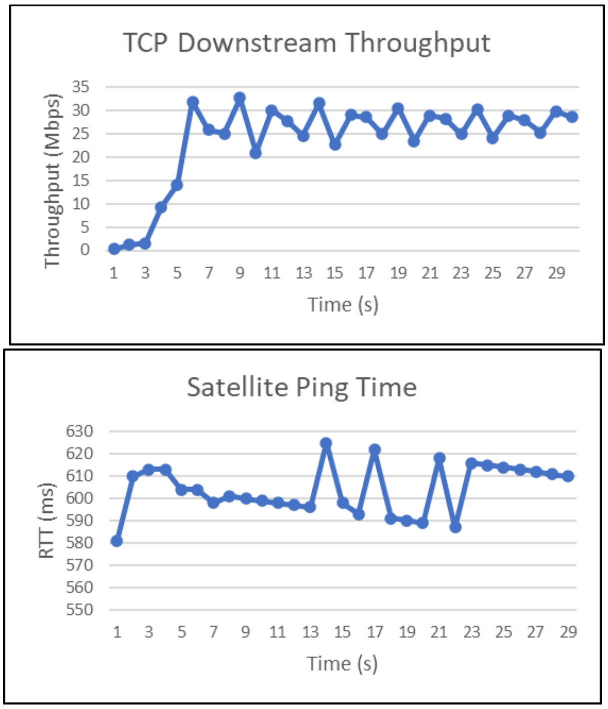

Figure 2: Satellite Downstream TCP Throughput and RRT

3.1.2 Fiber optic backhauling

Any use of ethernet-based or other type of links in the backhaul connection (even when no functional split is applied) burdens the end-to-end performance of a 5G network, since the target KPI values have been based on the theoretical calculation assuming the RAN domain only. To best measure the performance of an end-to-end system, the 5G NSA platform tied to minimize that impact by using fiber optics in the backhaul links. The deployment characteristics of the set up are depicted in Figure 3.

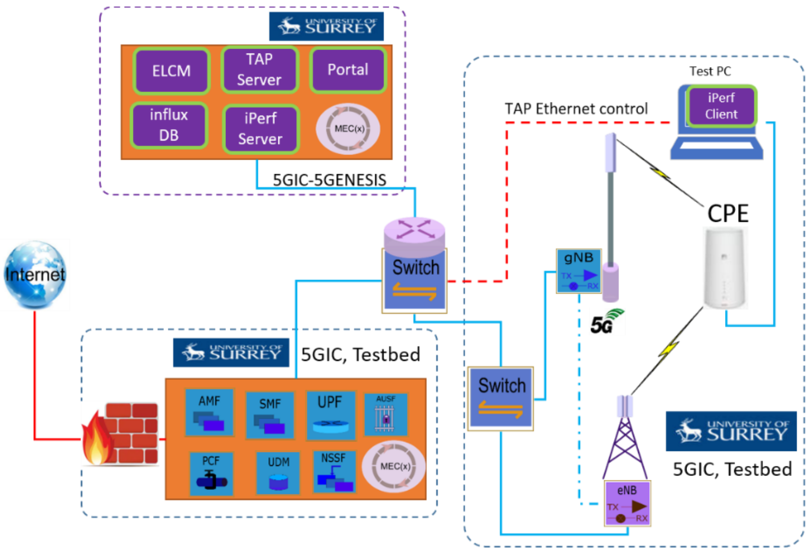

Figure 3: Architecture of the 5G NSA Surrey platform (blue lines depict the fiber links)

With the use of fiber optics, the measured RTT values are around 10ms (with peak value at around 8ms). Figure 4 depicts the actual RTT values perceived in 25 iterations of the experiment.

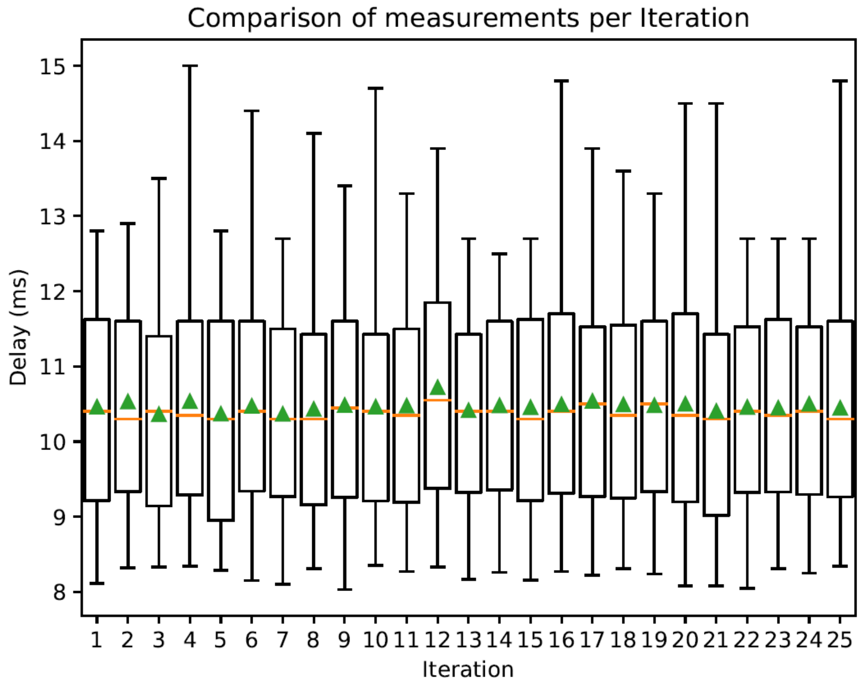

Figure 4: Round Trip Time – RTT test results with fiber optics backhaul.

Experiments with fiber optical backhauling were run also in Industry 4.0 verticals, in a specialised pilot. The transport network is used in this pilot to connect the Baseband Units (BBU) to the Radio Units (RU) of the radio cells, implementing centralized RAN architecture, where Baseband units serving several radio sites are centralized in a single location as depicted in Figure 5. A NSA (non-standalone) configuration is considered, thus both a 5G and a 4G radio base stations are needed to connect the User Equipment. On the left side there are the Baseband Units used for 4G (DU) and for 5G (BBU). Three clients, one CPRI Option 3 (2.5 Gb/s) link for 4G and two eCPRI links (10 GBE) for 5G are connected to a Hub node that performs switching and framing in the digital domain, and DWDM multiplexing functions. On the other end of the DWDM connection, realized with a fiber ring in order to offer physical protection, there is a Remote Node that performs wavelength selection with an OADM (Optical Add and Drop Multiplexer), and de-framing and switching in the digital domain. The three clients are then connected to Remote Radio Unit (4G) and Advanced Antenna System (5G).

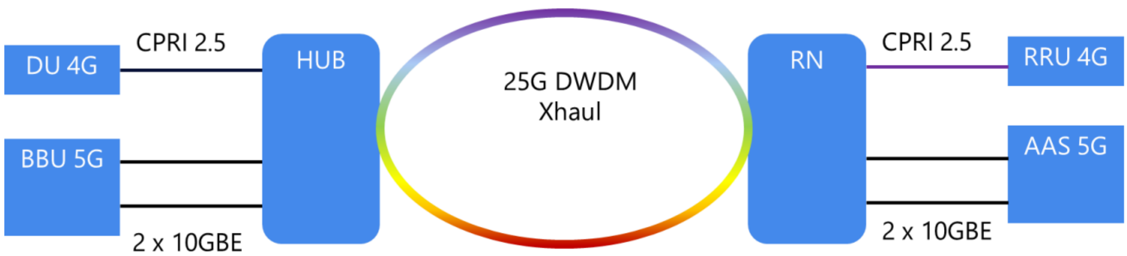

Figure 5: Transport network user plane

The set of measurements presented here are relative to the User Plane function of the transport network, to show that the performance of the network are adequate to support desired services without disruption. In that perspective, the reference methodology used is that described in IETF's RFC2544 35 [9].

The purpose of the measurements was to show that the performances of the transport network are adequate to support 5G and 4G fronthaul connections, based on two 10GBE eCPRI links (5G), and one CPRI link without introducing degradation or disruption. The following tests have been performed on the 10GBE links used for eCPRI:

- Latency: measures the time taken by a test frame to travel through a network device or across the network and back to the test port. Latency is the time interval that begins when the last bit of the input frame reaches the input port and ends when the first bit of the output frame is seen on the output port.
- Throughput: measures the maximum rate at which none of the offered frames are dropped by the device/system under test (DUT/SUT). This measurement translates into the available bandwidth of the Ethernet virtual connection.
- Back-to-back burst: measures the longest burst of frames at maximum throughput or minimum legal separation between frames that the device or network under test will handle without any loss of frames.
- Frame loss: defines the percentage of frames that should have been forwarded by a network device under steady state (constant) loads that were not forwarded due to lack of resources.

