5Genesis

5TH GENERATION END-TO-END NETWORK, EXPERIMENTATION, SYSTEM INTEGRATION, AND SHOWCASING

[H2020 - Grant Agreement No. 815178]

Deliverable D6.3

Trials and experimentation (cycle 3)

Editor	G. Gardikis (SHC)
Contributors	NCSR Demokritos (NCSRD), Fraunhofer FOKUS (FhG), Innovations for High Performance microelectronics (IHP), Humboldt University of Berlin (HUB), Karlstads Universitet (KAU), Simula Research Laboratory (SRL), RUNEL (REL), EURECOM (ECM), Infolysis (INF), Intel (INT), University of Surrey (UNIS), Fogus (FOG), Nemergent Solutions (NEM), University of Malaga (UMA), Cosmote (COS), Universitat Politècnica de València (UPV), Atos (ATOS), Space Hellas (Cyprus) Ltd. (SHC), FON Technology (FON)
Version	1.0
Date	January 31st, 2022
Distribution	PUBLIC (PU)

List of Authors

Fraunhofer	FOKUS-Fraunhofer Gesellschaft e.V., Institute for Open Communication Systems
A. Prakash, F. Eichhorn, O. Keil, M. Emmelmann	
IHP	Innovations for High Performance microelectronics/ *Leibniz-Institut für innovative Mikroelektronik*
J. Gutiérrez, N. Maletic, M. Scheide, M. Ehrig, E. Grass	
KAU	Karlstads Universitet
A. Brunstrom, M. Rajiullah, J. Karlsson, A. Rabitsch, K.-J. Grinnemo	
SRL	Simula Research Laboratory AS
G. Caso, C. Griwodz, Ö. Alay, M. Klausen	
REL	RUNEL NGMT LTD
I. Koffmann	
ECM	EURECOM
P. Matzakos, F. Kaltenberger	
INF	INFOLYSiS
V. Koumaras, C. Sakkas, G. Theodoropoulos, A. Papaioannou, E. Vergis	
INT	INTEL
V. Frascolla, T. Metsch	
UNIS	University of Surrey
S.Vahid, D. Triantafyllopoulou, R. Yogaratnam, F. Carrez, K. Moessner	
FOG	Fogus
D. Tsolkas	
NEM	NEMERGENT SOLUTIONS SL
P. Garrido and E. Atxutegi	
NCSRD	NATIONAL CENTER FOR SCIENTIFIC RESEARCH “DEMOKRITOS”
G. Xilouris, M. Christopoulou, H. Koumaras, T. Sarlas, T. Anagnostopoulos	
UMA	UNIVERSITY OF MALAGA
A. Díaz Zayas, I. González, P. Merino	
COS	COSMOTE KINITES TILEPIKOINONIES AE
I. Mesogiti, F. Setaki, E. Theodoropoulou	
UPV	Universitat Politècnica de València
A. Fornés	
SHC	Space Hellas (Cyprus) Ltd.

G. Gardikis, D. Lioprasitis, P. Ginatzis, I. Mertzanis, T. Prokopidis	
FON	FON TECHNOGY, FONLABS
I. Etxebarria, I. Pretel, E. Saiz	

Disclaimer

The information, documentation and figures available in this deliverable are written by the 5GENESIS Consortium partners under EC co-financing (project H2020-ICT-815178) and do not necessarily reflect the view of the European Commission.

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The reader uses the information at his/her sole risk and liability.

Copyright

The 5GENESIS Consortium consists of:

NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS"	Greece
AIRBUS DS SLC	France
ATHONET SRL	Italy
ATOS SPAIN SA	Spain
AVANTI HYLAS 2 CYPRUS LIMITED	Cyprus
AYUNTAMIENTO DE MALAGA	Spain
COSMOTE KINITES TILEPIKOINONIES AE	Greece
EURECOM	France
FOGUS INNOVATIONS & SERVICES P.C.	Greece
FON TECHNOLOGY SL	Spain
FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	Germany
IHP GMBH – INNOVATIONS FOR HIGH PERFORMANCE MICROELECTRONICS/LEIBNIZ-INSTITUT FUER INNOVATIVE MIKROELEKTRONIK	Germany
INFOLYSIS P.C.	Greece
INSTITUTO DE TELECOMUNICACOES	Portugal
INTEL DEUTSCHLAND GMBH	Germany
KARLSTADS UNIVERSITET	Sweden
L.M. ERICSSON LIMITED	Ireland
MARAN (UK) LIMITED	UK
MUNICIPALITY OF EGALEO	Greece
NEMERGENT SOLUTIONS S.L.	Spain
ONEACCESS	France
PRIMETEL PLC	Cyprus
RUNEL NGMT LTD	Israel
SIMULA RESEARCH LABORATORY AS	Norway
SPACE HELLAS (CYPRUS) LTD	Cyprus
TELEFONICA INVESTIGACION Y DESARROLLO SA	Spain
UNIVERSIDAD DE MALAGA	Spain
UNIVERSITAT POLITECNICA DE VALENCIA	Spain
UNIVERSITY OF SURREY	UK

Version History

Rev. N	Description	Author	Date
1.0	Final version	G. Gardikis (Ed.) (SHC)	31/1/2022

List of Acronyms

Acronym	Meaning
5G	5-th Generation of cellular mobile communications
5G NR	5G New Radio
5G-PPP	5G Public-Private Partnership
AP	Access Point
API	Application Programming Interface
APM	Access Point Manager
CN	Core Network
COAP	Constrained Application Protocol
COTS	Commercial-Off-The-Self
CPE	Customer Premises Equipment
DL	Downlink
DRAN	Distributed Radio Access Network
DRX	Discontinuous Reception
DUT	Device Under Test
DWDM	Dense Wavelength Division Multiplexing
E2E	End-to-End
ELCM	Experiment Life Cycle Manager
eMBB	Enhanced Mobile Broadband - 5G Generic Service
EMS	Element Management System
EPC	Evolved Packet Core
E-UTRAN	Evolved Terrestrial Radio Access Network
FPGA	Field Programmable Gate Array
HEVC	High Efficiency Video Coding
IaaS	Infrastructure as a Service
IED	Intelligent Electronic Devices
KPI	Key Performance Indicator
LCS	Location Service
LMR	Land Mobile Radio
LOS	Line of Sight
LTE	Long Term Evolution
MANO	Management & Orchestration
MCL	Maximum Coupling Loss

MCPTT	Mission critical push-to-talk
MCS	Mission critical services
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
mmWave	Millimetre Wave
MQTT	Message Queuing Telemetry Transport
N3IWF	Non-3GPP Interworking Function
NB-IoT	Narrow Band – Internet of Things
NFVO	Network Function Virtualization Orchestrator
NLOS	Non Line of Sight
NMS	Network Management System
NOC	Network Operations Centre
NSA	Non-Stand-Alone
OAI	Over the Air Integration
OSS	Operational Support Services
PFCP	Packet Forwarding Control Protocol
PLMN	Public Land Mobile Network
PSM	Power Saving Mode
PTMP	Point-to-Multi-Point
QoE	Quality of Experience
QoS	Quality of Service
RAT	Radio Access Technology
REST	Representational State Transfer
RRH	Remote Radio Head
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RTP	Real-time protocol
RTSP	Real-time Streaming Protocol
SA	Stand-Alone
SCT	Service Creation Time
SDK	Software Development Kit
SDN	Software Defined Networking
SINR	Signal to Interference plus Noise Ratio
SUT	System Under Test
TaaS	Testing as a Service
TAP	Test Automation Platform

TTN	The Things Network
TWIF	Trusted WLAN Interworking Function
UL	Uplink
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VNF	Virtual Network Function
VNFM	Virtual Network Function Manager
VPN	Virtual Private Network
VR	Virtual Reality
WIM	WAN Infrastructure Manager
WLAN	Wireless Local Area Network
WP	Work Package
WSAM	WiFi Slice Analytics Monitor
WSC	WiFi Slice Controller
WSMP	WiFi Slice Management Platform

Executive Summary

This deliverable presents the third and final cycle of trials and experimentation activities executed over 5GENESIS facilities. The document is the continuation of deliverables D6.1 and D6.2, in the sense that it captures tests carried out over the evolved infrastructures hosting 5GENESIS facilities following the methodology defined in the previous editions of this deliverable. The tests reported in this document focus on i) the final 5G infrastructure deployments that includes radio and core elements mostly in Stand-Alone (SA) deployment configurations based on commercial and open implementations, and ii) the various use cases/applications, some of them also involving field trials. Most of the tests described herein, especially the generic/lab ones are performed using the Open5GENESIS experimentation suite.

Following the approach of the preceding deliverables, the structure of this document is platform centric, hence it allows each platform to specify independently the group of executed tests and validations performed and the results presented and commented. More specifically, the following experiments were executed during this third cycle, some of them being vertical-agnostic ("generic" tests) and others related to a specific use case:

- Athens Platform - includes tests for throughput, RTT and area traffic capacity, as well as field trials and measurements for the use cases "Big Event", "Eye in the Sky", "Security as a Service"
- Málaga Platform - includes tests for throughput, RTT, area traffic capacity, location accuracy, reliability and speed KPIs, as well as field trials and measurements for the use cases "Wireless Video in Large Scale Event", "Multimedia Mission Critical Services" and "Edge-based Mission Critical Services"
- Limassol Platform - includes tests for throughput and RTT KPIs in various satellite/5G configurations, as well as field trials and measurements for the use cases "5G Maritime Communications" and "5G Rural applications"
- Surrey Platform - includes field trials and measurements for the use cases "Multi-RAT Support for Sensor Measurements", "Coverage Evaluation", "5G Wi-fi slicing (WSMP)", "CoAP over LTE/5G" and "APEX Integration".
- Berlin Platform – includes tests for throughput and RTT KPIs, as well as field trials and measurements on the "360^0 camera" scenario, along with the evaluation of 5G SA equipment at the FOKUS facility.

The main part of each platform section contains an overview presentation of the validated KPIs and measured metrics followed by commentary. Furthermore, the detailed test cases and result tables are available in the Annex of this document. It should be noted that the definition of the Test Cases are delivered as a separate Testing and Validation companion document[1]. This document includes all the test cases templated (i.e. the KPI measured, the System Under Test (SUT) definition, the measurement process and tools) that have been used throughout D6.3

[1] https://github.com/5genesis/5genesis_test_cases/blob/master/Experimenter%20Companion/5GENESIS_Test_Cases_Companion_v.2.0.pdf

Overall, deliverable D6.3 concludes the experimentation campaign of 5GENESIS, covering a wide range of KPIs and evaluating the benefits of 5G not only within the lab, but also in field trials for relevant vertical applications.

Table of Contents

Table of Figures

1. Introduction

This deliverable reflects the final phase of the work carried out within WP6, towards evaluating and validating 5G equipment and network deployments, by means of specific KPIs. The work in the field of KPI validation and performance evaluation is shared among the platforms, which act complementarily. A set of baseline KPIs (i.e. latency, throughput, service deployment time) is common to all platforms; yet, the diverse nature of the testbeds and the variety of capabilities and configurations makes it meaningful to measure the same KPIs across platforms, to show how different network configurations and capabilities affect these KPIs. In addition, each platform hosts a set of use cases/vertical applications which are selected to better highlight the value of its specific capabilities.

This deliverable describes the trials and experimentation results from the third and final testing cycle of 5GENESIS (M34-M42). Compared to the previous editions (D6.1, D6.2), this one embraces new testbed configurations (focusing on 5G SA setups), KPIs not previously covered, as well as a wide variety of vertical use cases.

1.1. Purpose of the document

The results presented in this deliverable are obtained from the experimentation procedures that were conducted over the five 5GENESIS facilities where the Open 5GENESIS suite was integrated. In this context, the related 5GENESIS deliverables are presented in Table 1-1.

Table 1-1 Related 5GENESIS Deliverables

id	Document title	Relevance
D2.1 [1]	Requirements of the Facility	The document sets the ground for the first set of requirements related to supported features at the testbed for the facilitation of the Use Cases.
D2.2 [2]	5GENESIS Overall Facility Design and Specifications	The 5GENESIS facility architecture is defined in this document. The list of functional components to be deployed in each testbed is defined.
WP4 Del.	Athens D4.3[3] Malaga D4.6 [4] Limassol D4.9 [5] Surrey D4.12[6] Berlin D4.15 [7]	These documents describe the platform setup, and capabilities after the end of their third integration cycle (Release C)
D6.1	Trials and Experimentation (cycle 1) [8]	This document presents the methodology and test results performed on the first release of the 5GENESIS platforms.

D6.2	Trials and Experimentation (cycle 2) [8]	This document presents the results from the phase 2 evolution at the testbeds and coordination layer framework. The results were obtained using the initial (D6.1) methodology and updated test descriptions and measurement procedures.

1.2. Structure of the document

The document is devoted to the presentation of the experimental results obtained in the third phase of 5GENESIS project, updating and/or complementing the results of D6.1 and D6.2. The first part of the document (main document) is devoted to experiments and trials that were conducted in each 5GENESIS platform; Chapters 3, 4, 5, 6 and 7 include the experiments in Athens, Malaga, Limassol, Surrey and Berlin platforms respectively. The Annexed part of the document is devoted to the detailed testing procedures and received results from each platform, as well as data anonymisation procedures applied to the Berlin platform experiments. Finally, the document is accompanied with an additional test companion (provided as a separate document) containing all the 5GENESIS Test Cases used for the presented experimental results.

1.3. Target Audience

The primary target audience of this third and final test report encompasses industrial and scientific stakeholders, exposing to them the procedure and results of 5G KPI validation.

More specifically, stakeholders that can benefit from the document include:

- Contributors to standardisation organisations
  Where the test cases can form the basis of test procedures.
- European Commission
  To evaluate the conduction and results of 5G experimentation.
- Industrial players
  To study the result of benchmarking involving COTS products.
- Academic and research stakeholders
  As basis for design decisions for 5G based frameworks and applications development.
- Non-experts interested in 5G opportunities
  To understand the capabilities and limitations of 5G technology.

2. METRICS AND TEST CASES

Most of the tests described in this document follow well-defined test cases, each of which addresses one or more KPIs. The definition of the test cases are included in the companion document "5GENESIS TEST CASES v.2.0"[2] and are presented in Table 2-1. It should be noted that this table also contains all test cases defined. The ones used for the tests of the present deliverable are marked with an asterisk (*)

Table 2-1 Test Case and KPI mapping

KPI	Test Case IDs
	GENERIC KPIs
Capacity	TC_CAP_AreaTrafficCapacity
Density	TC_DEN_MaxRegisteredUE_BER_001 TC_DEN_MaxActiveUE_BER TC_DEN_MaxNumOpReqProcessed_BER TC_DEN_OperProcessingDelay_BER TC_DEN_MaxRegisteredUE_BER_002
Energy Efficiency	TC_ENE_RANEnergyEfficiencyAVG TC_ENE_RANEnergyEfficiencyMAX TC_ENE_UEEnergyEfficiency TC_ENE_NBIoT_SUR
Latency	TC_LAT_e2eAppLayerLatency TC_LAT_PHYLatency_MAL TC_LAT_SmartGridControlMsgLatency_BER TC_LAT_APPLayerLatency TC_RTT_COAP_SUR (*)
Round Trip Time	TC_RTT_e2e (*) TC_RTT_e2eBGTraffic (*) TC_RTT_e2eRadioLinkQuality
Service Creation Time	TC_SCT_VMDeploymen_BER TC_SCT_5GConnSliceInstantiation TC_WiFi_SCT_e2e (*)
Throughput	TC_THR_Tcp (*) TC_THR_Udp TC_WiFi_Th_Rel_DoU
Ubiquity/Coverage	TC_UBI_RANCoverage TC_UBI_BHCoverage TC_UBI_NBIoTRAN TC_COVERAGE_DL_SURREY (*)
Location Accuracy	TC_Loc_Acc (*)

2 https://github.com/5genesis/5genesis_test_cases/blob/master/Experimenter%20Companion/5GENESIS_Test_Cases_Companion_v.2.0.pdf

Reliability	TC_Rel_e2e (*) TC_Rel_Thr_e2e (*)
	APPLICATION – SPECIFIC KPIs
Video Jitter	TC_JIT_VideoStreamJitter_MAL
IoT Application Latency	TC_IoT_PacketDelayHTTPPOST_SUR TC_IoT_PacketDelayMQTT_SUR_001 TC_IoT_PacketDelayCoAP_SUR (*) TC_IoT_PacketDelayMQTToverLORA_SUR TC_IoT_PacketDelay_WIFI_SUR (*) TC_IoT_PacketDelay_5G_SUR TC_IoT_PacketDelay_WIFI/5G_2SLICES_SUR (*)
Video QoE	TC_360LiveVideoStreamingQoE_BER (*) TC_360VideoStreamingQoE_Scalability (*)
MCPTT	TC_MCPTTAccessTime_MAL (*) TC_MCPTTAccessTimeIncCallEstablishment_MAL (*) TC_MCPTTMouthtoEarDelay
APEX integration CPU usage	TC_APEX_SURREY

3. Athens Platform Experiments

3.1. Overview

During the 3rd experimentation phase of the 5GENESIS Project, the Athens Platform focused on conducting experiments on its commercial 5G SA system, which is based on Amarisoft Callbox Classic (5G Core Rel.16 & RAN). The 5G RAN configuration parameters supported are presented in Table 3-1.

Table 3-1 Amarisoft SA RAN Configuration

Band	n78
Mode	TDD
Bandwidth	50 MHz
Carrier components	1 Carrier
MIMO layers	2 layers
DL MIMO mode	2x2
Max Modulation	256QAM
Beams	Single beam
Subcarrier spacing	30 kHz
TDD Config 1	5ms period, 7 DL Slots, 2 UL Slots, 1 Flexible Slot
TDD Config 2	2.5ms period 3 DL Slots, 1 UL Slots, 1 Flexible Slot

The KPIs that were evaluated in the final experimentation phase are listed in Table 3-2.

Table 3-2 KPIs evaluated in the Athens Platform

KPI to be evaluated at the Athens Platform according to DoA	Evaluated in Phase	Comment
Throughput	Phase 1, Phase 2, Phase 3	Based on iperf Measured on generic tests and UCs
RTT	Phase 1, Phase 2, Phase 3	Based on ping Measured on generic tests and UCs
Latency (One Way Delay)	Phase 2	Based on IxChariot traffic generator[3]

[3] https://www.keysight.com/zz/en/products/network-test/performance-monitoring/ixchariot.html

Area Traffic Capacity	Phase 3	Based on throughput
Service Creation Time	Phase 2	No changes in measurements since Phase 2
Additional 5G KPIs evaluated at the Athens Platform		
360° video streaming bit rate	Phase 3	Based on 360° video camera software
Video framedrop occurencies	Phase 3	

The set of experiments conducted are divided in two categories: generic tests and tests conducted during UC events.

3.2. Generic tests – Measurements and results

The generic tests that were conducted in Athens platform at Phase 3 of experimentation, include the following scenarios:

- Maximum achieved throughput measurements using iPerf as underlying tool
- E2E RTT measurements using ICMP traffic (ping)
- Area Traffic Capacity, based on throughput measurements

All baseline experiment measurements were received using the Open5Genesis experimentation framework for their definition and execution (i.e. the number of tests, the statistical analysis, the experiment execution, the probes used). Statistical analysis and visualization of the retrieved results was performed using the Open5Genesis Analytics Framework. The network topology that was used as the experimentation setup in Athens Platform during the gathering of measurements, is illustrated at Figure 3-1 Amarisoft 5GC & RAN testbed setup

.

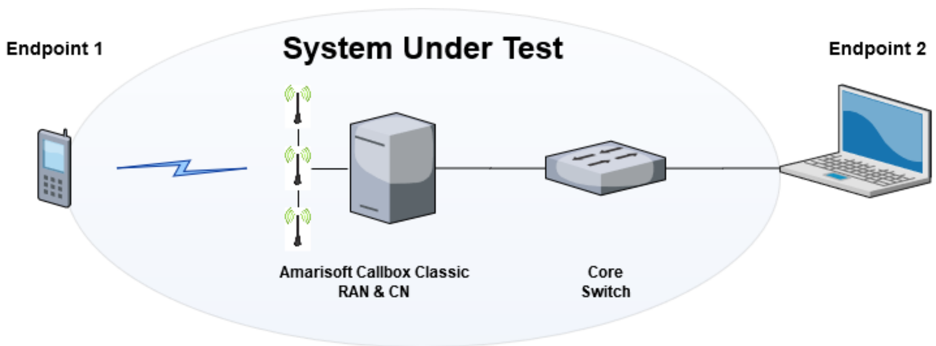

Figure 3-1 Amarisoft 5GC & RAN testbed setup

3.2.1. Throughput (SA)

In this section, the results of the throughput experiments are presented. The experiments were executed based on the test cases TC_THR_TCP and TC_THR_UDP and conducted in lab environments. The results gathered, include mean values of 293+/- 7.4 Mbps for TCP, and 314.42 +/-2.56 Mbps for UDP. The box plots presenting the TCP and UDP throughput results produced by the analytics framework, are shown in Figure 3-2 and Figure 3-3 respectively.

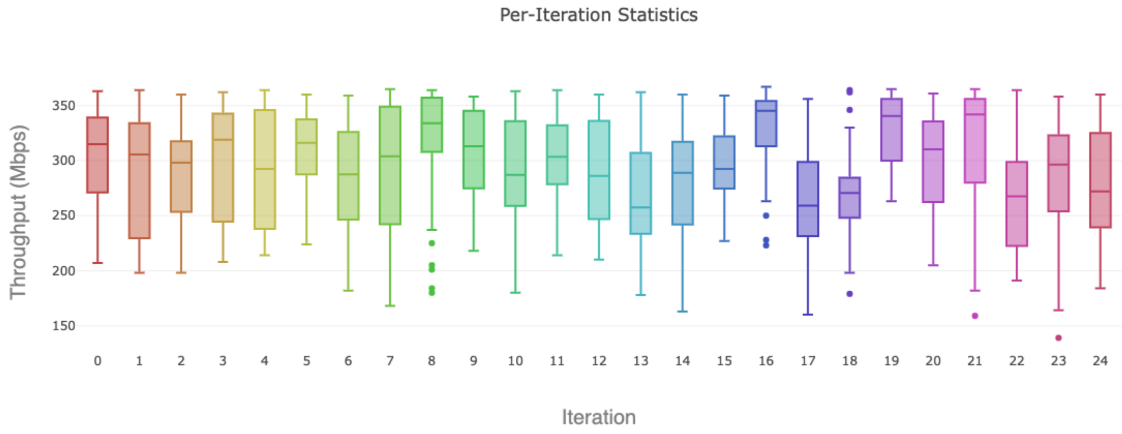

Figure 3-2 TCP Throughput - UE DL

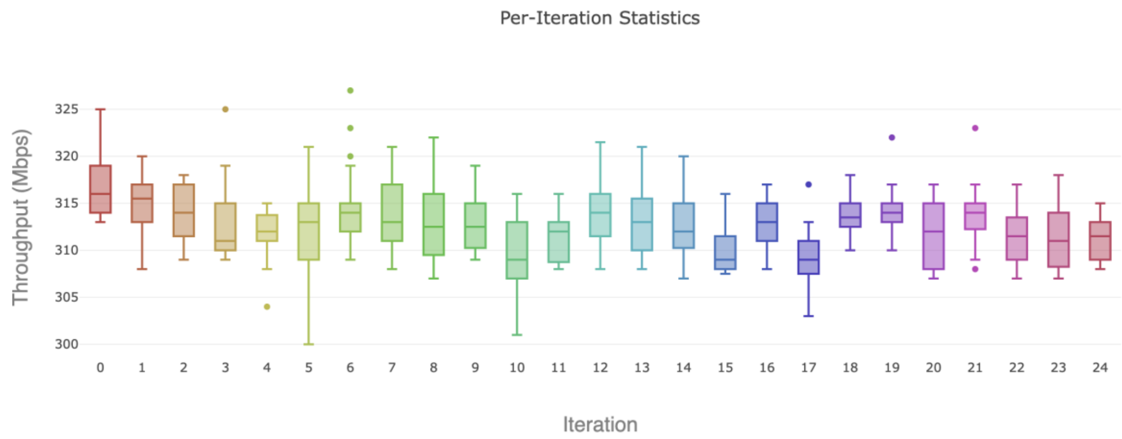

Figure 3-3 UDP Throughput - UE DL

It is worth mentioning that the max value reached via TCP is 361.88 +/- 1.14 Mbps, and the experiments were conducted in ideal radio conditions, with the UE maintaining a high downlink MCS value of 27. The radio equipment was configured to use the TDD configuration 1 from Table 3-1.

3.2.2. E2E Round Trip Time (SA)

This section contains the analysis of the round trip time (RTT) experiments that were performed in Athens Platform. All experiments were conducted using the Open5Genesis suite, and the tool that was utilized by the probes for receiving the measurements was ping.

For this KPI evaluation, the experiments included the creation of 2 configurations:

- The default RAN configuration characteristics (TDD config 1)
- A modified RAN configuration (TDD config 2) optimized for low latency, that prioritizes traffic and reduces idle time for the UE, when requesting uplink grants for the base station, to send uplink traffic.

Detailed information about the composition of reported TDD configurations, can be found in Table 3-1.

For the above configurations, all measurements were received according to the 5GENESIS experimentation methodology for a variety of ICMP traffic packet sizes, ranging from 32Bytes, to 512Bytes, which are indicative packet sizes for URLLC slices type. The direction of the traffic was kept the same for all experiments, with ICMP requests sent from the Dell Laptop (Endpoint 2) to the UE (Endpoint 1).

In this section, only the results of the 64Byte traffic are presented for both RAN configurations using the box plots generated by Analytics framework. The total of the results that were produced from the statistical analysis for all traffic profiles can be found in the Annex.

Figure 3-4 presents the RTT values for 64bytes packets, using configuration 1. The average measures RTT is 29.09 +/- 0.26 ms.

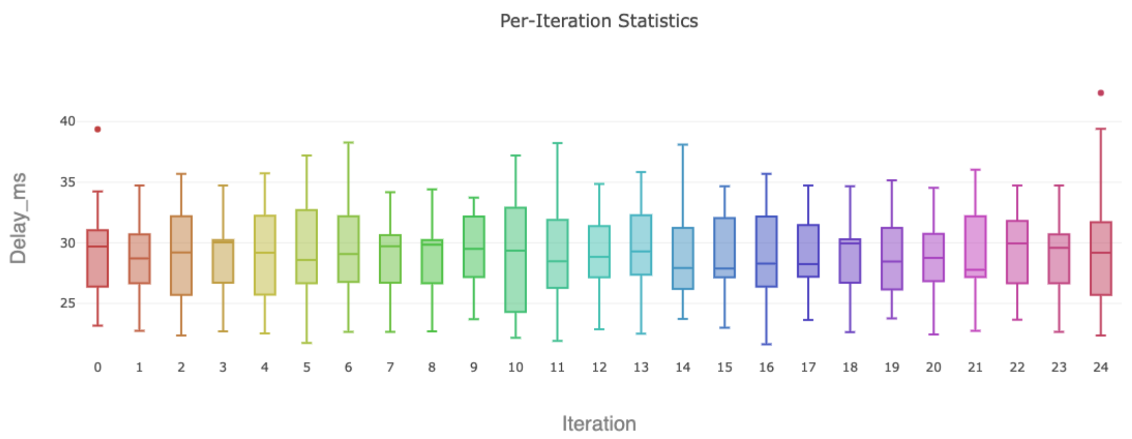

Figure 3-4 E2E RTT, 64Byte packets – RAN Configuration 1

Figure 3-5 presents the RTT values received for 64bytes packet using configuration 2. The average measured RTT values is 12.36 +/- 0.09 ms.

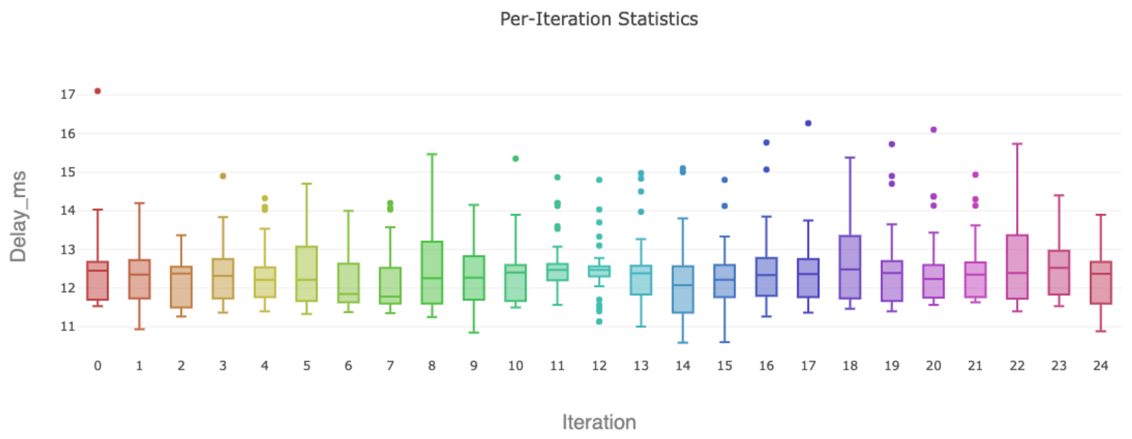

Figure 3-5 E2E RTT, 64Byte packets – RAN Configuration 2 (Low Latency)

An RTT improvement ranging from 46% to 58% (depending on packet size) when using configuration 2 (low latency radio configuration) can be achieved compared to the configuration 1. The Figure 3-6 summarizes all the RTT measurements with all configurations and packet sizes. The results show that using shorter, more frequent changing TDD patterns, and additionally reducing the interval of the scheduling request that the UE needs to wait before requesting an uplink grant, can provide significant latency gains.

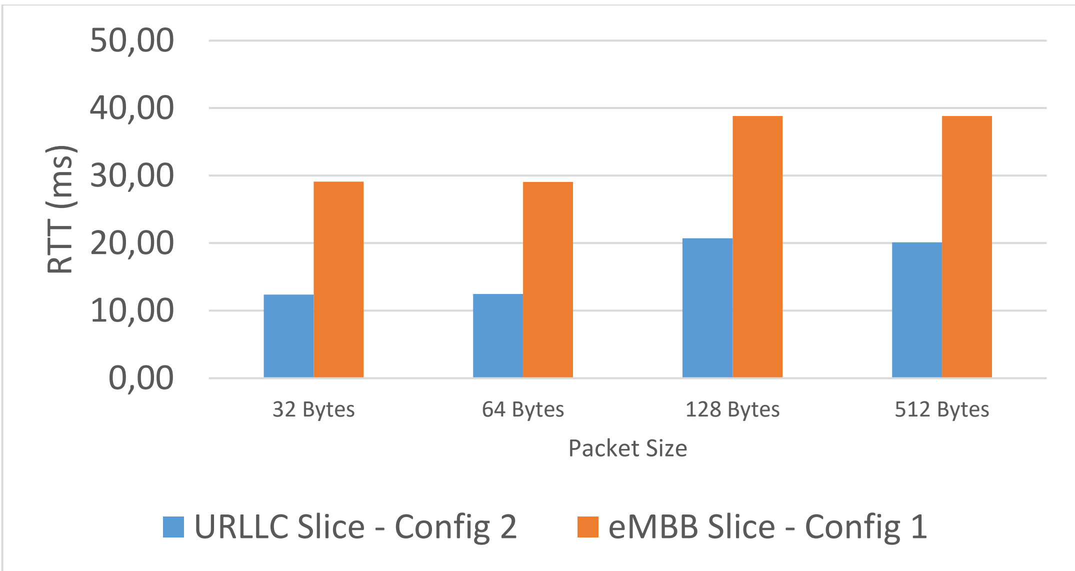

Figure 3-6 Average E2E RTT values achieved

3.2.3. Area Traffic Capacity

Area traffic capacity is the throughput provided over a specific geographical area, expressed in Mbps/m^2. This metric is tightly related to network planning and dimensioning in large scale deployments, but the underlying network technology capabilities can affect the maximum values that can be achieved. We have measured this KPI in one of the 5G SA deployments in Athens Platform including one cell, as an indicative example of an experimental deployment over an area of 40m^2 (5mx8m). Therefore, we set the value of TRxPs=1 (transmission and reception point - TRxP). There is no use of carrier aggregation in our case. Based on the average measurable UDP throughput within this area (using "full buffer" aspects), we calculated the estimated area traffic capacity at 13,9 Mbps/m^2, which is above the minimum requirement defined by ITU of 10 Mbps/m^2 for indoor hotspot-eMBB test environments. Table 3-3 presents the steps for calculating the estimated average area capacity.

Table 3-3 Area traffic capacity calculations

Parameters	Formula
Effective Bandwidth (Hz)	$BW_{eff} = 50\text{ MHz} \times \frac{7\text{ DL Slots}}{10\text{ Total Slots}} = 35\text{ ef. MHz}$
Average Aggregate Throughput (Mbps)	$389\ Mbps$
Average Spectral Efficiency(bit/s/Hz/TRxP)	$SE_{avg} = \frac{389\ Mbps}{35\text{ MHz} \times 12\ TRxPs} = 11.11\ (\frac{bps}{Hz}/TRxP)$
Area (m^2)	$\text{Area} = 5 \times 8 = 40\text{ m}^2$
Site density (TRxP/m^2)	$\rho = \frac{12}{40\ sq.meter}$ = 0.025 TRxP/m^2
Estimated average area traffic capacity (Mbps/m^2)	$C_{avg} = \rho \times W \times SE\_avg = 0.025 \times 50 \times 11.11 = 13.9\text{ Mbps/m}^2$

3.3. UC#1: Big Event – Measurements and results

The Big Event use case showcases the coverage of a big event i.e., football match using 5G infrastructure deployed at a stadium illustrated in Figure 3-7.