For what concerns the CPRI link, a functional test has been performed: the test showed that regardless the traffic load of the 10GBE clients, the 4G base station is up and running and no alarms are showed by RBS management system.

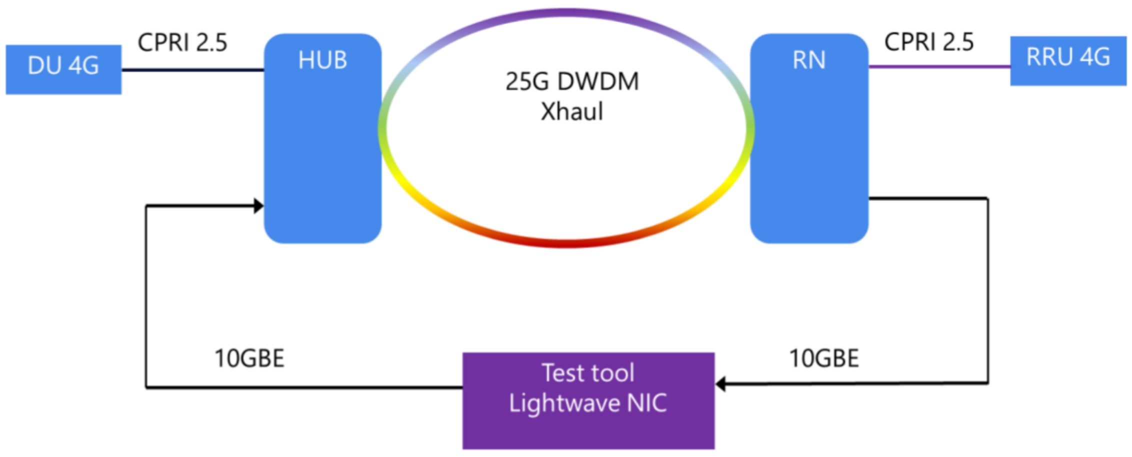

Figure 6: Test setup

A test instrument (NIC from Lightwave) has been used to perform RFC2544 compliant measurement on a 10 GBE link, while the other 10 GBE and the CPRI 2.5G links were fully loaded. The E2E latency experienced by the clients of the infrastructure, is constituted by the sum of the mobile contribution and the transport contribution.

The used transport network fiber span of 8.8 km introduces a delay of 44 µs due to light propagation in glass (5ns/m). The measurement reported in Figure 7 shows that the latency of the transport network is essentially due to the fiber span (44 µs); the additional delay due to the digital processing (i.e., switching and framing) is below the instrument precision (1 µs). Latency is independent of packet size or line load.

RFC Through Put Test

Stream Id	# 16	MAC Dest Addr	00-00-00-00-00-00
Trial Duration	10 (Secs)	Dest Port	Packet(same)
BW Ceiling	100.00 %	BW Floor	10.00 %

Accept Lost Rate – 0 % Resolution Rate – 1.00 %

Latency Iterations - 20

Size	Rate	Tx Packets	Rx Packets	Latency (usec) Min	Max	Avg
64	100.00	148809559	148809559	44	44	44
128	100.00	84463760	84463760	44	44	44
256	100.00	45291819	45291819	44	44	44
512	100.00	23497278	23497278	44	44	44
1024	100.00	11973693	11973693	44	44	44
1280	100.00	9615773	9615773	44	44	44
1518	100.00	8117317	8117317	44	44	44
9000	100.00	1385868	1385868	44	44	44

Figure 7: Latency test results

A shown in Figure 8, the actual throughput is equal to the theoretical one, regardless the frame length (100%).

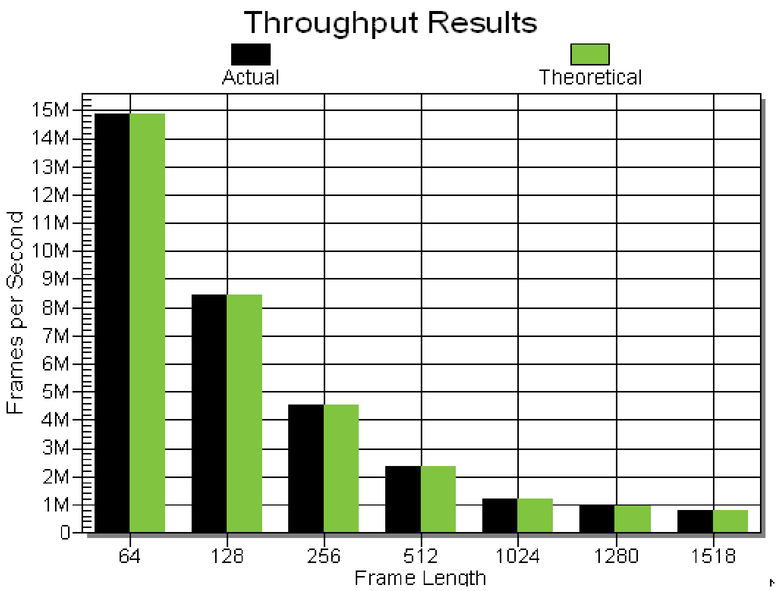

Figure 8: Throughput test results

Figure 9 shows that no frame loss has been detected in the optical communication, regardless the frame length or the traffic load. The back-to-back burst value is the number of frames in the longest burst that the DUT will handle without the loss of any frames. The test shows that no frames are lost regardless the burst length. In the following figure the results are shown for a 10-seconds long burst; longer bursts showed the same result.

Frame Size (bytes)	Max Possible Number of frames	Avg Observed B2B Frames
64	148809520	148809520
128	84459440	84459440
256	45289840	45289840
512	23496240	23496240
1024	11973180	11973180
1280	9615380	9615380
1518	8127420	8127420
9000	1385800	1385800

Figure 9: Back-to-back burst test results

In the end, the tests performed on the transport infrastructure, based on RFC2544 standard, proved that the network performance are more than adequate to support the architecture adopted in the pilot. The latency introduced by switching and framing layer is negligible for the application of interest. No frame loss or limited throughput has been observed during the tests. These considerations, related to transport performances, also applies to the other two use cases.

3.2 Core deployment edge vs central

One of the key new features introduced by the 5G Core is the possibility to offer "concurrent access to local and centralized services" ([20]). In practice that means to allow the deployment of services close to the end-user, in order to improve the end-to-end service performance indicators.

The way 3GPP SA2 ([21]) standardizes this feature is introducing the control and user plane separation (CUPS [22]) natively in the 5G Core architecture. With this split, User Plane functions can be placed independently from Control Plane functions. Additionally, 5G Core supports the concatenation of User Plane functions using a new interface (N9). All these features allow to improve the flexibility of the 5G Core and meets the requirement of supporting, at the same time, local and central services, as is described in the next picture:

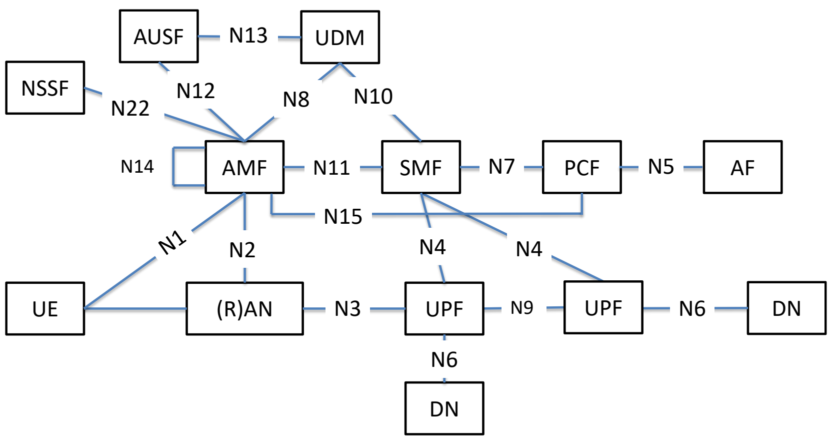

Figure 10: Concurrent access to local and central DN ([21])

The importance of the local services is mainly in their influence in the improvement of key performance indicators. One of the improved 5G KPIs improved is latency, as with a local deployment approach the service is closest to the end user in terms of distance, and the time required for data transmission is related to the physical distance: the light needs 1 millisecond to cover almost 200 km of distance (assuming glass fiber optical link – since we assume vacuum and index of refraction for glass equals to 1.5). A good approach that considers the latency of the processing systems, is to assume that 100 km introduces a One-Way Delay of 1 millisecond. Taking that into consideration, we can define an edge deployment by comparing the latency introduced due to distance with the latency introduced by the 5G NR:

- For eMBB services the requirement for 5G NR latency is 4 milliseconds. With that, we can define edge deployments for eMBB as those services deployed in the range of units and tens of kilometers up to one hundred kilometers away from the RAN.
- In case of URLLC, the latency requirement for 5G NR is 1 millisecond. For this reason, local services must be deployed for URLLC as close as possible to the end-user location (units of kilometers of order of magnitude)

Another KPIs improved by the use of an edge deployment of the 5G Core is the user data rate, both peak and user experience data rate. The rationale is that for short distances is easier and cheaper to have more bandwidth than for long areas and distances, allowing to have a peering or a point-to-point connectivity between the 5G Core edge and the local service.