Figure 3-7 Egaleo stadium Use Case deployment

The 5G infrastructure at the stadium encompasses a full 5GCore deployment and Edge Computing infrastructure. The stadium edge is connected to the NCSRD premises where additional computing resources are available as well as 5G Core for testing the latency impact when the core runs at the central office and there is no local breakout at the edge location. The big event is covered via a 360 camera that is connected via 5G link locally at the stadium. The 360-camera video stream is sent to a local video cache instance (VNF) in order to be available to the local to the stadium End Users as well as to remote content consumers. The overall use case deployment is illustrated in Figure 3-8. Alternatively, the video stream is sent to the Video Cache (VNF) instance running at the central office (NCSRD Premises).

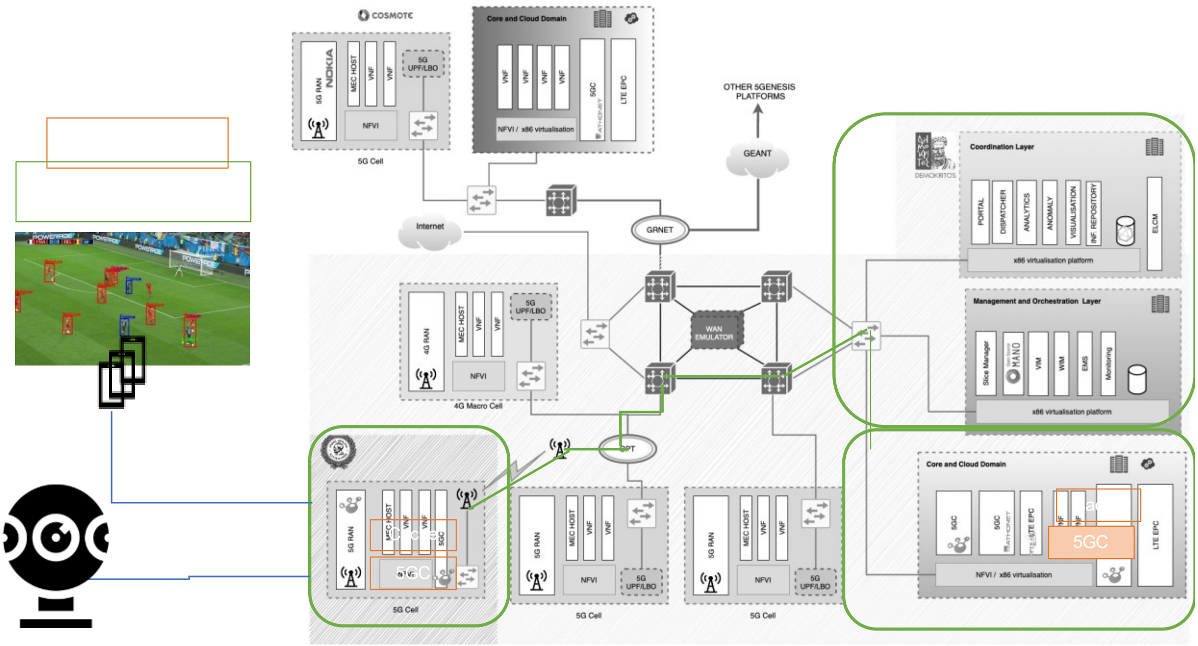

Figure 3-8 Core & Edge Slices Created, Slice Manager CLI

For this use case deployment, 2 cloud infrastructures were deployed:

- Core Cloud located at NCSRD datacenter and was based on OpenStack Rocky
- Edge Cloud located at Egaleo Stadium and was based on OpenStack Wallaby

Two slices were created, containing the required VNFs for the Video streaming service to run (i.e. Caching VNF), and were deployed to each location during the trial. All measurements were gathered with 5 second intervals for a test duration of approximately 14 minutes.

The 360º camera streamed video content in 8K resolution at 30 frames per second. For the duration of the video stream the mean values of the bitrate are presented in Figure 3-9. It can be observed from the box plots that when Core deployment is used, more outliers are observed. In addition, the achieved average received video bitrate at the UE side is the same.

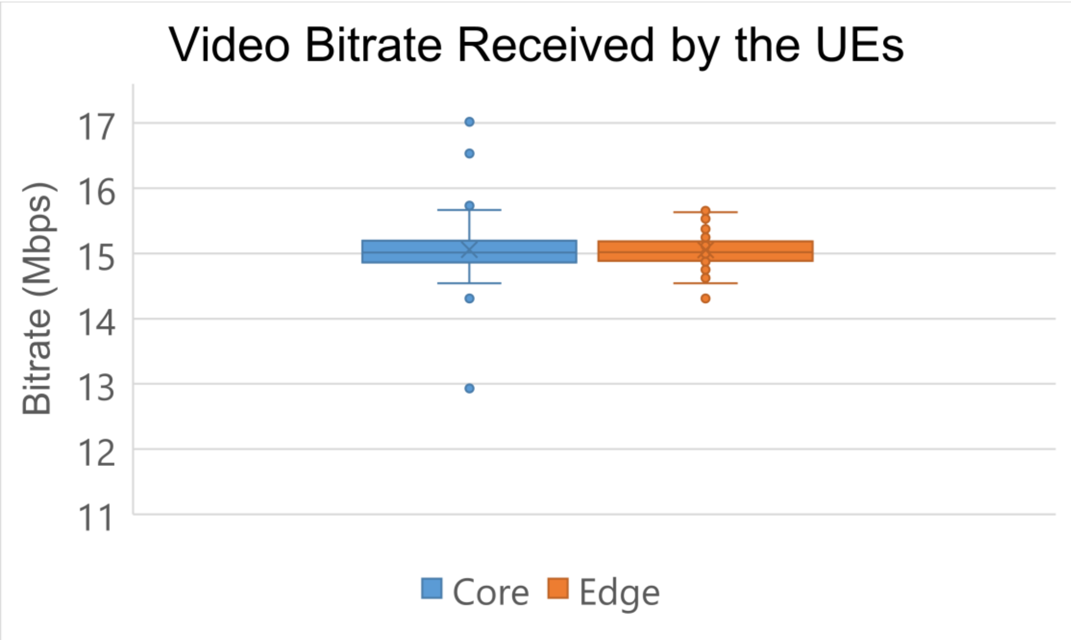

Figure 3-9 Measured Video Bitrate

At the same time using the timestamps of the video stream, latency measurements were also acquired. The UE is always connected at the Egaleo stadium side and consumes the stream either from the Core or the locally at the edge. The measured RTT for the link between Egaleo-NCSRD is approximately 3ms. As it can be observed from Figure 3-10, the average RTT is approx. 1.1sec. This includes the latencies imposed by the processing at the clouds which seems to be dominant. It is anticipated that the impact of additional traffic over the backhaul link would further affect the total latency and thus would more clearly differentiate the Cloud and Edge deployments with respect to latency.

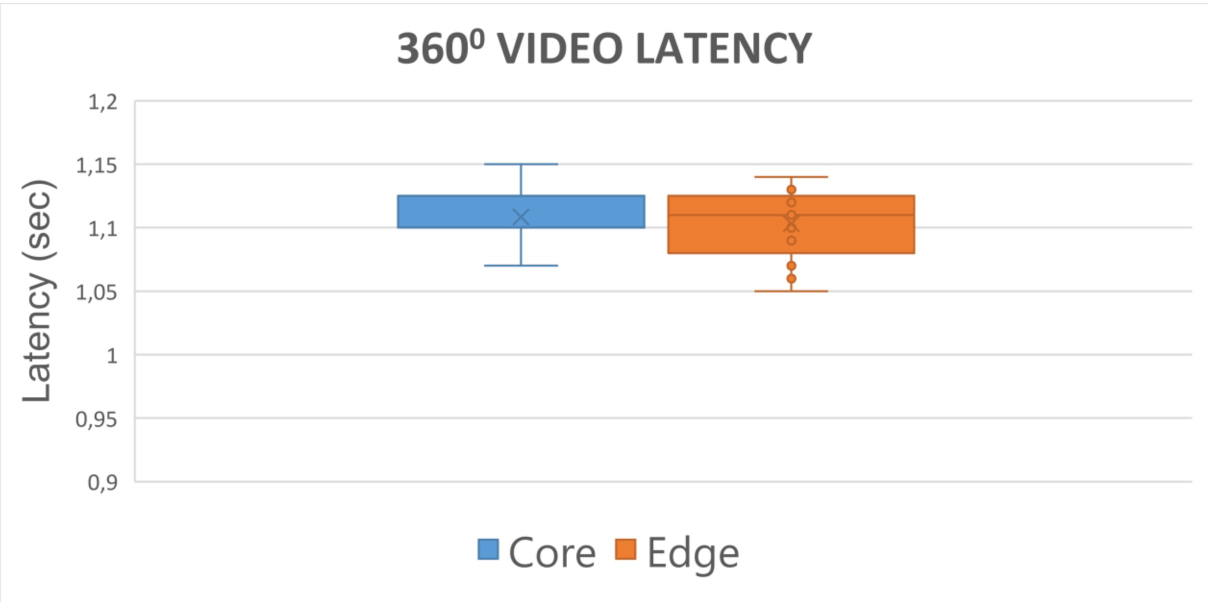

Figure 3-10 Video Latency Core & Edge Cloud Deployment

Video performance difference between the core and edge slice deployment was clearer when evaluating video frame drops. When using the edge deployment, frame drops count was measured at 15 times, all of which had a value of less than 2%, and they were barely observed by the content consumer. However, when using the core deployment, frame drops were measured 26 times, reaching 13% as maximum value. and are illustrated in Figure 3-11 and Figure 3-12

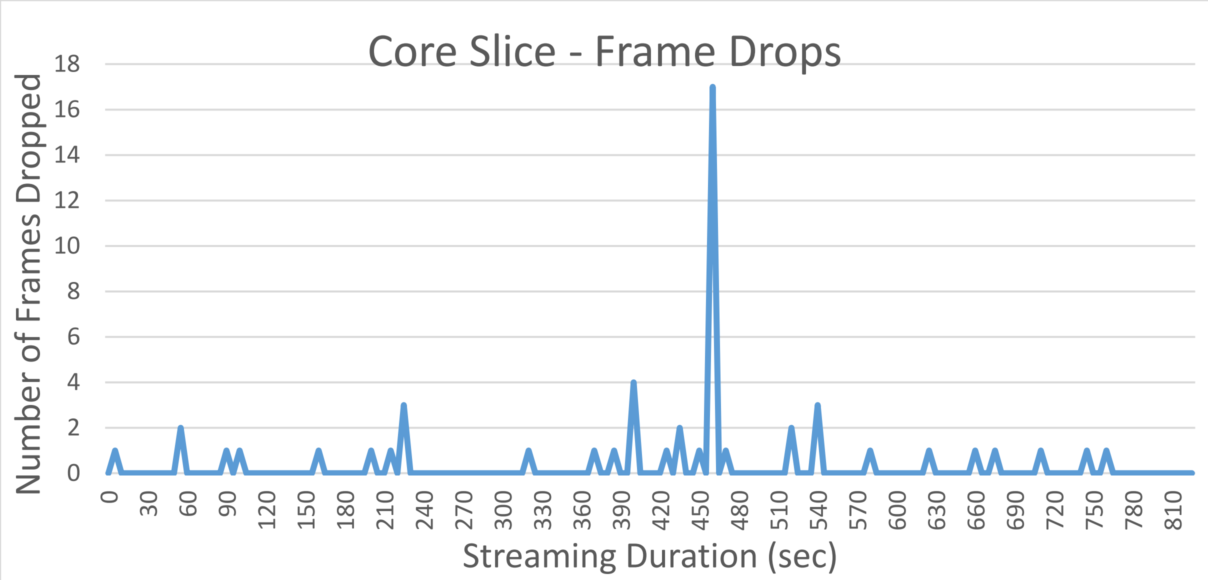

Figure 3-11 Frames Dropped - Core Slice

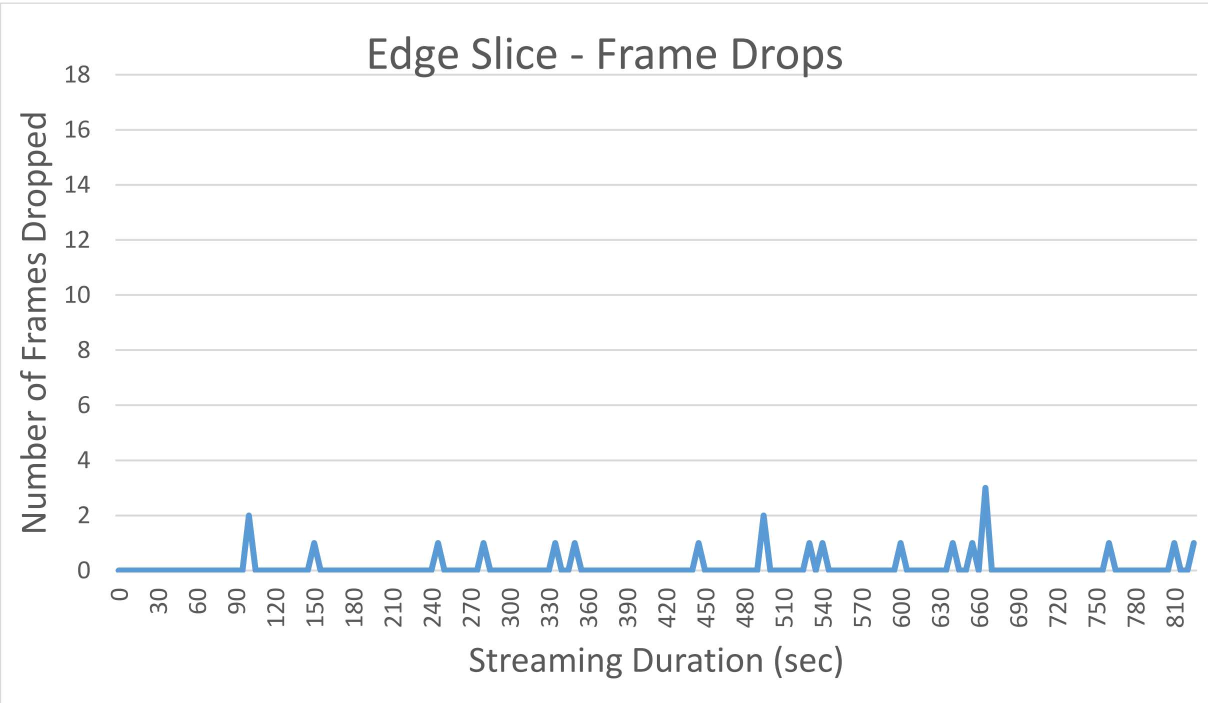

Figure 3-12 Frames Dropped - Edge Slice

3.4. UC#2: Drone Surveillance "Eye in the Sky" – Measurements and results

Considering the variety of advances that 5G offers to the UAV business, this use-case focuses on the demonstration of 5G MEC advantages that allow a UAV to offload the resource-demanding video processing/streaming at the edge of the network. On one hand, this provides energy efficiency to the UAV mission and on the other hand it enables sophisticated and resource demanding services to become possible for BVLOS missions.

The topology of the showcasing event is presented in Figure 3-13 where both sites of 5GENESIS Athens platform were used, namely NCSRD and COSMOTE, which further upgraded the initial plan of the DoA, where this showcasing event was planned to take place at NCSRD site only. More specifically, the RAN and MEC part of the infrastructure were deployed at COSMOTE site, based on NOKIA Airscale RAN and ATHONET's Local Break-Out (LBO) solutions, while the core part based on ATHONET Rel. 15 was hosted at NCSRD site. Therefore, this showcasing event was performed at a multi-domain infrastructure deployment. In the MEC implementation at COSMOTE two software components (in the form of VNFs) were deployed:

- the Ground Control software, which was used for delivering the C2 over 5G and controlling the drone using the 5G access network, and
- the Video Caching VNF for receiving the video-stream from the camera located on the drone with low latency over the LBO.

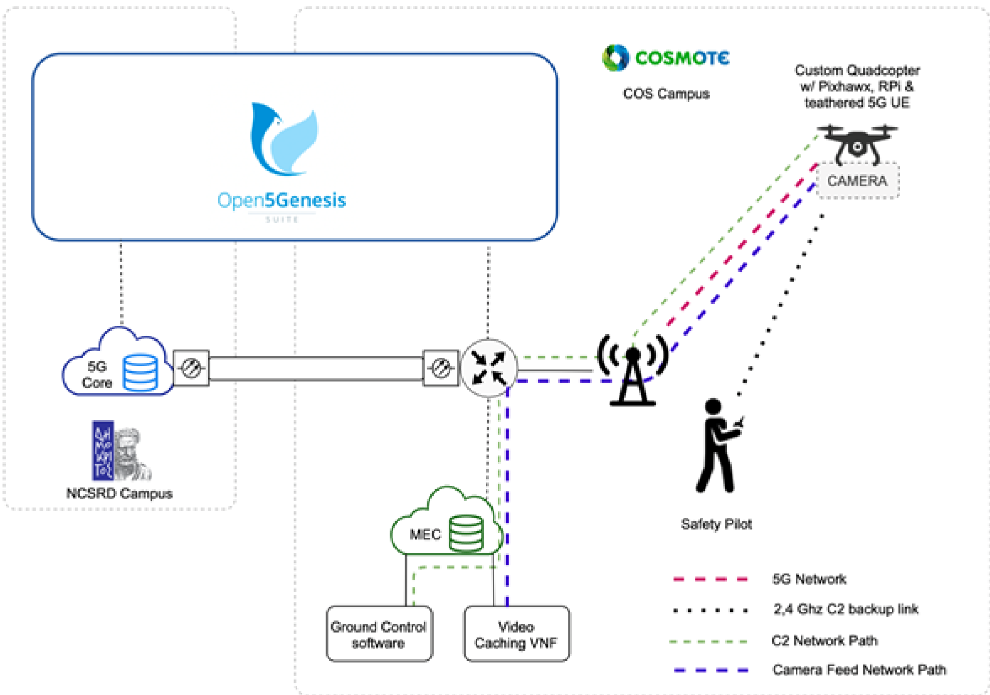

Figure 3-13 Drone trial topology

To facilitate the Video Caching VNFs needed for the use case, two cloud installations were used as depicted in Figure 3-13:

- a Core cloud deployed at NCSRD site based on OpenStack Wallaby, routing the control-plane via the optical transport network between NCSRD and COSMOTE.

- an EDGE cloud deployed at COSMOTE site based on OpenStack Ussuri, connected with the LBO gateway, to minimize transport network latency and complete the MEC setup.

The showcasing event took place at the COSMOTE site, on the 2nd and 3rd of July 2021. The aim had been to demonstrate the drone flight over 5G command and control (C2) and exhibit an eye-in-the-sky service using the LBO/edge of the COSMOTE facilities, testing video streaming from UAV to VNF hosted at the edge.

The conventional communication between the ground pilot and the UAV is realised with a direct link from the user to the UAV as depicted in Figure 3-14 (a). The pilot controls the UAV either moving the sticks of a remote controller or sending commands through a specified software (e.g., Mission Planner, QGroundControl etc.).

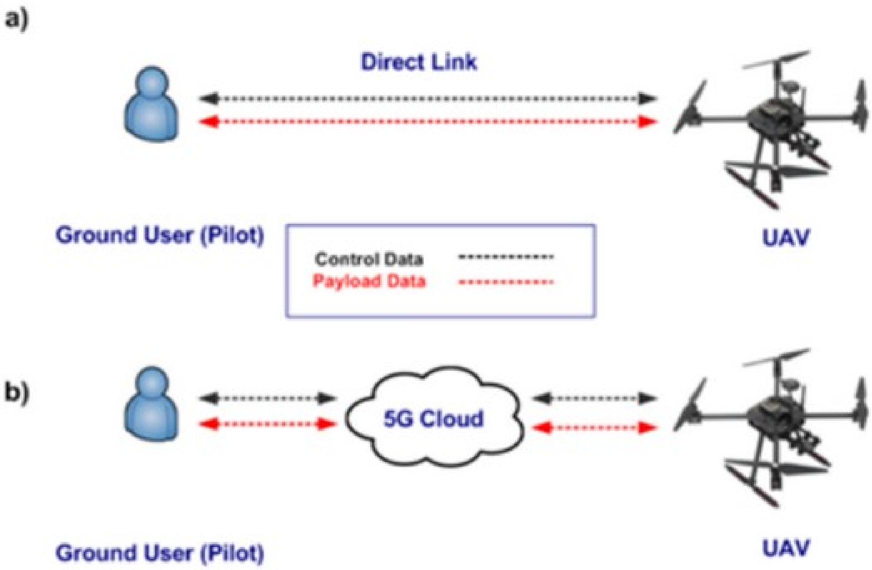

Figure 3-14 C2 link over 5G evolution

In both cases the channel performance is limited for serving demanding applications like UHD video streaming, aerial data relaying, long-distance real-time flight etc. Moreover, computational demanding applications are hosted on-board, resulting in intensive energy consumption that minimizes the flight time and the operational capacity of the UAV for long-distant missions or for sophisticated (e.g., AI-driven) image/video analytics.

As part of this demonstration, a 5G-enabled UAV prototype has been constructed which (i) realises a 5G communication link between the pilot and the UAV for both the control and the payload data, and (ii) offloads the softwarised flight controller and the streaming server at the edge of the 5G network, as shown in Figure 3-14 (b). The pilot/user and the UAV, both connected to the 5G system, exploit the 5G channel high-capacity capabilities and the low latency characteristics of the edge computing. Therefore, this setup and the proof-of-concept 5G-enabled UAV can support a wide range of new services and applications, while at the same time functionalities of the flight controller and the streaming server are deployed at the edge of the 5G network, alleviating the UAV from energy consumption and computational demanding tasks.

The standard communication between the UAV and the Ground Control Station (GCS) is realised using short-range 433 MHz telemetry and 2.4 GHz control links. In order for a 5G communication link to be established, a feature not available by commercial-of-the-self drones at the time of exhibition, a prototype UAV was built, including (see Figure 3-15: An open flight controller, an on-board computer, main/secondary battery and a USB-tethered 5G mobile phone.

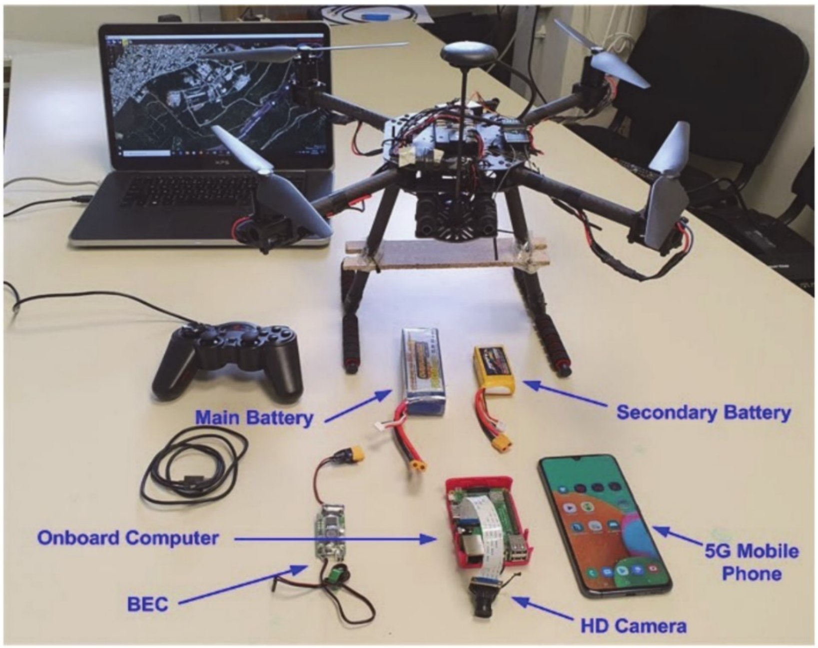

Figure 3-15 Equipment for 5G-enabled UAV

The following Table 3-4 summarizes the equipment that is used for providing the C2 channel over 5G.

Table 3-4 Equipment used for delivering C2 over 5G

Type	Description	Purpose
Main Board	Raspberry Pi 4 model b	Lightweight, low consumption computer
Camera	CCD Camera 5MP Model SEN0184	Live video streaming
Power Supply	LiPo battery 3Sp1 11,1 V - 1300MAh	Mobility
Voltage Regulator	Input: 2S-6S, Output: 5V/3A	Regulate/stabilize voltage to 5V/3A
Cable	USB to micro-USB	Physical UAV to Companion Computer connection
Operating System	Raspberry Pi OS	Companion Computer operation
UAV control software	MAVProxy	Companion computer to UAV communication Ground user to UAV communication

Ground Control Station

The QGroundControl (QGC) software[4] was used for the UAV control communication as depicted in Figure 3-16. Running on Windows laptop connected locally at the edge network of the 5G RAN and paired with a USB gamepad, it provided vehicle setup options and real-time full flight control for our 5G-enabled UAV prototype. QGC communicates over the MAVLink protocol with the MAVProxy, a UAV-control software, which is installed in the Raspberry Pi companion computer attached to the drone and provides:

- Setup, configuration, parameterization options for the vehicle components
- Map display with vehicle position, flight track, waypoints and vehicle instruments

Figure 3-16 Area of trials depicted via the QGroundControl Software

The following Table 3-5 summarizes the equipment used for the GCS and the setup of the architecture is depicted in Figure 3-17.

Table 3-5 Equipment of the Ground Control Station

Type	Description	Purpose
Laptop	Dell Latitude	Hosting UAV control software
Ground UAV Control Software	QGroundControl	Ground user-to-UAV direct communication Ground user- to-UAV communication through the companion computer Telemetry/Control data
Joystick/Transmitter	FlySky FS i6-X	Control the UAV (directly or through QGroundControl software)
Cable	USB Cable	Connect Joystick to Laptop

[4] http://qgroundcontrol.com/

Figure 3-17 Architecture of Ground Control Station

The 5G NSA Network outdoor deployment was based on NOKIA Airscale RAN solution (see Figure 3-18), located at COSMOTE site, paired with Athonet Rel. 15 Core located on NCSRD site and Athonet LBO solution deployed at COSMOTE site. The RAN configuration used is presented at Table 3-6.

Table 3-6 Nokia Airscale NSA RAN Config

5G Band	n78 (3700 MHz)
5G Bandwidth	100 MHz
5G Cell Max Modulation	256 QAM
4G Band	7 (2600 MHz)
4G Bandwidth	20MHz
4G Cell Max Modulation	256 QAM

The NOKIA RAN has a 2x2 MIMO configuration, operating with NR TDD DDDSUUDDDD (4+2+4). The TDD pattern was selected following the Greek National Regulation Authority (NRA) recommendations and following the commercial 5G Auction that took place at the end of 2020. For the experiment execution, the antennas were positioned outdoors, at 15m-30m from the area of the UAV flight.

Figure 3-18: 5G NR Antennas at COSMOTE

According to NOKIA, the 5G NR vendor, for the specific configuration and TDD pattern selected, the maximum values expected are 650 Mbps for downlink and 55Mbps for uplink for both TCP and UDP traffic. Practically these targets have been verified as part of the Phase-1 in lab-scale experiments as shown also in Figure 3-19 and Figure 3-20 below.

Figure 3-19: NPERF TCP Measurements with Public NPERF Server in Greece

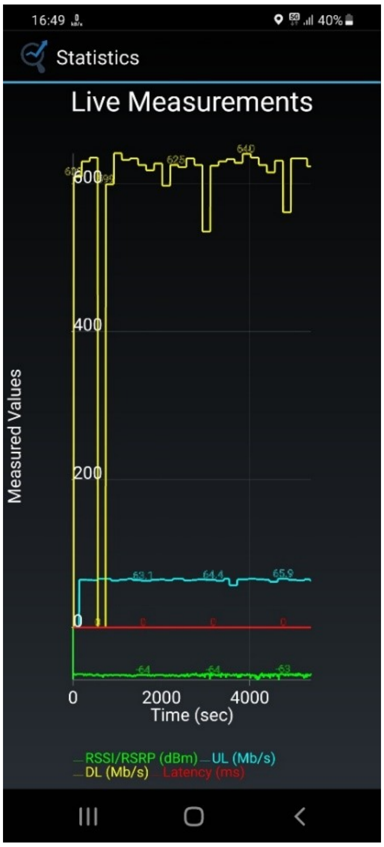

Figure 3-20: UDP, Latency and Signal Strengths baseline lab measurements

The automated experimentation tools by Open 5GENESIS were used for assessing the performance of the system during the C2-over 5G flight at different altitudes. The UAV flight in that case was manual with C2 over 5G, the camera was streaming live content using the streaming server deployed at the edge of the network and the execution of the experiment was performed in an automated way through the 5GENESIS platform. In more details the pilot controlled the UAV over the 5G network utilising the Ground Control Software connected at the MEC. Under the hood, the automation framework of OpenTap has been used for the sequential execution of commands.

Katana Slice Manager, part of the Open5GENESIS Suite, operating from NCSRD Core site, was used to provide the two slices necessary, differentiating the delivery of the C2 channel commands from the media-streaming channel as shown in Figure 3-21 and Figure 3-22. More specifically, the slice labelled *O5GCore_vcache_emmb* serves the core cloud deployment at NCSRD site and the *O5GCore_vcache_urllc* serves the edge cloud (LBO) located at COSMOTE premises.

```
SLICE_ID                                SLICE_NAME
7f670fe3-3aa7-44ce-ad66-f044dbef699f    O5GCore_vcache_emmb
1640fb1e-6018-498e-aacb-1d0f36265963    O5GCore_vcache_urllc
```

Figure 3-21 - Core & Edge Concurrent Slices, Katana CLI

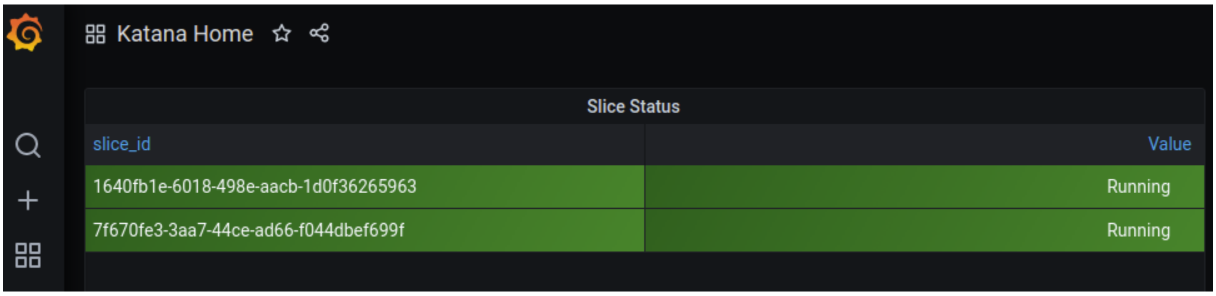

Figure 3-22 - Core & Edge Concurrent Slices, Katana Grafana Monitoring

The video streaming service was successfully delivered via the deployed streaming server at the MEC, where the UEs that were connected to the cell with the LBO capabilities were receiving the video stream with lower latency than the UEs that were connected to the non-LBO cell. Specifically, the mobile device connected to LBO receives the video faster in relation to the laptop that connects to the non-LBO cell.

Figure 3-23 End-user device receiving the video stream with approx. 2sec delay

Figure 3-24 End-user device receiving the video-stream during flight

A VM has been setup in the MEC infrastructure of COSMOTE attached to the LBO node, hosting the iperf3 server (probe) to collect measurements from the mobile application attached on the UE on-boarded to the drone. The iperf logs from both server and client were successfully stored into the InfluxDB provided by 5GENESIS. The performance of the system is depicted through the graphical representations of results as following:

- Reference TCP Download Throughput test directly with the UE located in proximity and at the same level of the 5G NR Antennas (Figure 3-25)
- TCP Download Throughput test with the drone and the on-board UE resting on ground at the flight field, located at a lower level than the Antenna level.
- TCP Download Throughput test with the drone and the on-board UE hovering at 3 meters altitude at the flight field, at a height level equal to the Antenna height (Figure 3-28)
- TCP Throughput test with the drone and the on-board UE hovering at 5 meters altitude at the flight field, at a height level equal to the Antenna height (Figure 3-29)
- Indicative latency measurements at the ground level

Figures 3-26, 3-28 and 3-29 are line graphs depicting measurements over time for the two agents used (blue line for iperf-client & orange line for iperf-server). The points depicted are `[UNIX-epoch timestamp, throughput value]` pairs and for convenience the X-axis is displaying the timestamps in UTC format. The measurement values were taken directly from the iperf-agents output, then stored to InfluxDB and for the purpose of this experiment the iperf reporting setting `(-i)` was set to 1-sec intervals. One important note here is that the captured timestamps are from two different physical devices which were not intended to be (and should not be considered as) synced to the same clock. Each graph depicts the two different outputs of the **same experiment**, one coming from the server's perspective and the other from the client's perspective. This also explains the slight differences which the reader can easily spot on every graph and is common when comparing iperf results from the two endpoints used. The average throughput results reported, which are depicted at the following table, give a clearer picture of how close they tend to be but since the experiments were taking place during actual drone-flight, the time was limited and performing many iterations that would have helped us generate more statistically solid results were out of scope.