Finally, for the same reasons as for the data rate improvement, and combined with the optimization in latency levels, the Reliability KPI is also improved, thus allowing better SLA agreements for local connectivity.

Table 3: Distance of the cloud environment depending on latency requirements

	Local Edge	Near Edge	Central office
	Units of Kilometers	Tens of kilometers	Hundreds of kilometers
URLLC	Performance: Optimal Investment: Very High Efficiency: High	Performance: Limited Investment: High Efficiency: Very High	Performance: - Investment: - Efficiency: -
eMBB	Performance: Optimal Investment: Very High Efficiency: Low	Performance: Optimal Investment: High Efficiency: Very High	Performance: Limited Investment: Low Efficiency: High

Putting it all together, as Table 3 reflects, and based on the measurements collected for a variety of deployment approaches for services and core, from the closest edge local deployments to central office model, the applicability to URLLLC and eMBB 5G services may be expressed and correlated to simple technical and economical criteria, namely Performance, Investment and Efficiency:

- Performance is related to the feasibility to meet the expectation levels for the type of 5G Service supported, as discussed above, with major focus on latency constrains.

- Investment is related to the involved CAPEX and OPEX for the considered deployment models, for both the Communication Service Provider and the Vertical/Enterprise.
- Efficiency is related to the comparative usage of resources (HW, links, bandwidth, energy consumption, etc.,) at the service of the expected level of performance.

The trade-offs implied in Table 3 can only be fully and properly assessed over concrete cases of application. That said, the thumb-rules derived from this analysis are:

- For URLLC-exclusive scenarios central office deployments are discouraged, and the decision to go for either local Edge or near Edge approaches bases first on actual latency requirement feasibility and secondly on economic factors.
- For mixed URLLC-eMBB scenarios the sweet spot approach is clearly near Edge model. Other approaches would significantly compromise either performance or investment and efficiency criteria.
- For eMBB-exclusive scenarios, the safest option to go for is also near Edge, although for some geographies and cases Central Office deployments could represent a better trade-off.

When it comes to the relative positioning of 5G NSA and 5G SA in this context, the outcomes of our experimentation can be summarized as follows:

- For local Edge deployment, performance-wise, 5G NSA and 5G SA deliver very similar figures for latency and throughput. But the key point here is the investment required, induced by the involved technology complexity, and 5G SA is clearly the option to favour for this type of deployment.
- For near Edge deployment, performance-wise, 5G SA delivers better figures for latency whilst 5G NSA is somewhat favoured for eMBB cases where UL User Data Rate is a critical KPI, thanks to the possibility to use 4G-5G aggregation approaches for optimizing it while preserving good average latency levels. 5G NSA and 5G SA are on par for this approach, even though for technical and economic sustainability and its advanced slicing features 5G SA shows better suited.

3.3 Bandwidth impact

The bandwidth sizes foreseen for the 5G NR system vary, as depicted in Table 4, and they can reach the size of 400 MHz in the Frequency Range 2 (FR2: frequencies above 24 GHz) for a single component carrier (CC), while up to 16 CC can be combined (i.e., in total 6.4 GHz bandwidth can be achieved).

Table 4: Main bandwidth sizes in 5G NR

<table>
<tr><th>FR</th><th>BW (MHz)</th><th>Max CC</th><th>Total BW (MHz)</th><th>Subcarrier spacing options</th><th>Slot frame indicator- SFI</th></tr>
<tr><td rowspan="2">FR1</td><td>50</td><td rowspan="4">16</td><td>800</td><td rowspan="2">15 kHz, 30 kHz, 60 kHz,</td><td rowspan="4">56 different combinations of DL, UL, and Flexible mode for the 14 symbols of a slot</td></tr>
<tr><td>100</td><td>1600</td></tr>
<tr><td rowspan="2">FR2</td><td>200</td><td>3200</td><td rowspan="2">120 kHz, 240 kHz</td></tr>
<tr><td>400</td><td>6400</td></tr>
</table>

As in any other wireless system, in the 5G system the bandwidth size is a primary factor that affects the performance perceived by the end users (mainly referring to the data rate KPI). However, 5G NR has introduced high flexibility in structuring the available bandwidth. Using different numerology [11] for the same size of bandwidth the achievable performance varies. Mainly, two key factors define the size of the available resources within a specific bandwidth, and thus affect the expected data rate values. First, in the frequency domain, it is the size of *subcarrier spacing* which affects the number of subcarriers to be used. For example, 50 MHz bandwidth can be used by 1600 subcarriers with 30 kHz spacing, while 40 MHz bandwidth can be used by 2600 subcarrier with 15 kHz spacing. Second, in the time domain and for TDD mode of operation, it is the *Slot Frame Indicator (SFI)* which defines how the available OFDM symbols are split to support DL and UL transmissions. For example, for SFI:28 heavy DL transmissions can occur (to support an eMBB slice), limiting the UL transmission resources. Also, in the two extremes, the slot can be only for DL or only for UL, when SFI:0 or SFI:1 is used, respectively.

From the perspective of the implementation options defined for the 5G system, the Non-Standalone option (addition of gNBs under the control of EPC – 4G core) is one of the most suitable options for a fast proliferation of the 5G in the market. Also, in EU, frequencies in the Frequency Range 1 (FR1: frequencies up to 6 GHz) are already assigned for 5G systems and they are being used in experimentation and commercial 5G deployments. In this context, the achievable KPIs for different bandwidths within the FR1 of a 5G NSA system are examined, as an effort to quantify data rates that end users can experience via the early 5G deployments.

For the experimentation process presented here, bandwidths of 40, 50 and 100 MHz were considered in band N78 (i.e., 3.5 GHz). DL and UL throughput measurements were conducted in 5G experimentation platforms [12], using UDP and TCP traffic through the Iperf3 tool. The results are depicted in Table 5. It is noted that the channel conditions were almost ideal, with UEs and gNB in fairly close distance.

Table 5: Throughput experimentation results for various bandwidth sizes in band n78

5G NR BW size (MHz)	DL/UL	Traffic Type	subcarrier Spacing	Throughput		Relevant theoretical peak values (achieved at the MAC layer, with SFI:28)
				Mean value	Peak value	
40	DL	UDP	15 kHz	264.74 Mbit/s	269.88 Mbit/s	322 Mbit/s
50	DL	UDP	15 kHz	369.27 Mbit/s	372.47 Mbit/s	402 Mbit/s
100	DL	TCP	30 kHz	492.08 Mbit/s	738.08 Mbit/s	814 Mbit/s
	UL	TCP	30 kHz	59.34 Mbit/s	73.40 Mbit/s	134 Mbit/s
	UL	UDP	30 kHz	119.55 Mbit/s	125.08 Mbit/s	

As can be observed in Table 5, the results of the experimentation process (measurements) are below the theoretical values, but very close to them. Considering that the measurements refer to transport layer (TCP and UDP have been used) it is safe to claim that the theoretical values are validated. Reasonably the TCP experiments have lower performance than the UDP ones due to that acknowledgement needed in the communication. This is mainly depicted in the UL measurements.

3.4 MIMO layers impact

The following measurements have been obtained at the Patras5G facility site deployed in Greece using the NGMN pre-commercial trials framework V3.0 definition for peak user throughput. The scope of this sub-section is to discuss the impact of the different MIMO configurations on peak user throughput, defined as the maximum DL/UL data rate achievable for a single user located at the best location within a cell.

Measurements have been obtained with two different systems under test, namely a CPE and a UE, provided by different equipment vendors. Furthermore, measurements were obtained for NSA and partly SA configurations.

Table 6: NSA Test Configuration with CPE (Conf. 1 (eMBB service), Location: Indoor, Distance: ~3m)

BS	Mode	CPE	Commands
Vendor not disclosed	NSA/TDD 7:2	Vendor not disclosed	iperf3 -s -i 1 –u iperf3 -c 192.168.3.2 -i 1 -u -b 300M -t 100

Bandwidth (MHz)	MIMO	DL UDP (Mbit/s)	UL UDP (Mbit/s)	DL TCP (Mbit/s)	UL TCP (Mbit/s)
50	2x2	272	18.5	50.6	18.7
50	1x1	185	8.7	48.7	13
40	2x2	215	20.4	51.2	24.9

Table 7: NSA Test Configuration with UE NSA (Conf. 1 (eMBB service), Location: Indoor, Distance: ~3m)

BS	Mode	UE	Commands
Vendor not disclosed	NSA/TDD 7:2	Vendor not disclosed	iperf3 -s -i 1 –u iperf3 -c 192.168.3.2 -i 1 -u -b 300M -t 100

Bandwidth (MHz)	MIMO	DL UDP (Mbit/s)	UL UDP (Mbit/s)	DL TCP (Mbit/s)	UL TCP (Mbit/s)
50	2x2	291	32.6	162 (multi)	22.1
50	1x1	184	17.3	180 (multi)	16.8
40	2x2	213	23.4	146 (multi)	25.1

Table 8: SA Test Configuration with CPE (Conf. 1 (eMBB service), Location: Indoor, Distance: ~3m)

BS	Mode	CPE	Commands
Vendor not disclosed	SA/TDD 7:2	Vendor not disclosed	iperf3 -s -i 1 –u iperf3 -c 192.168.3.2 -i 1 -u -b 300M -t 100

Bandwidth (MHz)	MIMO	DL UDP (Mbit/s)	UL UDP (Mbit/s)	DL TCP (Mbit/s)	UL TCP (Mbit/s)
50	2x2	255	40.5	252 (multi)	40
50	1x1	143	14.3	53.2	15.1
40	1x1	98	10.7	48.4	10.1

Table 9: SA Test Configuration with UE (Conf. 1 (eMBB service), Location: Indoor, Distance: ~3m)

BS	Mode	UE	Commands		
Vendor not disclosed	SA/TDD 7:2	Vendor not disclosed	iperf3 -s -i 1 –u iperf3 -c 192.168.3.2 -i 1 -u -b 300M -t 100		
Bandwidth (MHz)	MIMO	DL UDP (Mbit/s)	UL UDP (Mbit/s)	DL TCP (Mbit/s)	UL TCP (Mbit/s)
50	2x2	325	16.9	148 (multi)	20.2
50	1x1	Not obtained	Not obtained	Not obtained	Not obtained
40	1x1	101	15.2	30.8	25.7

Overall, the performance figures need to be understood in the context of the available bandwidth (40 and 50 MHz) which restricts the peak achievable values. Furthermore, the equipment used for the measurements does not represent typical carrier grade equipment and should be considered as indicative performance values achievable with early prototype equipment.

The variation between different MIMO configurations (2x2, 1x1) measured for the same configured parameters otherwise, show that the gain is smaller than expected. In the case of DL UDP transfer in NSA mode at 50 MHz, 1x1 achieves a throughput of 185 Mbit/s, while 2x2 achieves 272 Mbit/s; a gain of about 50%. In the case of NSA this is consistent independent of whether a CPE or a UE is used and exhibits the same targets also for SA mode at 50 MHz. However, for the case of UL UDP transfer the gain is about 100%. The measurements in SA mode are generally less consistent and currently under investigation to isolate the root cause for the variation.

3.5 UL vs DL intensive patterns

High throughput with low latency, massively connected devices and effective utilization of spectrum can be realized by adopting the fifth-generation new radio air interface, known as 5G NR. To address these challenges, 5G NR uses different multiple access and modulation techniques.

From the perspective of service and network deployment, 5G uplink performance is critical to ensure diversified services development and guarantee user experience such as HD video, online games, big data collection, intelligent surveillance, or AR/VR live video.

The uplink is typically the limiting factor in LTE mainly due to differences of the transmit power, the TDD carrier frequency link budget and number of antennas deployed in the LTE node B versus devices. 5G in Sub-6 GHz FR1 spectrum provides significant advantages for users and carriers, increasing the device uplink transmit power to afford significant uplink coverage extension, improving the TDD competitiveness with respect to FDD deployment, providing a 3dB power increase, improve cell-edge spectral efficiency using higher order modulation and transport block size and enhancing the overall cell-edge performance.

As most low-range bands have been allocated to 2G/3G/4G mobile communication networks or other systems, 5G spectrum is mainly located in the mid-bands as 3.5 GHz. Due to the nature of radio signal propagation, the higher the frequency band is, the higher path loss and penetration loss it has, which further limits the coverage.

For example, at the same distance, the path loss of 3.5 GHz is 4.4 dB higher than that of 2.1 GHz, and in outdoor-to-indoor scenario, where wireless signals penetrate through glass windows or walls to reach indoor users, the penetration loss of 3.5 GHz is about 3 dB higher than that of 2.1 GHz.

Advanced technologies such as Massive MIMO have been introduced into 5G network to offset the propagation loss, this technique with the use of dynamic Time-Division Duplexing (TDD) enables flexible adjustments of uplink (UL) and downlink (DL) resources according to the instantaneous traffic load. This flexibility gives us the possibility to efficiently use the spectrum where all the 5G networks in the same area will be synchronized: All base stations will transmit in the same time slots and all the devices will transmit only in the same dedicated time intervals.

Frequency-Division Duplexing (FDD), that is the technique used for GSM, WCDMA and mostly in LTE, was a greater global uptake that has the advantages of not needing tight synchronization, as the transmitter and receiver operate at different carrier frequencies. This is useful in the case of symmetric traffic, since it does not waste bandwidth during the switch-over for transmitting to receiving and also radio planning is easier and more efficient, since base stations do not need to "hear" each other (as they transmit and receive in different sub-bands) and normally not interference each other.

Time-Division Duplexing (TDD) has a strong advantage in the case where there is an asymmetry to the uplink and downlink rates. The channel reciprocity TDD has against FDD, make the uplink measurements can be reused for downlink measurements, so as the amount of uplink data increases, more communication capacity can dynamically be allocated to that and, as the demand shrinks, capacity can be taken away.

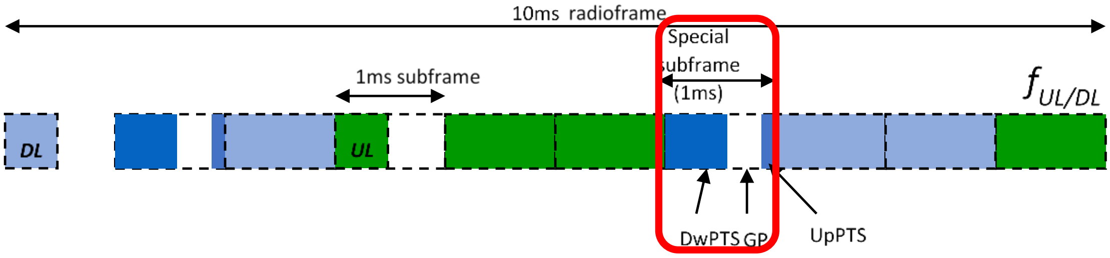

Figure 11: Time-division duplexing

5G NSA uses the solution EN-DC (E-UTRAN NR Dual Connectivity) or also called Architecture Option 3x, where the control signalling of 5G Radio is done by the LTE CORE using and 4G node.

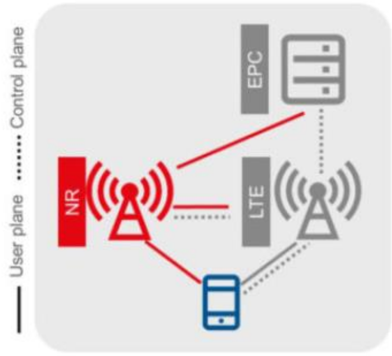

Figure 12: NSA Dual Connectivity

This NSA solution allows us to use the Dual Connectivity Uplink aggregation feature that permits to use the LTE leg for user data and aggregate both traffics bearers to have a big uplink throughput. This feature does not impact on LTE/NR nodes and provides improved reordering buffer handling introducing a timer that makes reordering process robust and efficient.

The peak Uplink throughput was tested for LTE, NR and NR+LTE UL traffic aggregation, obtaining the following values:

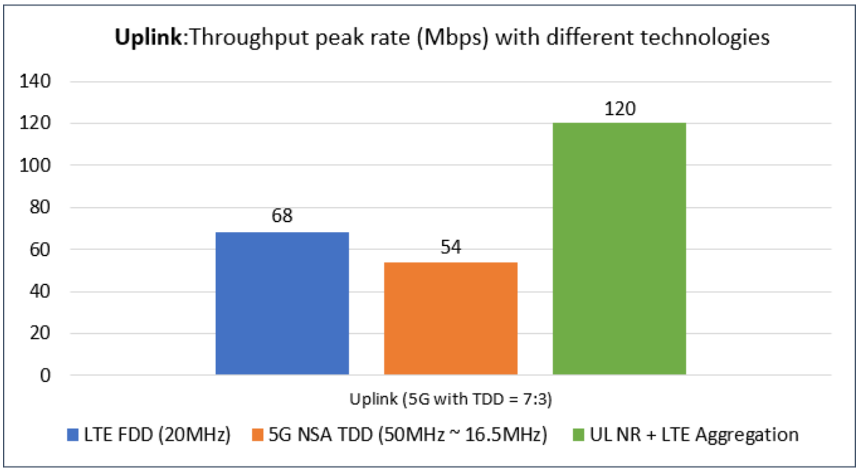

Figure 13: Uplink throughput peak rate

LTE uplink is higher than NR NSA uplink due to the TDD technique vs FDD. Even in NR we tested with bigger bandwidth: 50 MHz in comparison with 20 MHz, we are using the same band for DL and UL in an asymmetric traffic adding the switch-over for transmitting to receiving, however the TDD solution is more efficient in channel in a long communication as we can use the measurements to estimate the quality of the channel in order to use MIMO technology and synchronize the frames with contiguous base stations.