Table 3-7 iperf endpoints reports: average throughput measured at different altitudes located at the flight field 30m from 5G-NR

	Ground level	3m	5m
iperf server	259 Mbps	261 Mbps	261 Mbps
iperf client	279 Mbps	282 Mbps	281 Mbps

Figure 3-27 Drone Flight over 5G at COSMOTE Edge Site

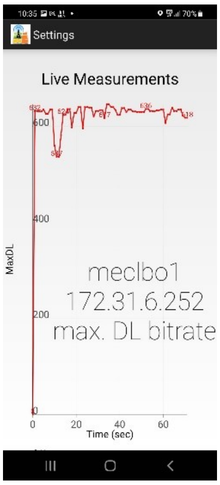

Figure 3-25: Baseline metrics from close (3-5m) proximity to 5G NR

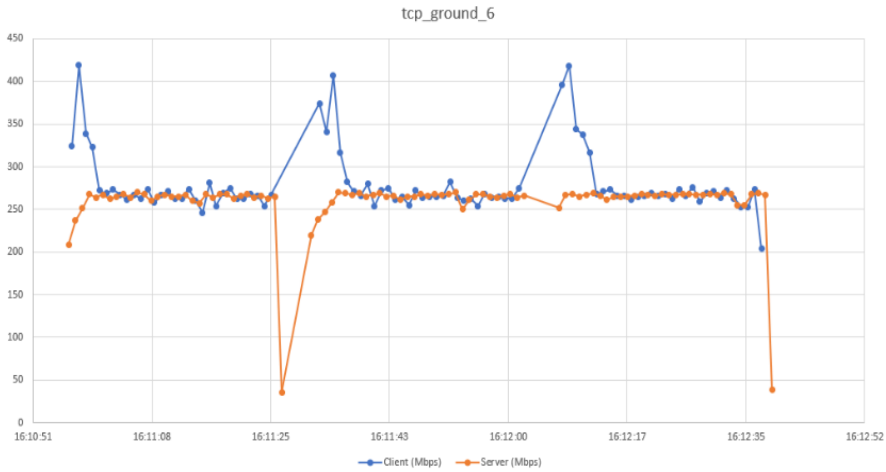

Figure 3-26 Drone TCP Throughput Test, Ground, located at the flight field 30m from 5G-NR

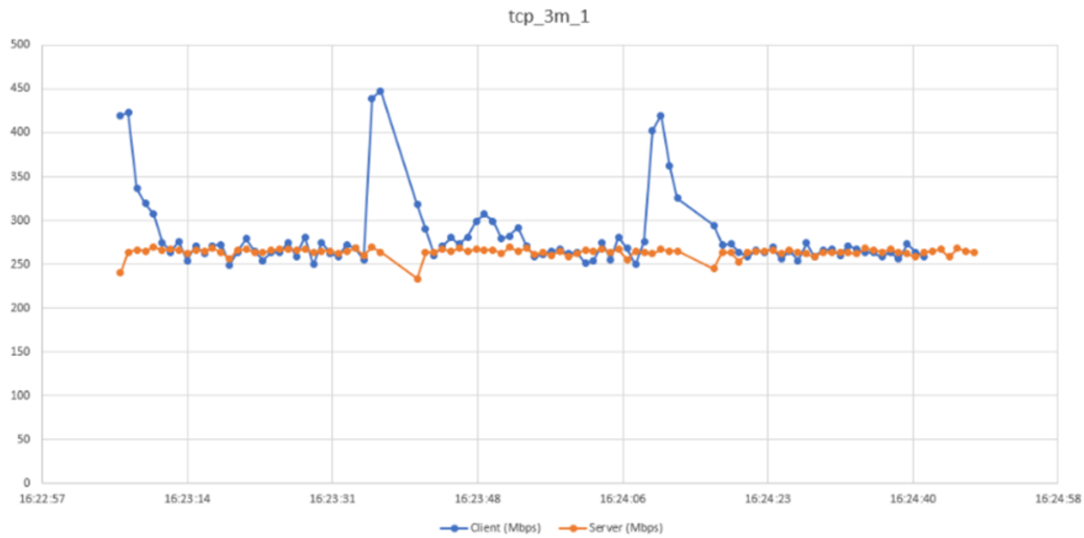

Figure 3-28 Drone UE TCP Throughput test, 3m Altitude at the flight field located 30m from 5G-NR

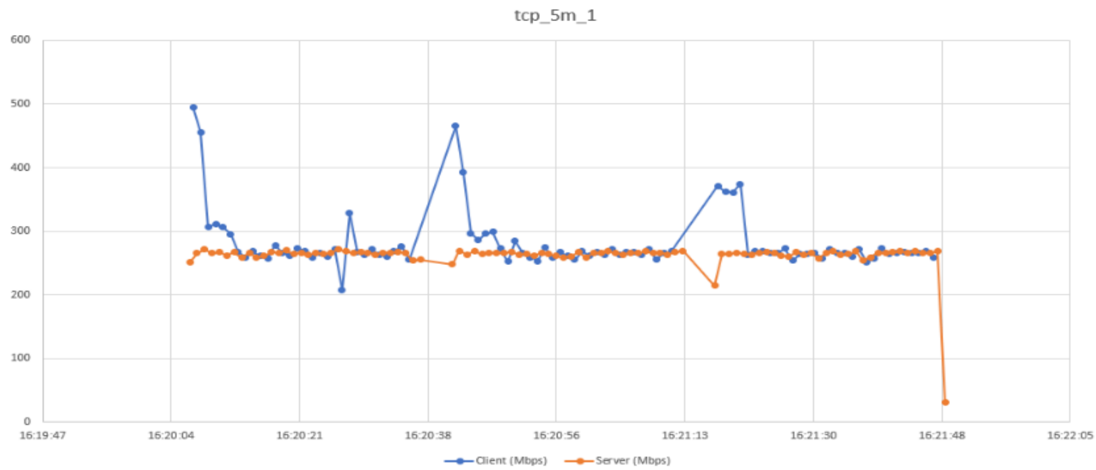

Figure 3-29 Drone TCP Throughput test, 5m Altitude

There is a noticeable degradation of performance from 650Mbps to 300MBps download, as we move away from the 5G NR antennas towards the parking lot standing as the flight field that can be attributed to the distance from the antennas of approximately 30m and the obscured, by a huge tree, Line of Sight.

On the end-to-end latency measurements, collected at random repetitions from the flight field, a range between 14ms and 42ms was observed, accumulating the radio, LBO, and application processing time. As such, this measurement is considered satisfactory.

Figure 3-30: Latency measurements from flight field

3.5. UC#3: SecaaS – Measurements and results

The Security as a service (SecaaS) use case scenario has been refined to be more fitted to the 5G network capabilities. The initial aspiration to demonstrate the use case in an actual e-learning environment at OTEACADEMY was not achieved due to COVID-19 restrictions. However, the scenario was demonstrated in a controlled environment at NCSRD premises. The scenario involves four roles at its minimal configuration, two students accessing the e-Learning class through 5G UEs, the Instructor who is using a broadband wired connection and the attacker who is also connected via 5G and is using DoS attack to disrupt the lesson. As soon as the attack is activated, the deployed SecaaS service employs heuristic and ML based anomaly detection methods to detect and the mitigate the attack. Following the mitigation, the service recovers successfully. The layout of Use Case 3 is illustrated in Figure 3-31.

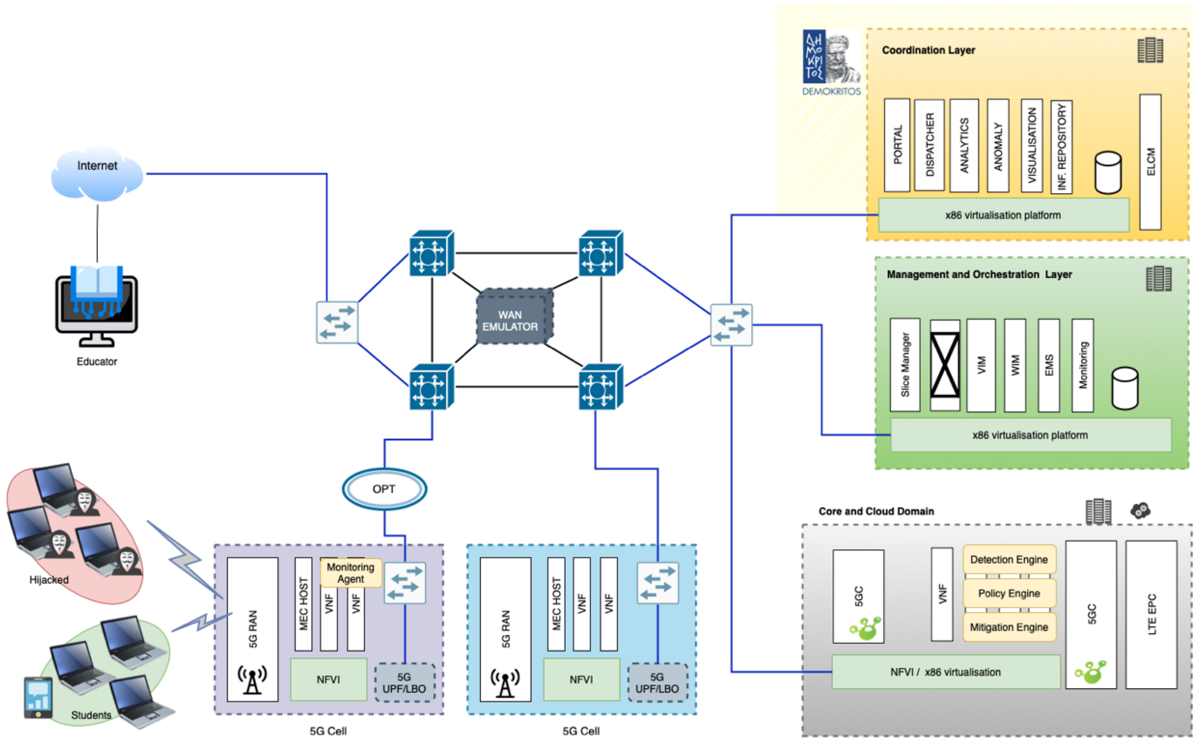

Figure 3-31 SecaaS Use Case

The SecaaS service comprises the following components:

- Monitoring Agent – Responsible for collecting monitoring information by inspecting the network traffic and sending the extracted information to the Detection Engine
- Detection Engine – Responsible for detecting network anomalies and attacks based on introduced policies, data collected by the monitoring agent and detection algorithms. In this case the detection engine, can detect traffic patterns used as DoS attack and identify the cause of anomaly.
- Mitigation Engine – Responsible for processing the information on offending flows detected by the Detection Engine and interaction with the 5G Core to issue a remedy based on specific actions. In this case the Mitigation Engine maps the offender's IPs to the IMEI information and order the immediate logout of these devices from the 5G network.
- Policy Engine - Contains specific service policies and allow/deny lists targeted to specific user groups.

Then framework is based on Wazuh[5] platform with additional implementation of rulesets, policies, and remediation/mitigation strategies.

In the executed scenario the following workflow is examined. Several students are connected via 5G to an external Educator reached via the Internet. The students are participating via video and audio in the lecture. The quality of the service, prior to the DoS attack is the best (Figure 3-32).

[5] https://wazuh.com

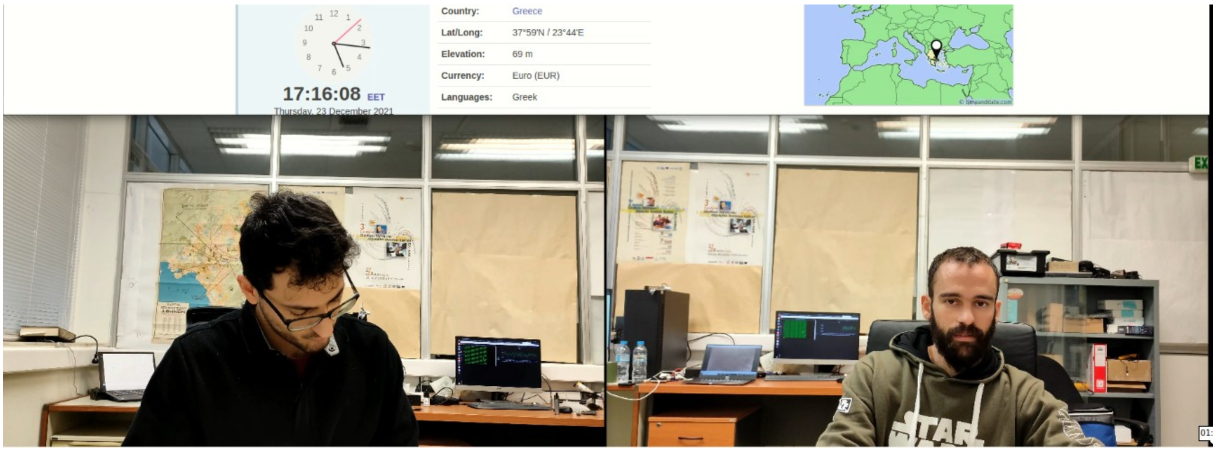

Figure 3-32 Video Streaming from students

The next step is for DoS traffic to be injected in the 5G Network from a 5G terminal. The attack attempts to flood the uplink channel of the 5G cell to disrupt the communication of the students with the educator. Figure 3-33 depicts the attacker screen during the attack.

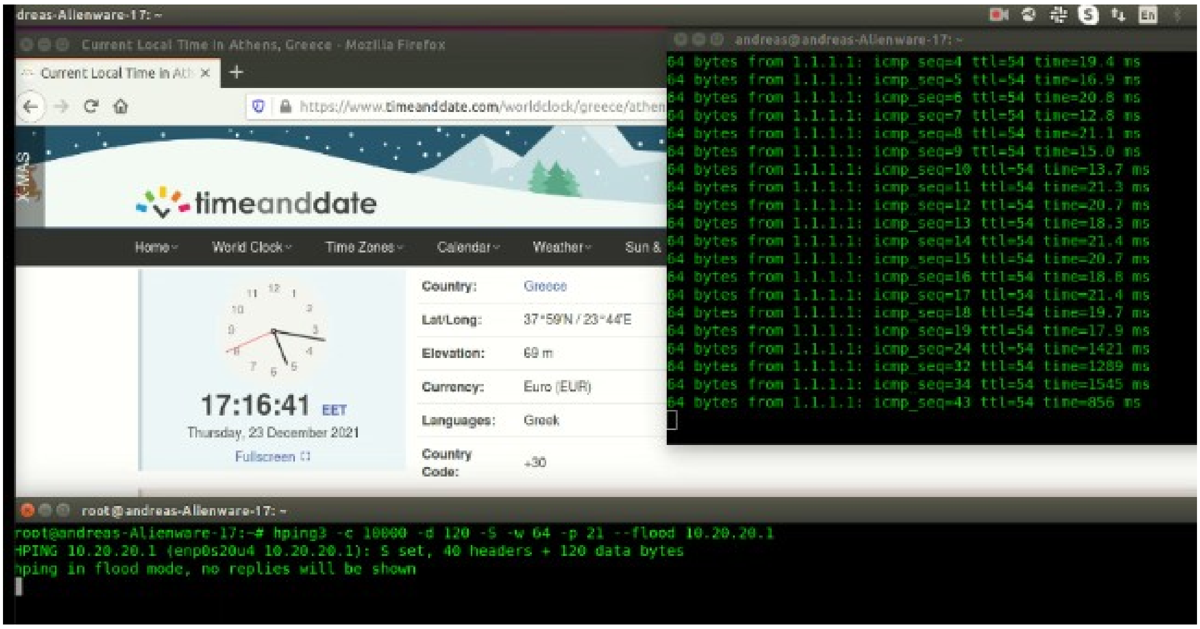

Figure 3-33 Attacker screen during attack

During the attack the decrease in the available uplink can be seen in Figure 3-34. Depending on the number of DoS flows created by the attacker (i.e., DDoS), the impact on the available uplink can be severe. During the attack, the noticeable effect on the Educator terminal will be for the video to completely freeze.

Figure 3-34 Uplink bit rate before, during and after the DoS attack being mitigated.

Due to the almost immediate response and mitigation, the detection engine and therefore the whole detection-mitigation chain was not operational. At some point the detection application is activated. The detection and identification are almost done simultaneously as Figure 3-35 presents.

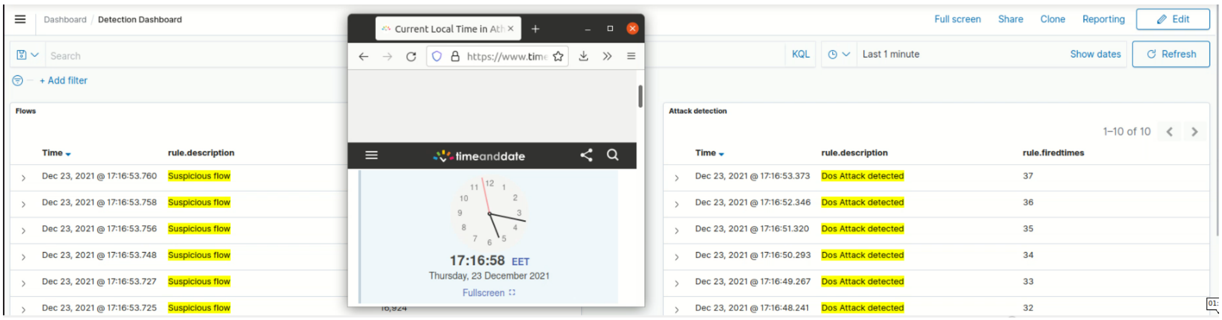

Figure 3-35 Detection Engine GUI.

The left column in Figure 3-35 provides view on the Suspicious Flows detected by the Detection Engine and the right column is providing view on the identified attacks and the number of matches. ***The attack identification and the command to mitigate it takes less than 1 sec***. Upon mitigation the offending UEs (as identified by the mitigation engine), are de-associated and are disconnected from the 5G Network.

3.6. Additional exploratory use cases

This section presents two additional experiments conducted in the Athens platform, that were not initially identified as stand-alone use cases, but were proposed and formed during the development process of the 5GENESIS software components. More specifically, they are built around the integration of the 5GENESIS Slice Manager with two different policy engines, namely Adaptive Policy Execution (APEX) and New Evolutive API and Transport-Layer architecture (NEAT). Although they are not part of the three original use cases defined for the Athens Platform, they offer some valuable results and insights to the concurrent network slicing field. For the exploratory use cases, we rely on use case specific measurements to illustrate performance, as presented in the following sub-sections.

3.6.1. NEAT – Concurrent Slicing

This section presents the results from the experiments regarding the integration between the Katana slice manager and the NEAT policy system. The results illustrate how this integrated solution can make dynamic use of multiple concurrent slices. The NEAT system runs on the UE and decouples applications from the underlying network stack. It allows the choice of transport protocol, its configuration, and the choice of network slice usage to be made dynamically based on the application's requirements, as well as the current network configuration and status, obtained from the slice manager.

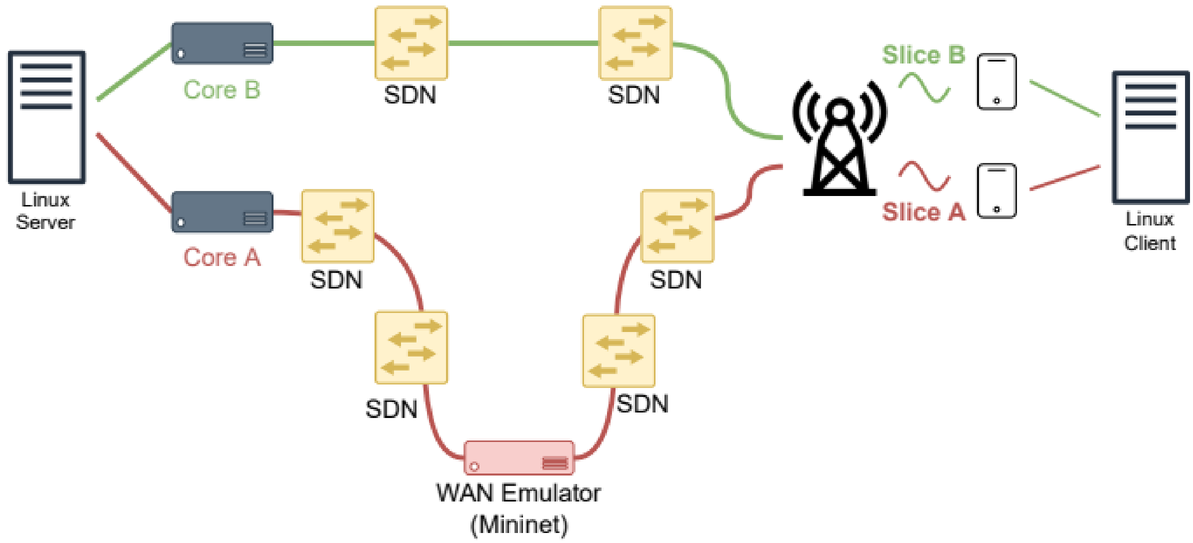

Figure 3-36 Network topology and testbed setup

Figure 3-36 depicts the test setup for the experiments conducted in the Athens platform. It consists of a Linux client machine, which is connected to a Linux server over two different network slices; Slice A, a best effort slice that is used as the default slice, and Slice B, which is optimized for latency sensitive services. Two mobile phones provide network connection to the client machine via tethering. The core and radio components are deployed in 5G standalone mode. QoS and priority policy is applied to the slices using different QCI/5QI radio bearers. NR band n78 (3500 MHz) is used in TDD mode, with 50MHz bandwidth and up to 256-QAM.

Figure 3-37 shows the RTT measured over the two slices **in a drone use case scenario**. The traffic used for the evaluation was comprised of two distinct traffic flows: a latency sensitive control

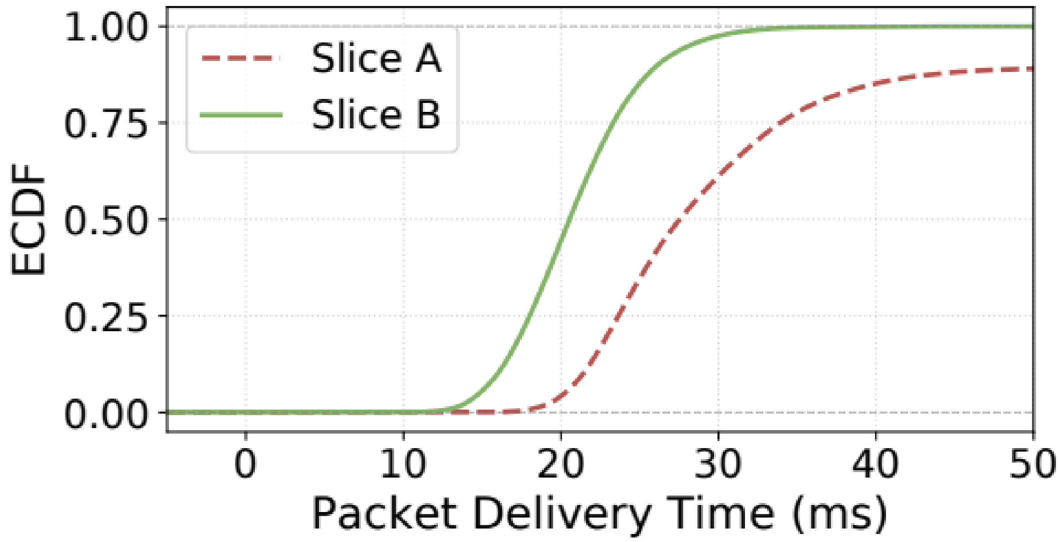

Figure 3-37 Empirical Cumulative Distribution Function of the end-to-end RTT

flow, and a video stream. For the latency sensitive flow, we replayed a packet trace originally captured from the drone PixHawk 4.0 autopilot. The trace is comprised of navigation commands transmitted over the MAVLink protocol between the drone and a controller. For the video flow, Real-Time Messaging Protocol (RTMP) was used to stream a video file to the server. In one scenario, the two flows contend for resources over Slice A, whereas in the other scenario an intelligent choice was made by the NEAT system such that the latency sensitive flow was directed over Slice B instead.

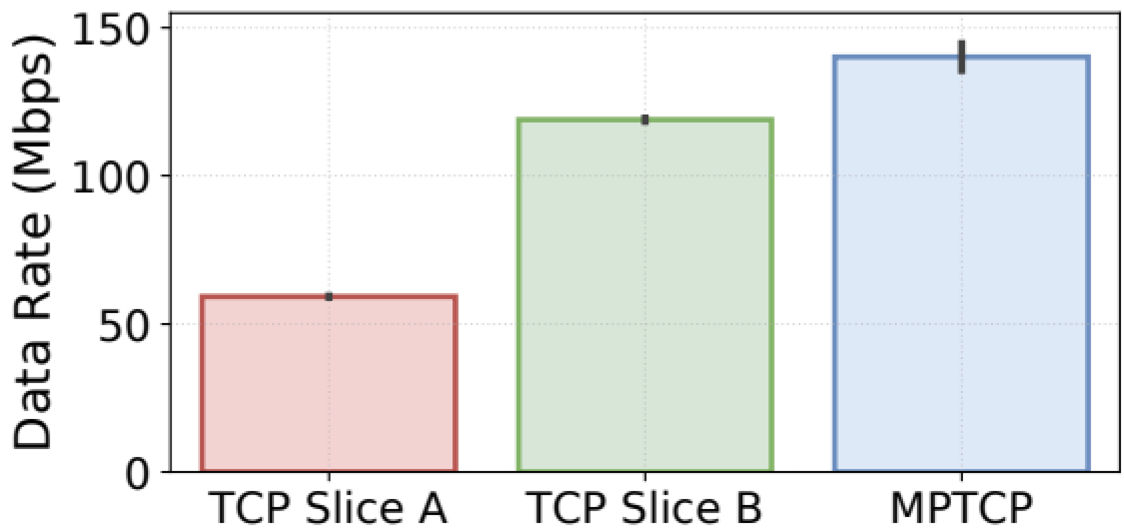

Figure 3-38 Average throughput for single-path TCP over Slice A and Slice B, and MPTCP over both.

Next, we consider a bulk traffic scenario, the traffic was generated using iperf. Figure 3-38 shows the average downlink throughput obtained over each of the two slices as compared to the average downlink throughput achieved by allowing NEAT to enable Multi Path TCP (MPTCP) on the UE. This decision is based on information from the slice manager that the two slices take different paths over the network.

3.6.2. APEX – Slice Manager

As previously described in D3.4, the Adaptive Policy Execution (APEX) engine has been integrated into the Katana slice manager software stack. APEX is triggered by Prometheus, running as part of the Slice monitoring module when it detects that a specific part of the slice is failing or misbehaving. Prometheus then generates an alert and sends it to the APEX policy engine via the internal Kafka message bus. APEX processes the incoming request and recommends a series of appropriate actions to Katana, based on the alert and the underlying system conditions. Such actions may involve interacting with other systems, such as instances of the NEAT policy engine running on the UEs. In this section, we present a use-case where a specific virtual Network Service (NS) fails. Prometheus detects the failing NS and sends an alert to APEX. APEX decides on the actions that must be executed and sends feedback to the Slice Manager.

The APEX policy used for this use-case considers three scenarios: (1) if the concerned slice has no previous history of failure, the slice manager attempts to correct the issue by restarting the misbehaving NS, (2) if the slice has a history of failure, the slice manager will restart the slice with new parameters and restrictions, such as relocating the NS on a new infrastructure component, and also notifies the NEAT system on the UE of the failure, which can then take immediate action to try and resolve the issue at the UE side. Once the slice manager has resolved the issue, NEAT is once again notified, and (3) if the slice has a history of failure, as well as a history of failed attempts to correct the issue, the slice manager is instructed to notify the system admin as well as NEAT on the UE and terminates the whole Slice. In scenarios 2 and 3, a NEAT policy will redirect the network traffic over the failing slice to another slice, until the issue has been resolved.

The testbed setup for these tests is depicted in Figure 3-36. In this experiment, two concurrent slices are used for providing connectivity to the UE where the NEAT policy is running, namely the “fast_path” acting as Slice A and the “slow_path” acting as Slice B. This use-case has been tested through the dynamic steering of a downlink QUIC flow. The QUIC flow is started over the primary slice, Slice A. After 100 seconds, the termination of a specific NS on Slice A is initiated in order to simulate failure (scenario 1). Note that the termination of the NS is not immediate, as the NS is gracefully shut down, which may take at least 40 seconds. After 200 seconds, the NS is terminated again (scenario 2), and therefore the QUIC flow is temporarily redirected to another slice, Slice B, until Slice A has recovered. After 300 seconds, the NS is terminated a final time (scenario 3), which deletes Slice A, and the traffic is permanently directed to Slice B. For scenario 2 and 3, NEAT exploits QUIC's Connection Migration feature in order to migrate a connection from one slice to another.

Figure 3-39 illustrates how the data of the QUIC flow is distributed over the two slices over time. At the 250 seconds mark, an NS part of the Slice A is terminated due to a software error, causing the connectivity through the Slice A to be lost. The Slice Manager monitoring system detects the NS failure and applies the policy instructed by APEX; it restarts the said NS and immediately notifies the NEAT components on the UEs connected to this Slice. On the UE side, NEAT intervenes and reconfigures the UE connectivity over the backup Slice B until the affected NS is restarted and connectivity on Slice A is restored. However, the software error persists, causing the NS to fail again. At that point, APEX applies another policy on the Slice Manager, terminating the Slice A with the malfunctioning NS. Moreover, NEAT is once again notified and permanently reroutes the traffic over the backup Slice B. As a result, we achieve the minimum connectivity disruption while achieving the maximum throughput by returning to the "fast_path" slice whenever possible.

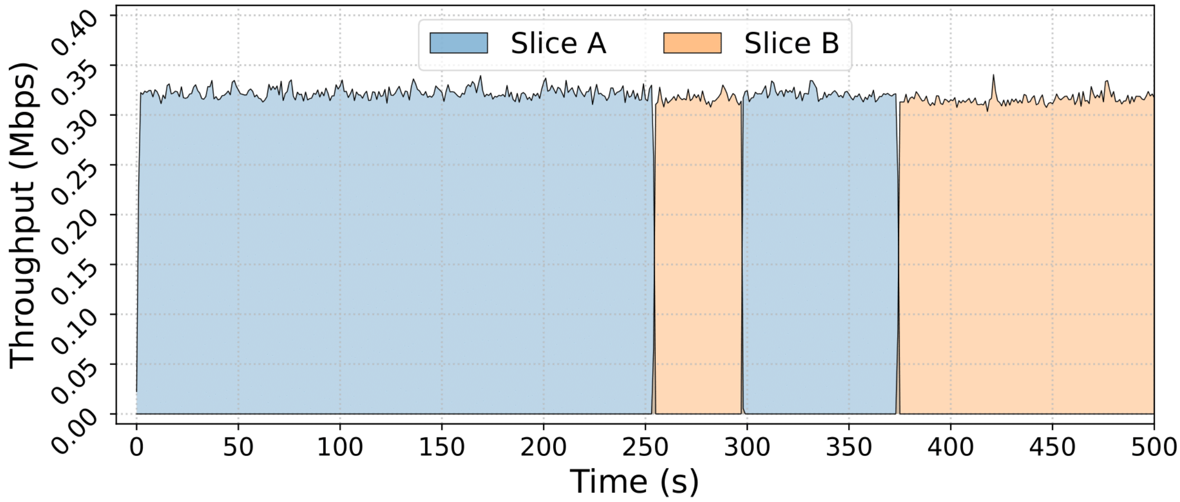

Figure 3-39 Dynamic steering of a single QUIC flow over two slices using NEAT, assisted by APEX and Katana

3.7. Portable platform: OAI SA end-to-end performance results

In the current section, we provide the obtained performance results from testing the OAI end-to-end SA setup on the portable platform environment. This setup is composed of the OAI 5G-NR gNB (software running on top of a general purpose x86 server connected with a 5G RU), a containerized deployment of the OAI 5G CN and a COTS UE module. The implementation details of the OAI 5G SA setup have been described in [18]. The objective here is to describe the performance improvements that have been achieved during the final trials phase of the project.

Table 3-8 summarizes the main hardware equipment and software configuration parameters of OAI that were used during the tests. The experiments were conducted in indoor lab environment with very good LOS channel conditions. In practice, this means that for both DL and UL the highest allowed MCS was selected.

Table 3-8 Main hardware equipment and OAI software configuration parameters for SA end-to-end tests

5G RU	ETTUS USRP N310
COTS UE module	Quectel RM500Q-GL[19] or SIMCOM SIMCOM8200EA[20]
Channel Bandwidth	60 MHz
SCS	30 kHz
Downlink allocated PRBs	162
Uplink allocated PRBs	80

Number of scheduled DL slots per 10 slots	6
Number of scheduled UL slots per 10 slots	2
Max modulation scheme DL/UL	64-QAM

The throughput and RTT KPIs were measured with iperf and ping tools respectively. For the case of throughput measurements, the communication endpoints were the COTS UE and the SPGWU component of the core network. For the RTT measurements the endpoints were the UE and a remote internet host.

Table 3-9 shows the performance results with respect to end-to-end downlink and uplink throughput and RTT, based on the configuration provided in Table 3-8. The measured DL throughput of *141-142 Mbps* is close to the expected theoretical throughput (*~150 Mbps* for the used configuration). Similarly, the measured UL throughput of *13.6-14 Mbps* is close to the expected theoretical value (*~15Mbps*). Both throughput measurements were stable throughout the experiments, without significant difference between the min and max values. RTT measurement results are also satisfying based on the used configuration and the range between the min and max values is acceptable given the end-to-end nature of the experiment.

Table 3-9 DL/UL throughput and RTT based on OAI end-to-end SA setup

KPI	Min	Average	Max
DL throughput	141 Mbps	**141.47 Mbps**	142 Mbps
UL throughput	13.6 Mbps	**13.89 Mbps**	14 Mbps
RTT	9.88 msec	**12.1 msec**	15.33 msec

3.8. Summary and conclusions

At the Athens 5GENESIS facility over the course of the project both 5G deployment options have been validated, i.e., 5G NSA and 5G SA. The current document presents validation results and trials as well as specific experimental scenarios using 5G SA deployments. The validation results illustrate the benefit especially in terms of latency that is achieved with the 5G SA deployment. The presented results for 5G performance KPIs achieved, depend heavily on the features and COTS equipment availability. In most of the KPI validation scenarios executed in the Athens platform the added value of 5G vs 4G is distinctly demonstrated, so does the adoption of 5G SA vs 5G NSA. The 5G NSA results have been discussed in D6.2. For easy comparison

- E2E RTT: Using 5G and the flexible TTIs, we reached approximately an E2E RTT of 12ms (64 bytes), including the core network. This value is lower than the E2E RTT in 4G, measured in Phase 1, which was approximately 37ms (64 bytes, D6.1 – Trials and

Experimentation, Cycle 1). The added value of 5G reflected in the measurements is that we are able to provide more use cases in the Athens Platform, by using flexible numerology.