The use of feature NR+LTE traffic aggregation has a big impact in the throughput peak, as we benefit of adding both traffics bearers. This improvement was very important for the experiments performed with Verticals as the they demanded almost 100 Mbps in uplink to perform it successfully.

Note that a small variation in Uplink throughput is more significant than in Downlink throughput. Note the following graph, that represents the Downlink throughput peak in the LTE, NR and NR+LTE DL traffic aggregation cases. The improvement is from 712 Mbit/s to 875 Mbit/s = 163 Mbit/s and this represent an increase of 23% of downlink throughput peak.

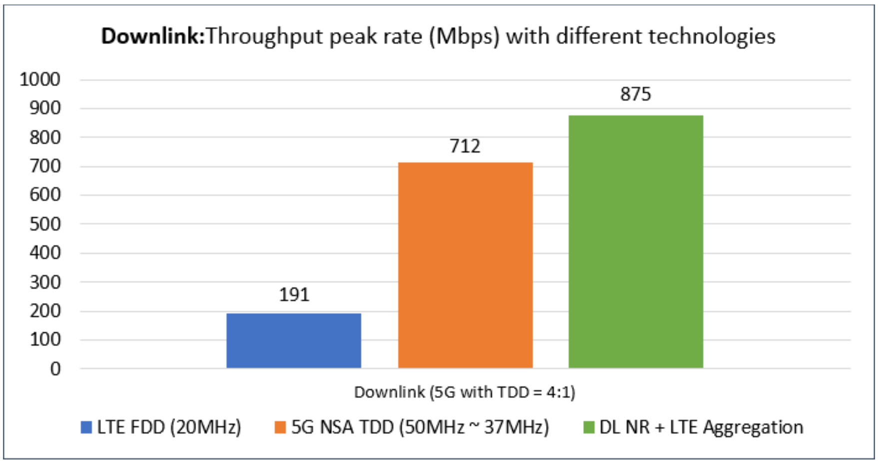

Figure 14: Downlink throughput peak rate

However, if we take the data from the Uplink table, the improvement from 54Mbps to 120Mbps = 66Mbps and that represents and improvement of 122% of uplink throughput peak.

3.6 Scheduling impact

Proactive scheduling is a valuable feature available in Nokia gNodeBs that provides significant reduction in the communication latency. As a short introduction, when this feature is enabled, the scheduler generates additional uplink scheduling grant in excess of those explicitly requested by the connected mobile device whose latency we want to reduce. In consequence, the benefits introduced by this capability are dependent on the uplink activity pattern at MAC level, which correlates with the nature of the end-to-end uplink traffic.

Let's analyse in more detail the underlaying radio technology aspects and use an example to illustrate how this feature can be of help, particularly at intervals where the traffic pattern is more disperse.

To understand how to improve scheduling to reduce latency, we need to understand how uplink grant allocation works. In 5G, as in LTE radio access technology, the allocation of uplink resources is fully controlled by the radio network. More specifically, the mobile devices do not have the authority to decide on their own which resources to use for uplink data transmissions, but only the network scheduler owns that authority.

To allocate an uplink grant, the network will typically provide the related information in a control message in the PDCCH control channel. These control messages are known as uplink DCI (Downlink Control Information). There are different formats or types of DCIs, but to convey uplink grant related information, typically DCI-0-0 or DCI-0-1 are used. Among other information, the uplink DCIs may indicate the allocated frequency resources defining the number of PRBs (Physical Resource Blocks) and the PRB offset where the mobile device must transmit. Additional fields can be used to control the time offset between the transmission of the uplink DCI and the time slot where the device will be allowed to transmit. However, before allocating any resource for a mobile device to use in uplink transmission, the network will normally need to know if/when this device wants to transmit.

For that purpose, Scheduling Request and Buffer Status Reporting are two key concepts involved in the provision of uplink resources for mobile devices to use in their transmissions. The first one, scheduling request, is a control mechanism used by a connected mobile device, to request time and frequency resources where to transmit uplink data. The full detail of the scheduling request procedure is beyond the scope of this document and are defined in TS 38.321 [13] section 5.4.4.

Let’s just mention, that the mobile device notifies to the network, using the Physical Uplink Control Channel (PUCCH), that it wants to transmit uplink data when provisioned with uplink resources.

However, in general, the mobile device cannot apply immediately for uplink resources sending the scheduling request, but only at predefined periodicity and offset for scheduling request opportunities (so as to reduce potential contention in the use of the PUCCH channel). As a reference, the periodicity may vary from 2 OFDM symbols up to 640 slots in between scheduling request opportunities. As defined in TS 38.331 [14] section 6.3.2 for the SchedulingRequestResourceConfig Information Element.

In an extreme worst case, if the maximum scheduling request periodicity is configured by the network, a mobile device may need to wait up to 640 slots for a scheduling request opportunity. Unfortunately, the scheduling request periodicity is not the only contributing factor to the uplink latency, as the actual transmission and reception of the scheduling request takes time. Additionally, upon reception of the scheduling request, additional time will be needed to transmit the decided uplink grant, that will also happen few slots in advance of the actual PUSCH uplink transmission.

The Buffer Status Reporting, is a mechanism used by the network to understand if after an uplink data transmission, a mobile device keeps data pending of further transmission. When the UE transmits the actual data, it can interleave a related control element to inform the network how much data remains available for future transmissions if additional resources are granted. In scenarios where the amount of available data, and the rate of uplink data arrival for transmission, exceeds the capacity of the allocated uplink resources, it would be possible for the network to keep allocating continuously uplink grants every uplink time slot. However, in practice the nature of the user data traffic is bursty or spare in nature at intervals. In those scenarios, for example when spare packets are received at the start of a TCP connection, or after a gap between traffic bursts, the mobile device may report that no additional traffic is buffered for transmission. In that case, the network may allocate a single uplink grant to the device for a single time slot, what will cause the mobile device to wait for the following scheduling request opportunity before triggering a new uplink data transmission.

When using proactive scheduling, the network can benefit from its authority in terms of scheduling decisions, and assume that a mobile device may need to transmit in the one or more slots after a provisioned uplink grant. In assuming so, the network can provide additional uplink grants for transmission in a sequence of consecutive uplink slots of configurable duration. Of course, the additional allocation of additional uplink resources may cause in some slots the available uplink data not to completely fill the provide resources, or even to have granted uplink slots where no data is available at all. However, this is a valid behavior in terms of radio operation and conformance to the 5G NR, as the mobile device will fill up all the unused payload with padding. Indeed, this is a very frequent scenario in Radio Frequency (RF) conformance testing, as defined in test procedures within TS 38.521-1 [15], where the mobile devices send uplink mac padding since there is no uplink payload.

Depending on the value and priority of the traffic, and the need to reduce latency, we can conclude that proactive scheduling may be an interesting option to consider.

Table 10 shows the results of executing a RTT test case with and without proactive scheduling and also the values of the most representative radio parameters. These radio values are very similar for both scenarios so we can compare the RTT results. The RTT test cases consists of repeating 25 times 2 minutes ICMP ping sessions between the UE (User Equipment) and the Serving Gateway. The statistical analysis of the RTT results reveals a significant decrease of the delay (more than 50%) and a higher stability of the results.

Table 10: RTT 5GNR NSA Sub 6 GHz Proactive Scheduling vs non Proactive Scheduling

Parameter	Indicator	Scenario	
		Proactive Scheduling	Non Proactive Scheduling
RTT [ms]	Average	12.450 +/- 0.066	28.597 +/- 0.241
	Median	12.190 +/- 0.088	28.256 +/- 0.151
	Min	10.020 +/- 0.021	17.916 +/- 1.254
	Max	20.784 +/- 1.626	75.948 +/- 18.589
	5% Percentile	10.237 +/- 0.042	22.638 +/- 0.280
	95% Percentile	15.885 +/- 0.244	33.949 +/- 0.321
	Standard deviation	1.816 +/- 0.081	6.035 +/- 1.349
SINR [dB]	Average	21.4	22.3
RSRP [dBm]	Average	-58	-59
RSRQ [dB]	Average	-10.8	-6.7

3.7 Coverage aspects

3.7.1 Coverage impact to trials

Pre-trials in Greece and Finland have provided the results presented in this subsection [18]. The coverage in this context could means an area where 5G services are available. The coverage is in some places related to available spectrum area, which is, at least in Finland, strictly controlled and there are clear limits. This cause under usage of transmitted RF power, which affect to received 5G services level like throughput etc.