- Throughput, the average was measured for NSA and SA modes using UDP and also TCP. The NSA mode yielded a performance of 363.28 +/- 1.00 Mbps for 5G NSA and 314.42 +/-2.56 Mbps for 5G SA. For TCP the values is quite lower, as expected, however the max value of throughput measured during TCP is 361.88 +/- 1.14 Mbps. This provides some consistency with the expected maximums based on theoretical values provided by the vendor (Amarisoft). Also as expected the impact on throughput between NSA and SA is negligible for similar configurations and any observed differences are based on vendor HW implementation.
- For Core vs Edge placement of components, the results reveal that the processing delays imposed by the service components have more severe impact in the overall latency. However, the impact of Core vs Edge starts to affect the overall performance when the backhaul connection is congested. Especially regarding our 360 streaming video use case the most impacted metric was drop incidents and number of dropped frames.
- Slice manager/NEAT Slice manager is part of the Open5GENESIS suite, with capabilities to create slices over the whole infrastructure and operate them concurrently. As the original implementation of the scheduling and placement components did not change the Service Deployment Time measurements presented in D6.2 are still valid. However additional experiments exploiting concurrent slicing and day 2 operations on slices plus monitoring, reveal the ability to deploy innovative services across 5G infrastructure.

In addition, the previous validations, three trials took place, although due to COVID restrictions the planned deployments were affected. However, two of the trials took place at the actual setting that was planned slightly modifying the scenario. The three trials included 5G configurations with eMBB and URLLC slice deployments exploiting core or edge component deployments. The key takeaway from these trials is that 5G technology is flexible and may serve different vertical requirements that may require either low-latency or high throughput using the same infrastructure. All trial scenarios were deployed via Open5GENESIS Platform via creation of specific network slices configured differently for each scenario. The validations could not provide results at the full extend due to the limitation imposed by the deployment that was possible at the time.

4. MALAGA PLATFORM EXPERIMENTS

4.1. Overview

At the Malaga platform the experimentation in phase 3 has been focused on the validation of the new 5G deployment based on 5G SA picocells deployed in the Ada Byron building located at the University Campus, while most of the trials have been organized at the city center under the umbrella of the 5G NSA deployment available there. The city center represents a more realistic environment where the police used the 5G NSA network for the following use cases:

- To add new mobile cameras for video surveillance in addition to the fixed cameras connected by fiber.
- To deploy a mobile command center closer to the location of massive events.
- To equip cops with 5G devices which included Mission Critical Push to Talk Communications (MCPTT), access to video surveillance cameras and streaming of live videos recorded during their actions.

Trials have been reported in D7.4 and in this document we include technical details.

In this third phase we have evaluated again throughput and latency KPIs for the new SA deployment. Service creation time, evaluated in Phase 1, has not been repeated because the virtual infrastructure used in Phase 3 hasn't changed significantly. Reliability, capacity, speed and location accuracy have been measured in this last round of experiments. Table 4-1 summarizes the KPIs evaluated at the different phases of the project.

Finally, baseline results of millimeter wave experiments are included in this report. These results belong to the setup based on the 5G gNode emulator from Keysight, which offers support for millimeter wave.

Table 4-1 5G KPIs evaluated at the Málaga Platform

KPI to be evaluated at the Málaga Platform according to DoA	Evaluated in Phase	Comment
Throughput	Phase 1, Phase 2, Phase 3	Based on iPerf
Latency	Phase 1, Phase 2, Phase 3	Based on RTT
Service Creation Time	Phase 1	Base on the slice manager
Reliability	Phase 3	Based on RTT
Capacity	Phase 3	iPerf
Location accuracy	Phase 3	Specific
Speed	Phase 3	iPerf
Density of users	N/A	UEs emulator is required to measure this KPI

Additional 5G KPIs evaluated at the Málaga Platform		
MCPTT Access time	Phase 1, Phase 2, Phase 3	MCPTT app-
MCPTT End-to-end Access	Phase 1, Phase 2, Phase 3	MCPTT app
Content distribution streaming services: Video resolution, Time to load first media frame	Phase 1, Phase 2, Phase 3	Video streaming client app

Phase 3 experimentation and trials have been conducted at the infrastructure described in Table 4-2 The Network Functions Virtualization Infrastructure (NFVI) of the testbed is managed by two different technologies distributed in two different domains: OpenStack for the main data center and OpenNebula for the edge infrastructure. In the main data center, there are three dedicated servers to host and manage the network service instances using OpenStack (Rocky release): the controller, the compute and the storage nodes. In the Edge NFVI, an Open Nebula version 5.8.1 is available. On top of both virtual infrastructures there is a single OSM Release 6 orchestrator handling the NFV deployments. The main data center is also hosting the band base unit (BBU) for the 4G, 5G NSA and 5G SA deployment and the core networks.

The most relevant components added in Release C are the 5GCs, the 5G emulator from Keysight and the new SA devices.

Table 4-2 Release C of the Malaga infrastructure

Deployment Parameters	5G Products/Technologies Options			
Setup ID	1. Full E2E 4G & 5G	2. Keysight 4G & 5G emulator	3.Indoor 5G ECM	4.Indoor 5G REL
Description	Indoor & outdoor E2E 4G & 5G (NSA and SA)	Indoor full 4G & 5G network emulator	5G setup with ECM OAI solution	5G setup with RunEL solution
Core Cloud	Yes - OpenStack	No	No	No
Edge Cloud	Yes - OpenNebula	No	No	No
# Edge Locations	1	NA	NA	NA
Slice Manager	Yes - Katana	NA	NA	NA
MANO	OSM v6	NA	NA	NA
NMS	TAP	TAP	TAP	TAP
Monitoring	Prometheus	NA	NA	NA
3GPP Technology	4G LTE+, 5G NSA, 5G SA	4G LTE+, 5G NSA, 5G SA	5G	5G
3GPP Option	NA	NA	NA	NoS1
Non-3GPP Technology	NA	NA	NA	NA

Deployment Parameters	5G Products/Technologies Options			
Core Network	Athonet Rel. 15 vEPC and 5GC Polaris NetTest EPC and NetCore 5GC Rel. 15	Keysight 4G and 5G UXM wireless test platform	Athonet Rel. 15 vEPC Polaris NetTest EPC Rel. 15	No Core
RAN	Nokia Airscale System (indoor and outdoor)	Keysight 4G and 5G UXM wireless test platform	OAI eNB/gNB	RunEL eNB/gNB
UE	COTS UE	COTS UE	COTS UE	RunEL UE Emulator
Relevant Use Cases	Use Cases 1, 2, 3	NA	NA	NA

The most important achievements of the third experimentation phase have been the throughput reached in the 5G SA deployment, which is higher than 1 Gbps, the successful trials organized to support Police actions during massive events, the millimeter wave tests conducted with the 5G emulator from Keysight and the local breackout setup deployed at the Edge Data Center.

4.2. Generic tests – Measurements and results

Generic tests have been applied to validate the 5G SA. The 5G SA deployment is composed by 2 pico RRHs connected to a 5GC from Athonet. The 5G SA deployment is located at the Ada Byron research building of the University of Málaga.

Table 4-3 5G SA network configuration

Band	n78
Mode	TDD
Bandwidth	100 MHz
Carrier components	1 Carrier
MIMO layers	4 layers
DL MIMO mode	4x4
Max Modulation	256QAM
Beams	Single beam
Subcarrier spacing	30 kHz
Uplink/Downlink slot ratio	1/4

4.2.1. Throughput

This test is devoted to the measurement of the throughput in the downlink between the main compute node and a 5G UE. The test has been executed automatically via the 5GENESIS Coordination Layer, iPerf TAP plugins and the iPerf agents developed in WP3. There is a direct line of view between the 5G UE and the picocell. The average CQI (Channel Quality Indicator) is 13, which indicates a good quality (the maximum value for the CQI is 15). The configuration has a bandwidth of 100 Mhz, which means an increase of 60 MHz respect the 5G NSA configuration used is D6.2, also, in this configuration, four (4) layers are available, in contrast with the two-layer available in the 5G NSA setup. These improvements and the good radio propagation conditions, as indicated by the CQI, have enabled to reach a throughput well in excess of 1 Gbps as shown in Figure 4-1

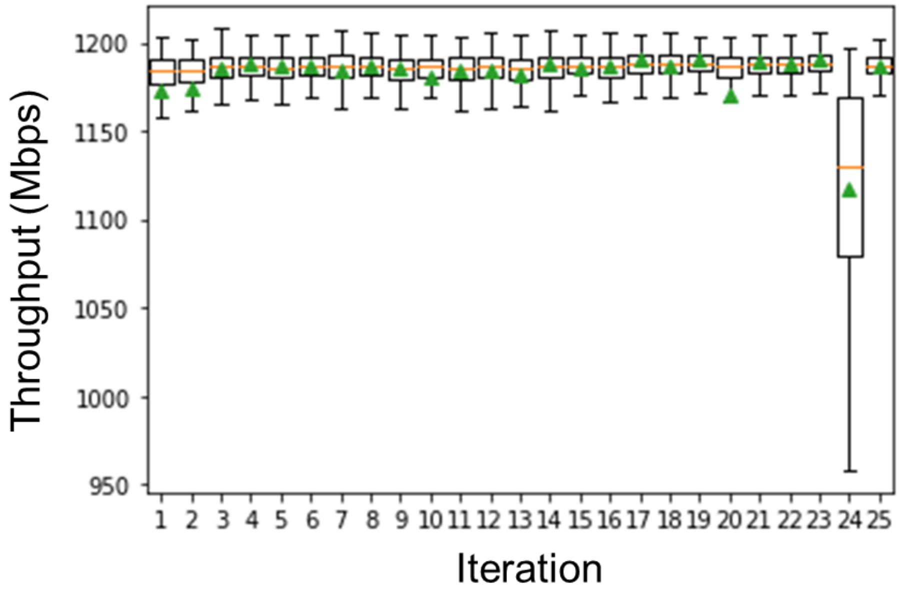

Figure 4-1 5G SA MIMO 4x4 100 MHz 256 QAM throughput

4.2.2. Round trip time

This test is devoted to the measurement of the RTT between a 5G UE and the Packet Data Gateway of the EPC. The test has been executed automatically via the 5GENESIS Coordination Layer, ping TAP plugin and the ping agent developed in WP3.

The proactive scheduling was also activated in the 5G SA configuration and mean round trip time obtained is similar to the value obtained in the 5G NSA scenario and lower that the 24.8 ms reported in D6.1 for ideal conditions in an LTE radio access emulator.

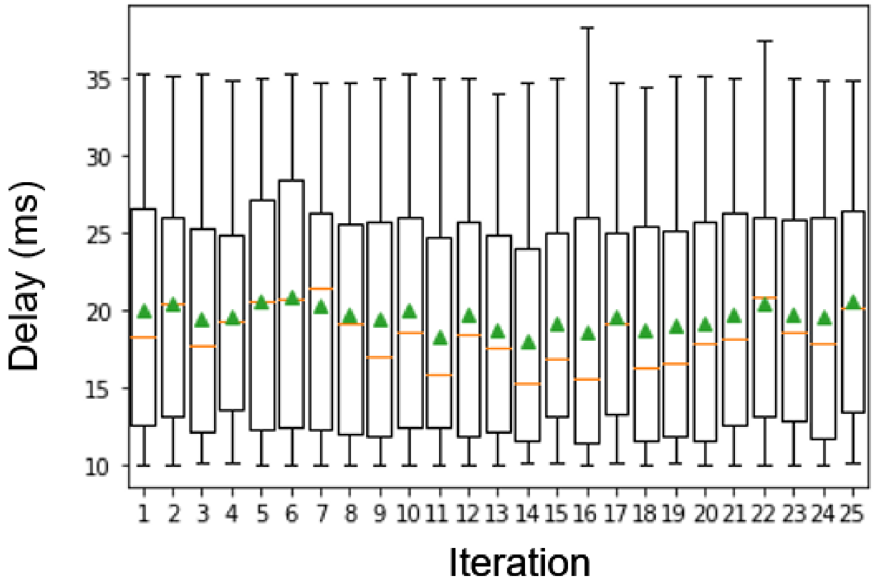

Figure 4-2 SA MIMO 4x4 TDD 100 MHz 256 QAM RTT

4.2.3. Area traffic capacity

The area traffic capacity refers to the total traffic throughput within a certain geographic area, and it is expressed in Mbps/m². The 5G SA deployment has only two picocells, the value obtained for this deployment has a radius of 5 meters per picocell. For a real deployment with a higher number of picocells the throughput measurement will degrade due to interferences from neighbor cells. In such a dense deployment, further reducing the distance between picocells will increase capacity and vice versa, increasing the distance will reduce the capacity (per square meter) as the served area per cell will be larger. We have measured in one of the cells under the assumption that the results are similar in the other one, so we use $TRxPs$ =1 in the calculations below.

Parameters	Formula
Effective Bandwidth (Hz)	$BW_{eff} = 100 \text{ MHz} \times \frac{4 \text{ DL Slots}}{5 \text{ Total Slots}} = 80 \text{ ef. MHz}$
Average Aggregate Throughput (Mbps)	$1182{,}187\ Mbps$
Average Spectral Efficiency(bit/s/Hz/TRxP)	$SE_{avg} = \frac{1182{,}187\ Mbps}{100 \text{ MHz} \times \left(\frac{4\ DL\ slots}{5\ Total\ slots}\right) x\ 1\ TRxPs} = 14{,}78$
Area (m²)	$\text{Area} = 3{,}14 \times 5 \times 5 = 78{,}5 \text{ m}^2$
Site density (TRxP/m²)	$\rho = \frac{1}{78{,}5\ sq.meter}$ = 0,012 TRxP/m²
Estimated average area traffic capacity (Mbps/m²)	$C_{avg} = \rho \times W \times SE\_avg = \frac{1}{78{,}5\ sq.meter} x$ 14,78 x 80 ef.MHz = 15,05 Mbps/m²

The estimated average area traffic capacity is **15,05 Mbps/m²**. This value is aligned with the theoretical value proposed for indoor hotspot deployments in the literature [16].

4.2.4. Location accuracy

This test is devoted to the measurement of the Location accuracy KPI as defined within the 5Genesis context. In order to support the testing and measurement of this KPI, the Málaga Platform has integrated an Location Service (LCS) solution named LocationWise provided by Creativity Software. Since it does not appear to be any commercial LCS solution supporting 5G location procedures, Málaga Platform has focused on this solution, its possibilities and functionality, and also on the measurement of the location accuracy using 4G procedures with different location algorithms as a baseline for future 5G location experiments. It is important to note that the location experiments have been performed over Málaga Platform's 5G network, and that just the location procedures are the ones standardized for 4G.

The most basic positioning method supported by the LCS solution is the so called Cell ID, in which the area for the estimated location is equal to the coverage area of the serving cell the phone being located is attached to, as shown in Figure 4-3

Figure 4-3. Example of Cell ID location area estimation (blue) in cell coverage (green)

This method evolved into Enhanced Cell ID, E-CID, which uses additional information to provide a more precise location. Two variants of E-CID have been used in these experiments:

- E-Cell ID + Timing Advance (CITA), which uses the values of Timing Advance type 1 or type 2 provided by the cell to estimate the distance from the UE to the cell, and hence limiting the estimated area of location, as seen in the figure below. It is important to note that the further the UE is from the cell, the bigger the estimated location area gets, due to the geometry of the coverage area.

Figure 4-4. Example of CITA location area estimation (blue) in cell coverage area (green)

- E-Cell ID + Geo Multilateration (Geo), in which the distance between the UE and the cell is also estimated, but additionally the information from neighbour cells is used, specifically its overlapped coverage area, as can be seen in Figure 4-5. In the case of Geo, the overlap of the different cells in the network is a very important aspect that will set how good the location estimation is in the different locations, being spots with reduced area and multiple overlaps the ones that will provide the best accuracy for the location.

Figure 4-5. Example of GEO location area estimation (blue) in cells coverage areas (green and yellow)

The location accuracy KPI experiments executed at the Málaga Platform consist of four sets of measurements. Two sets correspond to the experiment executed for the CITA algorithm, and the other two correspond to the GEO algorithm, and for each algorithm one set is in a geographical spot where just two cells were serving the UE and the other one for a spot where three cells were serving the UE. An important remark is that the spot of the two cells was closer to the RRHs than the spot of the three cells, as seen in Figure 4-6. This has an impact on the results as will be explained.