Some trials, for example, must be performed due to trial nature on non-optimal place for coverage point of view, which may also affect to other 5G KPIs [16]. Secondly, the coverage is depended of the spectrum allocation of 5G Frequency Range 1 (FR1) under 6 GHz frequency bands and Frequency Range 2 (RF2) above 24.25 GHz frequency bands as depicted below in Figure 15.

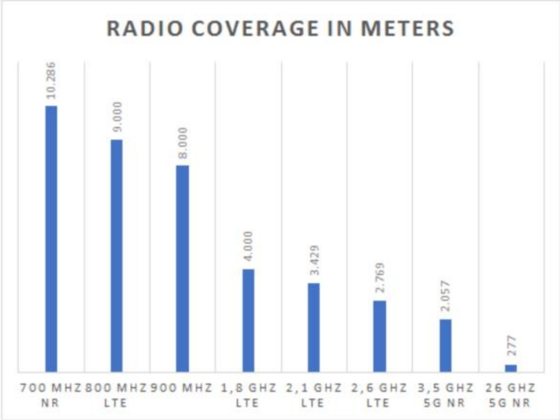

Figure 15: Estimation of radio coverage for several 5G NR frequencies [17]

In the Athens Trials [18] feasibility tests for controlling drone flights over the 5G network. These test experience shows how the coverage should not be considered only on the horizontal level, as with the drones the vertical coverage is equally important. To overcome the coverage limitations in the tests, several manual flights were performed to find the good coverage and the spots where the 5G coverage is lost.

As a conclusion, the network coverage planning is important as closer to the cell edge the trials are performed, the throughput is weaker. Secondly, it makes sense to use some efforts to find the good coverage areas on the trial sites and avoid possible weaker areas.

3.7.2 Mm-wave coverage measurement

In this subsection, a series of measurements that focus on how the environment can influence coverage at mm-wave frequencies are presented. The tests were done using a commercial grade 5G base station with antennas installed on the roof of a building approximately 17 meters above ground. Four carriers with 200 MHz bandwidth each and central frequencies between 26.6 and 27.2 GHz were used. The base station transmitted 16 highly directional beams on each of the four frequencies. The measurements were performed without any traffic in the network.

The measurements were made with a scanner that could monitor all relevant channels using an omnidirectional antenna. The scanner measured the average power of the resource elements that carry the Secondary Synchronization Signal (SSS) transmitted within a Synchronization Signal Block (SSB). Since these are time multiplexed for the different beams, the measurements could be performed without any inter-beam interference.

A detailed description of the tests and the results can be found in [19].

Vegetation will have an important influence on the coverage of mm-wave systems. Figure 16 shows the results of a series of tests with different trees blocking the Line-of-Sight (LoS) path between the BS antenna and the scanner. All measurements were done at the same distance of 220 meters from the BS antennas.

The time series plot in Figure 16 shows the received Reference Signal Received Power (RSRP) as a function of time. The average signal attenuation caused by the trees ranged from 17.7 to 26.4 dB. The time series plot also shows that the trees caused a significant variation of the received signal strength, which is due to the motion of the branches and leaves in the wind.

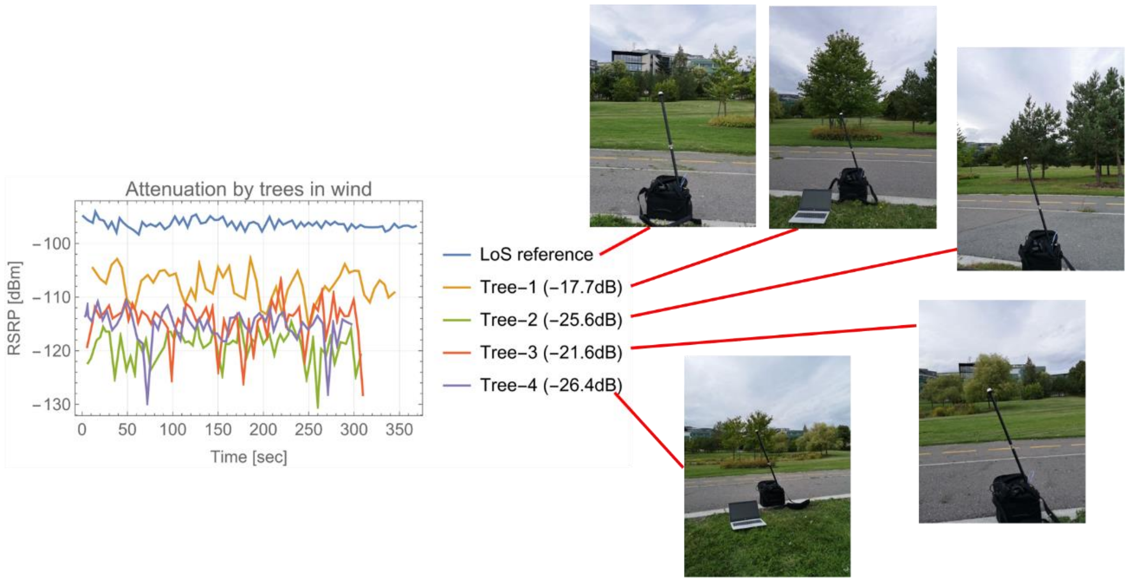

Figure 16: Attenuation of mm-wave signals by trees

Human body attenuation of mm-wave signals is an interesting and important effect influencing the coverage. To measure this the scanner was placed at a location with good line of sight to the BS antenna as shown in the photo in Figure 17. The measurements were done in the following way. First, we withdraw from the scanner for some time to obtain the received RSRP under unblocked LoS condition. Then a person placed himself in front of the scanner blocking the LoS path to the BS antenna. Finally, after first withdrawing from the antenna for some time, a person folded his hands around the top of the scanner antenna blocking signals from all directions.

A time series plot of the received RSRP for this experiment is shown in Figure 17. When the person was blocking the LoS path to the BS antenna, the received signal level was reduced by about 20 dB. However, when the person folded his hands around the top of the scanner antenna, the received signal level fell below the sensitivity of the scanner. This indicates that the human body causes a very large attenuation of mm-wave signals. The residual signal strength measured when the person was blocking the LoS path to the BS antenna is probably from signals reflected from objects in the environment and not from signal going through the person's body.

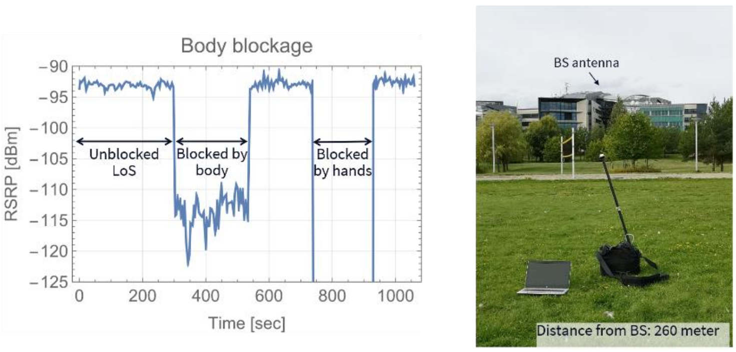

Figure 17: Attenuation of mm-wave signals by human bodies

We also performed some measurements related to transmission of mm-wave signals over water. This is interesting in several use cases. One example is fish farms where high speed communication is needed to transfer sensor and high-resolution videos from fish cages to shore. Test were done both with the scanner located by the water edge and with the scanner located about 6 meters above the water. Figure 18 shows the measurements obtained in the latter case.

The most important observation from these measurements is the great variability of the received signal strength, which can be seen from the time series plot in Figure 18 for the 6 meter above water case. For LoS conditions over ground the signal strength was quite stable as shown for example in Figure 16. The standard deviation of the received RSRP for over ground LoS scenarios was about 0.3 dBm, while for the over water measurements the received RSRP had a standard deviation ranging from 2.3 to 4.7 dBm. This increased signal variability must be considered when planning mm-wave systems in over water scenarios and will result in increased link margin requirements.

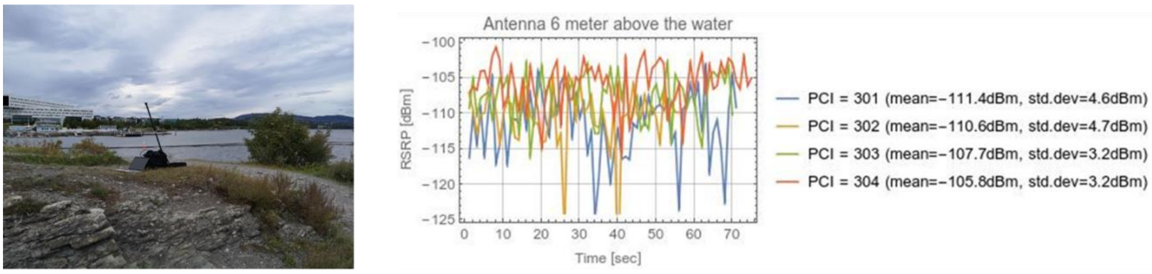

Figure 18: Transmission of mm-wave over water

3.8 Transport layer protocol impact

The traditional internet protocols were designed with a specific underlying network infrastructure in mind. 5G networks offer multi-gigabit speeds and very low latency (in some cases around 1 ms). Thus, the expectation of the mobile link being the bottleneck is no longer true. Further, early measurements reveal that existing transport layer protocols may not be able to fully utilize the potential of 5G networks.

In this work we plan to identify inefficiencies of TCP congestion control algorithms when used over 5G. We evaluate 1) the loss-based TCP CUBIC and 2) TCP BBR, which relies on the estimation of Bottleneck Bandwidth and Round-Trip Propagation Time. We use a UDP transfer from a well-connected server as a baseline, meant to assess the potential speed of the link. Our measurements are performed during August and September 2020. We anonymize the results by referring to the operators as “5G Deployment” 1 to 3.

3.8.1 Set up

Our UE is a laptop connected over gigabit ethernet to a 5G CPE from a large vendor. The achieved speeds are always well below 1 Gbit/s, so the ethernet connection does not limit the measurements. Our dataset consists of file downloads for every combination of network operator and transport layer protocol. Packet capture allows us to monitor the fluctuations of speed during each download. We get speed estimates by dividing time into 1-second buckets and monitoring the number of bytes received during every bucket. Finally, we perform a basic filtering in the buckets to discard estimates that fall into the slow start phase of TCP. Since we want to study protocol behavior, we only use one flow per download. This is in contrast to the typical approach of speedtest-like applications, which launch several parallel flows in an effort to maximize throughput. The relatively low speeds presented in the sequel can be partially attributed to the above.

3.8.2 Transport layer performance

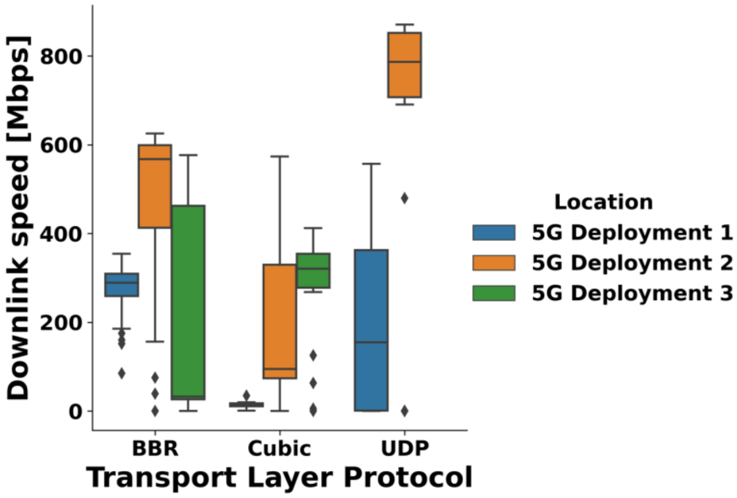

Figure 19: Distributions of the achieved downlink speed for single flow downloads, per transport layer protocol.

As mentioned above, we split each download into 1-second buckets and perform basic filtering to remove artifacts, such as slow start. Each of the remaining 1-second buckets acts as a speed sample. In Figure 19, we present the distributions of the speed samples for the different protocols and locations. Each network is deployed in a different city. Note: We do not have UDP samples for 5G Deployment 3. As expected, UDP achieves the highest possible speed per location. The UDP traffic source generates 1 Gbit/s constant bit rate traffic, but we observe a big variability on the received traffic, which cannot be attributed to protocol behavior. The measurements are static and performed soon after the networks launched 5G, thus signal strength and congestion are unlikely to affect them. The variability may be caused by overflowing buffers and traffic shaping.

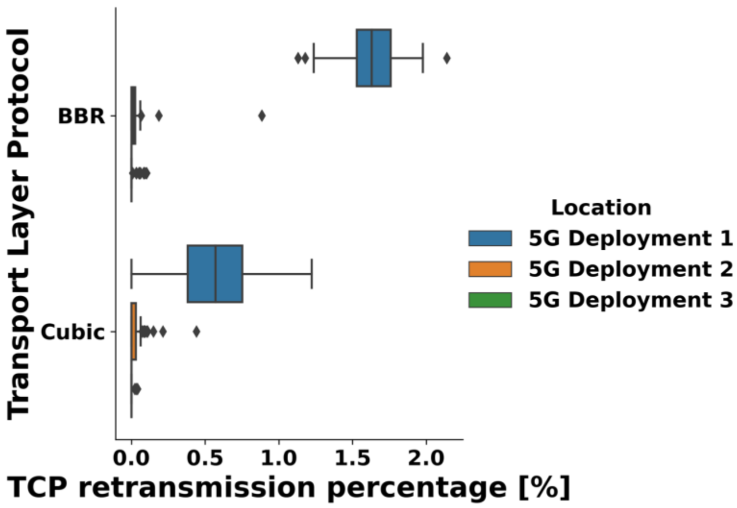

Figure 20: Percentage of packets that are retransmissions within each 1-second bucket.

BBR significantly outperforms Cubic in two of the three locations measured. A possible reason is packet loss. We evaluate packet loss by monitoring the percentage of packets marked as retransmissions within each 1-second bucket. Figure 20 presents the distribution of these samples. The loss-based Cubic is expected to overact to packet losses, whereas BBR is designed to focus on other metrics. 5G Deployment 3 has almost no retransmissions and therefore is the only location where Cubic performs better. On the other hand, if there is even minor packet loss BBR is more performant.

3.8.3 CDN

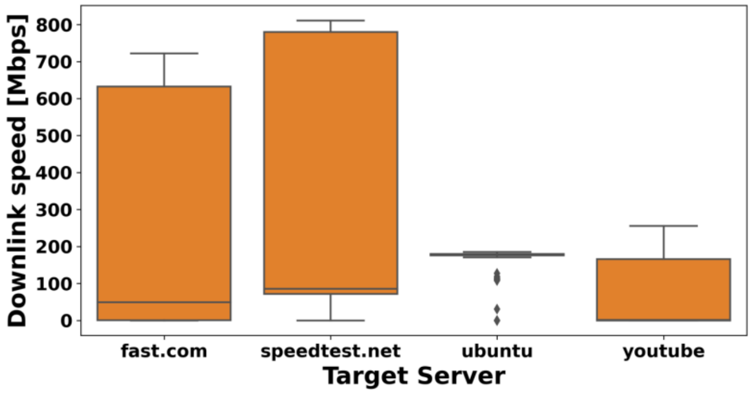

Figure 21: Downlink speed when downloading files from CDNs.

We briefly evaluate how CDNs perform by downloading files hosted in several well-known domains over the 5G network at 5G Deployment 2. Figure 21 presents the distributions of the samples which were created using 1-second buckets as above. In the case of YouTube, we used the youtube-dl utility instead of the website to avoid traffic pacing, which is typical in streaming scenarios. Services targeted towards testing download speed may momentarily max out the connection, but the median values for all the measured CDNs are consistently low. A possible reason could be the use of TCP Cubic, which does not perform very well over 5G Deployment 2 and is the default congestion control algorithm of the Linux kernel.

4 Conclusions

The motivation for this whitepaper was the summarization of the major 5G performance results from related 5G PPP projects, as well as the analysis of the results on the basis of a) identifying the main impacting factors and b) comparing the achieved values against the theoretical ones. In this direction, the collection and analysis of 5G performance results was started from the 5G PPP ICT-17 projects, since they have entered their final phase and they have achieved interesting insights on the actual 5G performance, while they have produced significant outcomes based on their lessons learned.

In general, the performance impact factors can be classified into three main categories: a) The deployment and configuration aspects; b) The scenario under which an experiment / measurement is performed; c) The testing/experimentation procedure, referring to the methodology used for collecting measurements and calculating the KPI values. This paper focus on (a), while factors (b) and (c) are already addressed in a previous TMV whitepaper. In (a) the focus is on two deployment factors, namely the transport network characteristics, and the network core deployment type, as well as on five configuration aspects, namely the bandwidth size, the MIMO layers in RAN, the UL/DL intensive patterns, the scheduling approach, and the target coverage. Based on the aforementioned identification of the major impact factors, several indicative results (for each impact factor) mainly from the three 5G PPP ICT-17 projects are presented and explained, while recommendations are given when possible. The objective is to clarify the details behind the performance numbers and provide a series of interpretation guidelines in trying to explain to the wide audience (e.g., Verticals) what they can really expect from 5G.

Some indicative finding can be summarised below:

- Impact of transport network characteristics: Fiber optic transport network shows negligible latency (mainly dependent on the fiber length) and support for the required throughput without packet loss. Satellite link adds about 600ms RTT latency and supports on average a rate of 25 Mbps.
- Core deployment edge vs central: URLLC-exclusive scenarios require close Edge or near Edge approaches. Mixed URLLC-eMBB scenarios is clearly near Edge model. In eMBB-exclusive scenarios, the safest option to adopt is also near Edge (central office also possible).
- Bandwidth impact: Throughout results in FR1 are close to theoretical values. TCP experiments have lower performance than the UDP due to ACK/retransmissions.
- MIMO layers impact: In NSA mode, DL UDP transmissions with 2x2 MIMO achieves gain of about 50%. In UL UDP case the gain is about 100%. Results from SA is under further analysis.
- UL vs DL intensive patterns: In uplink, the use of NR+LTE traffic aggregation has a big impact in the throughput peak rate (~122% improvement). The downlink case showcases improvement of 23% in NR+LTE scenarios compared to NSA.
- Coverage impact: Vegetation have important influence (~19 to 26 dB signal attenuation). Significant impact when person is blocking the LoS path (~20 dB attenuation). Great variability of the received signal strength (6m) over water (standard deviation ranging from 2.5 to 5 dBm)

As next steps, the TMV WG will extend the current study by collecting and analysing 5G performance results from the ICT-19 projects and all the other 5G PPP projects that are currently running the deployment or execution phases of 5G trials. The plan is to provide additional results on the impact factors and insights by end 2021.

Contributing Projects

Contacts

WG Chair: Evangelos Kosmatos, WINGS ICT Solutions vkosmatos@wings-ict-solutions.eu

WG Co-chair: Michael Dieudonné, Keysight Technologies michael_dieudonne@keysight.com

5G PPP WGs: 5g-ppp.eu/5g-ppp-work-groups

Contributors

Name	Company/Institution/University	Project	Country
	Editors		
Evangelos Kosmatos	WINGS ICT Solutions, TMV Chair	5G-EVE	Greece
Michael Dieudonné	Keysight Technologies, TMV Co-chair	5G-VINNI	Denmark
	Contributors and Reviewers		
Manuel Lorenzo	Ericsson	5G-EVE	Spain
Dimitris Tsolkas	Fogus Innovations & Services P.C.	5GENESIS	Greece
Anastasius Gavras	Eurescom	5G-VINNI	Germany
Andrea F. Cattoni	EANTC	5G-VINNI	Germany
Luca Valcarenghi	Scuola Superiore Sant'Anna	5GROWTH	Italy

Almudena Diaz Zayas	University of Malaga	5GENESIS	Spain
Ole Grøndalen	Telenor	5G-VINNI	Norway
Gokul Srinivasan	Robot Expert	5G!Drones	Finland
Ilkka Kansala	Nokia	5G!Drones	Finland
Lars Mikkelsen	Keysight Technologies	5G-VINNI	Denmark
Drissa Houatra	Orange	5G-VICTORI	France

References

[1] 'Verizon Launches World's First Commercial 5G Smartphone Service', Apr. 2019, https://www.forbes.com/sites/jeanbaptiste/2019/04/04/verizon-launches-worlds-first-commercial-5g-smartphone-service/?sh=43f221201961

[2] NGMN, Definition of the Testing Framework for the NGMN 5G Pre-Commercial Network Trials (Version 2), https://www.ngmn.org/fileadmin/ngmn/content/downloads/Technical/2019/190111_NGMN_PreCommTrials_Framework_definition_v2_small.pdf

[3] 3GPP, TS 28.552: Management and orchestration; 5G performance measurements, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3413

[4] Final Evaluation Report From the 5G Infrastructure Association on IMT-2020 Proposals IMT-2020/14, 15, 16, Parts of 17, document 5D/50-E, 5GIA, Feb. 2020.

[5] M. Fuentes et al., "5G New Radio Evaluation Against IMT-2020 Key Performance Indicators," in IEEE Access, vol. 8, pp. 110880-110896, 2020, doi: 10.1109/ACCESS.2020.3001641.

[6] 5GPPP White paper: Validating 5G Technology Performance Assessing 5G architecture and Application Scenarios, https://5g-ppp.eu/wp-content/uploads/2019/06/TMV-White-Paper-V1.1-25062019.pdf

[7] 5Growth COMAU pilot, https://5growth.eu/wp-content/uploads/2019/06/D4.2-Verification_methodology_and_tool_design.pdf

[8] TR 103 761, https://portal.etsi.org/webapp/WorkProgram/Report_WorkItem.asp?WKI_ID=59575

[9] RFC2544, "Benchmarking Methodology for Network Interconnect Devices", https://tools.ietf.org/html/rfc2544

[10] 5GENESIS project, https://5genesis.eu/

[11] 3GPP TS 38.211 version 15.2.0 Release 15

[12] D6.2 5GENESIS (https://5genesis.eu/wp-content/uploads/2020/08/5GENESIS_D6.2_v1.0_FINAL.pdf)

[13] 3GPP TS 38.321: "NR; Medium Access Control (MAC) protocol specification", https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3194

[14] 3GPP TS 38.331: "NR; Radio Resource Control (RRC); Protocol specification", https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3197

[15] TS 38.521-1: "NR; User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Range 1 standalone",

https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3381
[16] https://5gtn.fi/5g_coverage_3d_visualization/
[17] View on 5G Architecture, 5G-PPP-5G-Architecture-White-Paper_v3.0_PublicConsultation.pdf
[18] https://5gdrones.eu/wp-content/uploads/2020/12/Feasibility-Tests-Egaleo-Stadium-Oct-2020.pdf
[19] Salman Mohebi, Foivos Michelinakis, Ahmed Elmokashfi, Ole Grøndalen, Kashif Mahmood and Andrea Zanella, "Sectors, Beams and Environmental Impact on Commercial 5G mmWave Cell Coverage: an Empirical Study", arXiv:2104.06188 [cs.NI], April 2021
[20] 3GPP, TS 23.501: 5G; System Architecture for the 5G System (Specification # 23.501 (3gpp.org)) https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144
[21] 3GPP SA2 – Architecture (SA2 - Architecture (3gpp.org)) https://www.3gpp.org/specifications-groups/sa-plenary/sa2-architecture
[22] 3GPP Control and User Plane Separation (Control and User Plane Separation of EPC nodes (CUPS) (3gpp.org)) https://www.3gpp.org/cups

Abbreviations and acronyms

3GPP	3rd Generation Partnership Project
5G IA	5G Infrastructure Association
5G NR	5G New Radio
5G PPP	5G Public Private Partnership
AR	Augmented Reality
BBU	Baseband Unit
CAPEX	Capital Expenditures
CC	Component Carriers
CDN	Content Delivery Network
CP	Control Plan
CPE	Customer Premise Equipment
CPRI	Common Public Radio Interface
DL	Downlink
DUT	Device Under Test
DWDM	Dense Wavelength Division Multiplexing
eMBB	enhanced Mobile Broadband
FDD	Frequency Division Duplex
FR	Frequency Range
GEO	Geostationary Orbit
ICMP	Internet Control Message Protocol
ICT	Information and Communications Technology
IMT	International Mobile Telecommunications
ITU	International Telecommunication Union
KPI	Key Performance Indicator
LOS	Line of Sight
LTE	Long-Term Evolution
MAC	Medium Access Control
MEC	Multi-access Edge Computing
MIMO	Multiple-Input and Multiple-Output
mMTC	massive Machine Type Communications
MNO	Mobile Network Operator and
NEM	Network Equipment Manufacturer

NGMN	Next Generation Mobile Networks Alliance
NSA	Non-Stand-Alone
OADM	Optical Add and Drop Multiplexer
OFDM	Orthogonal frequency-division multiplexing
OPEX	Operating Expenses
PDCCH	Physical Downlink Control Channel
PEP	Performance Enhancing Proxy
PRACH	Physical Random-Access Channel
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
R&D	Research and Development
RAN	Radio Access Network
RB	Resource Block
RF	Radio Frequency
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTT	Round Trip Time
RU	Radio Unit
SA	Stand-Alone
SDU	Service Data Unit
SINR	Signal-to-Interference-plus-Noise Ratio
SSB	Synchronization Signal Block
SSS	Secondary Synchronization Signal
SUT	System Under Test
T&M	Testing and Monitoring
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TMV	Test, Measurement, and KPIs Validation
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UP	User Plane
URLLC	Ultra Reliable Low Latency Communications
VR	Virtual Reality