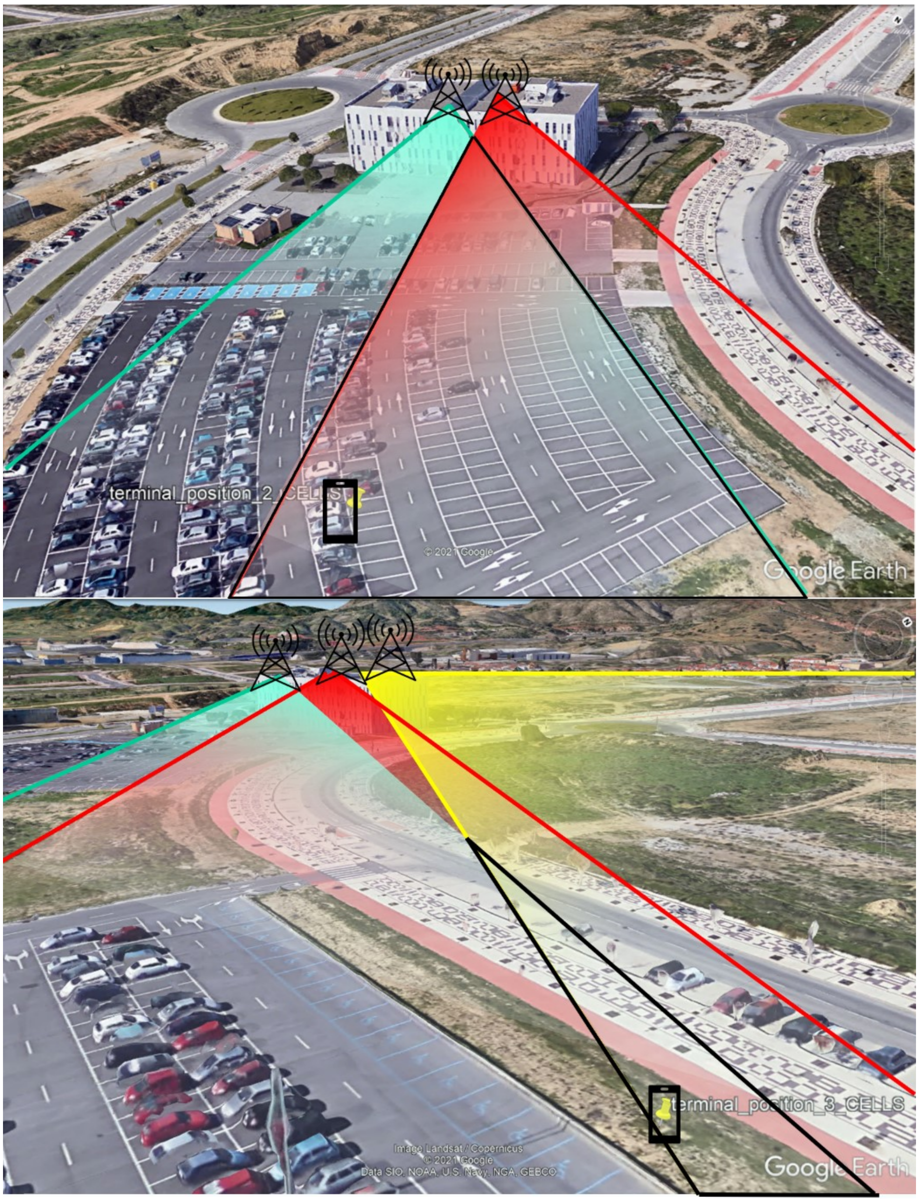

Figure 4-6. Real position of the UE and cells estimated coverage areas for the experiments

Another important detail to note is that the area with coverage from three cells is extremely narrow in the platform's deployment, and hence the coverage of one of the cells was unstable in the three cells spot, meaning that at some moments the UE could just be served by two cells. This can be spotted in the following results graph in the line for GEO algorithm and three cells, where some spikes (at iteration 13 or 16 for example) in the line correspond with moments with just two cells.

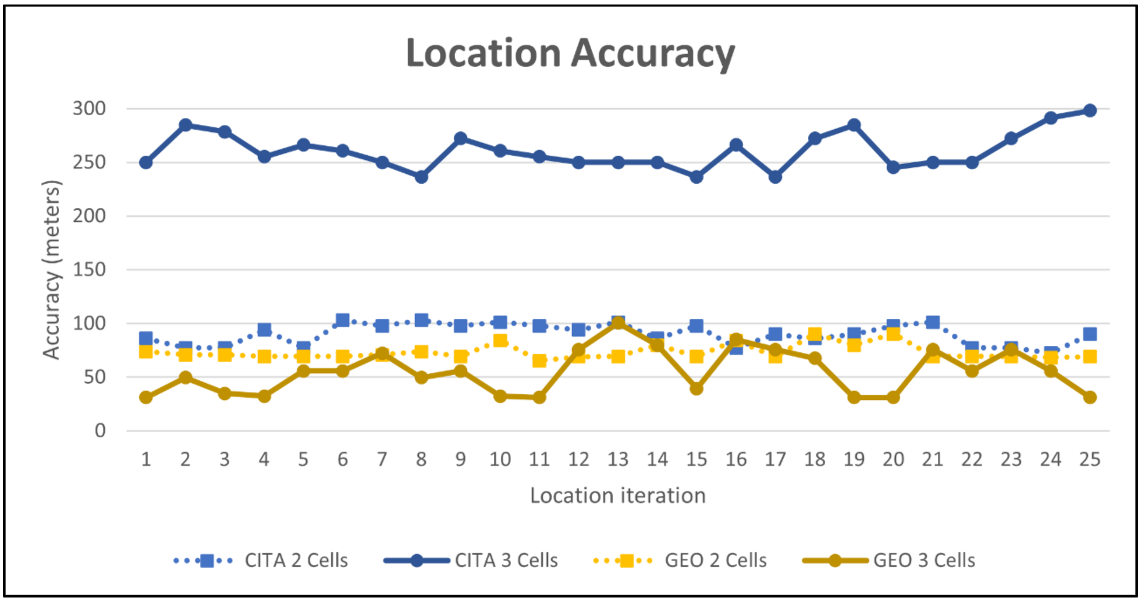

Figure 4-7. Location accuracy experiments results

Completing the experiments we have confirmed the expected behaviour of the different methods, and have extracted three specific conclusions:

- CITA method suffers a degradation in its accuracy with an increase of the UE distance to the cell. This has been confirmed comparing the experiments for CITA for two and three cells, since the spot for three cells was further away from the cells than the spot for two cells, and hence for CITA the results for the three cells scenario present a worse accuracy result.
- GEO method improves its accuracy when the UE discovers more cells. This has been also confirmed comparing the scenarios with two and three cells for Geo, and in this case the spot where the UE was under the coverage of three cells presented a better value of accuracy. Taking into account that this method uses the overlapping between different cells, it was expected that being under the coverage of more cells would improve the location, but it is important to note that it also depends on how the different cells coverage overlaps (e.g., a full overlap of two cells would not improve the location estimation).
- Even if the location estimation area reported by the LCS system is smaller, the location is not necessarily better, since the estimated area could be located some meters away from the true location. So the conclusion is that apart from achieving a small precise estimated area, it is also important that the parameters reported to the LCS system and its processing bring to an accurate positioning of this estimated area.

A general conclusion that has also been achieved thanks to these experiments is that the behaviour and results of the locations performed by the LCS system depend enormously on the radio network deployment, e.g. the position and density of the cells in the area of interest. This makes difficult to be able to measure the location accuracy KPI, or at least to measure what would be the best possible result that the LCS system can achieve, since that result would vary from one network to another with a different radio deployment.

Nevertheless, the results obtained for our radio deployment are quite satisfactory, achieving a location accuracy of around 50 meters for favourable scenarios, which should be the typical in a commercial urban scenario where the density of cells is high. In the future the Málaga Platform will evolve to integrate 5G based location once available, and then we will again perform experiments to measure the location accuracy and assure that the technology comply with the initially established objective of 1 meter location accuracy.

4.2.5. Reliability

The reliability test case refers to the percentage value of the amount of sent network layer packets successfully delivered with the time constraint required by the targeted service divided by the total number of sent network layer packets.

During the tests we have identified 5 different time intervals, reaching a 99.99 % of successful packets received for a RTT < 40 ms. This value of the end-to-end latency is suitable for mission critical use cases such as user plane push to talk voice, mission critical data and mission critical delay sensitive signalling and for other industrial use cases such as process automation which request a maximum RTT of 60 ms and a reliability of 99.9 %. For more stringent requirements regarding the end-to-end latency additional Release 16 should be applied. Detailed results are provided in the Appendix.

4.2.6. Speed

In order to quantify the impact of speed on throughput performance drive tests have been conducted around the Ada Byron building. In this case we used the 5G NSA outdoor deployment at the Ada Byron building, which is located at the university campus in an urban area. Due to the speed restrictions in this area the maximum user speed during the tests was 40 km/h. The maximum throughput in static and ideal conditions for this 5G NSA deployment described in Table 4-4 is 287 Mbps.

Table 4-4 5G NSA configuration

Band	n78
Mode	TDD
Bandwidth	40 MHz
Carrier components	1 Carrier
MIMO layers	2 layers
DL MIMO mode	2x2 Closed Loop Spatial Multiplexing
Modulation	256QAM
Beams	Single beam
LTE to NR frame shift	3 ms
Subcarrier spacing	30 kHz
Uplink/Downlink slot ratio	1/4

From the NGMN 5G whitepaper [17] 50 Mbps in the downlink is expected on minimum for use case mobile broadband in vehicles. As we can see in Figure 4-8 the PDCP (Packet Data Convergence Protocol) throughput reached during the drive test is, in average, higher than 50

Mbps. The lower values of throughput correspond with locations close to the edge of the cell, as we can appreciate in the RSRP values depicted in Figure 4-8.

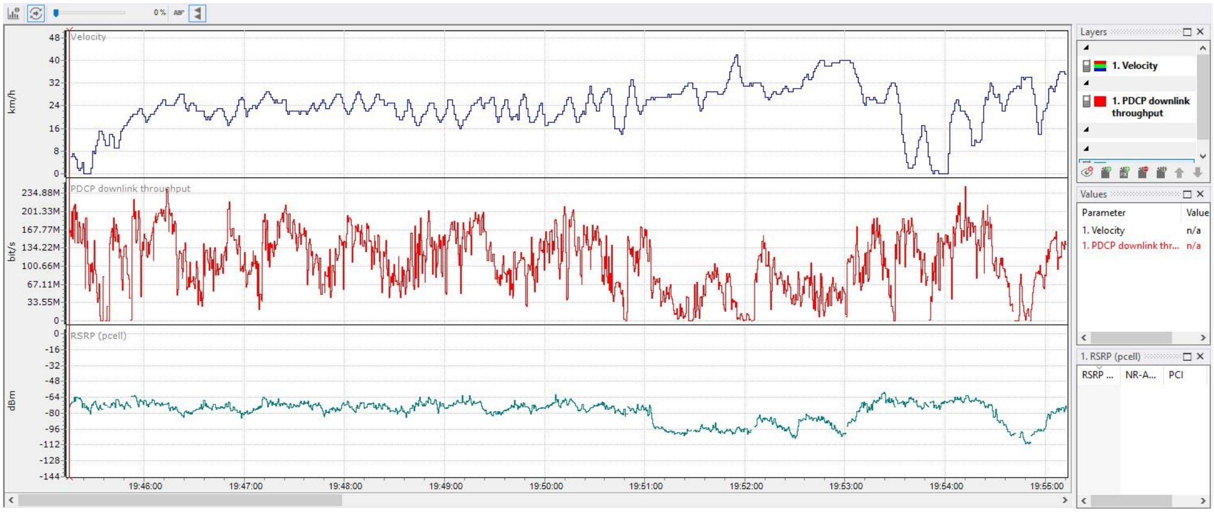

Figure 4-8 Drive test results

4.3. UC#1: Wireless Video in Large Scale Event – Measurements and results

4.3.1. Content distribution streaming services

The test cases defined in TRIANGLE project for content distribution streaming services have been executed to quantify the performance of video streaming services in the 5G SA deployment. The first test case calculates the time to load first media frame, a very important KPI in public safety applications. Figure 4-9 shows the current results obtained in the 5G SA deployment and the results obtained in the 5G NSA and reported in D6.2. This KPI can be affected by the computational resources available in the mobile devices. Figure 4-10 depicts time to open the video streaming application for one the 5G NSA devices used in D6.2 (blue line) and the 5G SA device used in D6.3. The results reveal that One plus 9 offers a better performance than One Plus 7 which can explain the peaks reported in D6.2 for 5G NSA when measuring Time to load media frame.

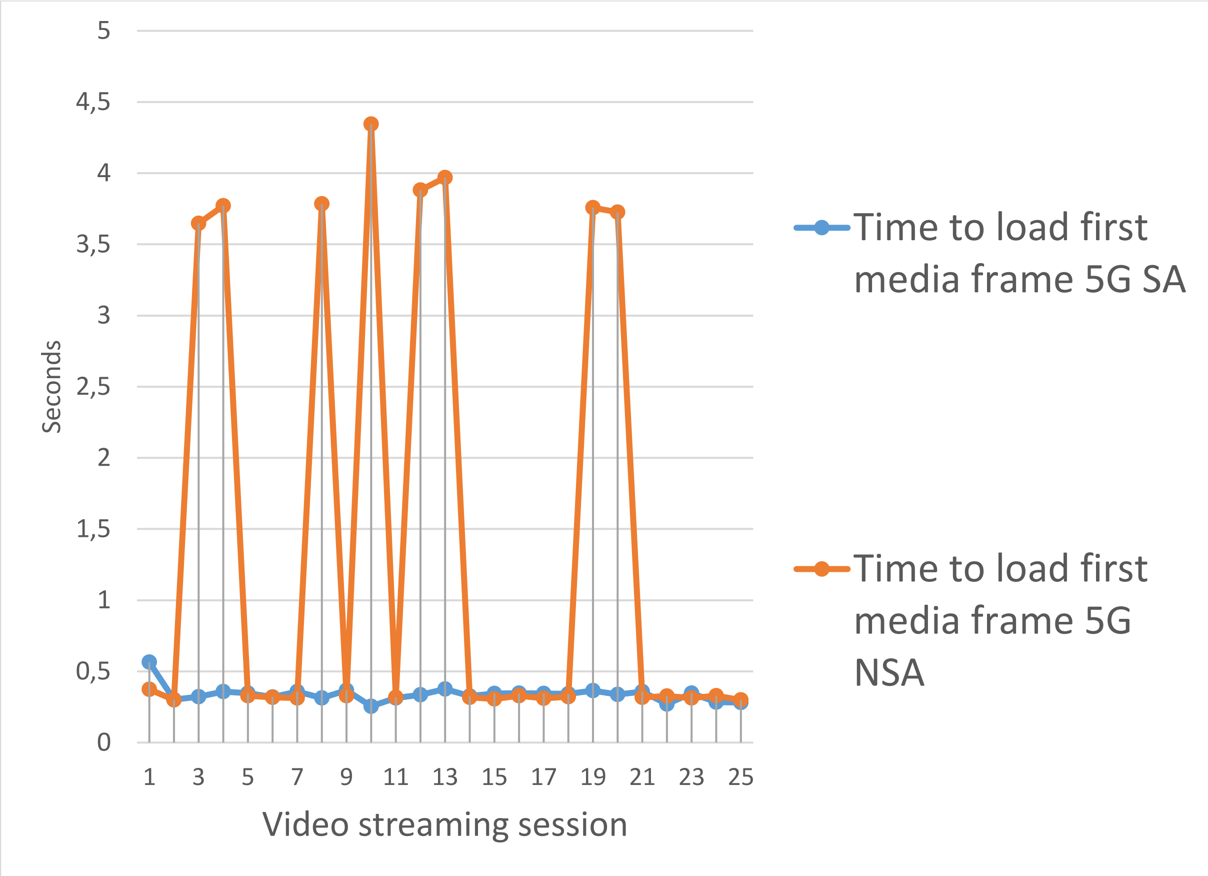

Figure 4-9 5G SA Time to load first media frame

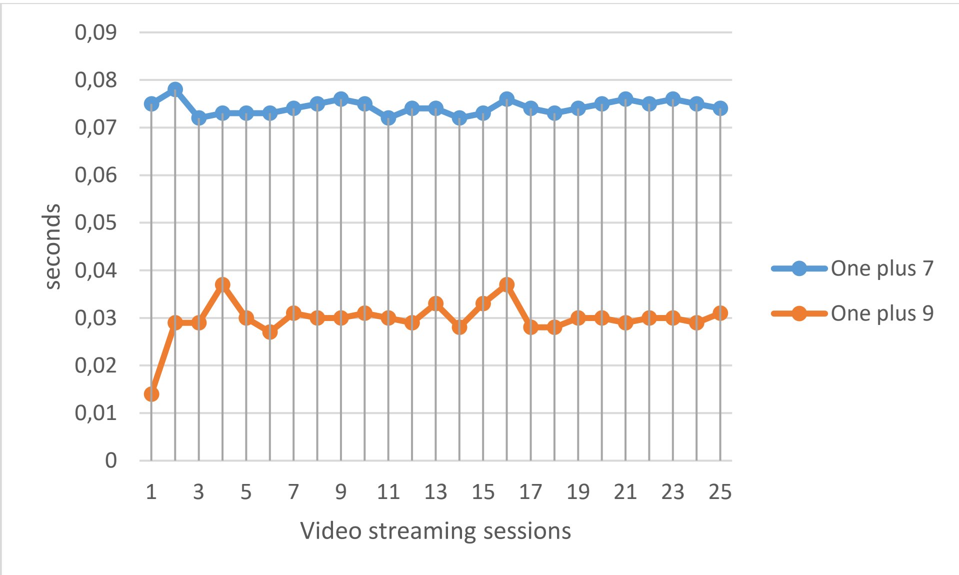

Figure 4-10 Time to open video streaming application.

The second test case measures the resolution of the video. Figure 4-11 shows the results obtained for the current deployment under test, 5G SA, and 5G NSA. The results are very similar.

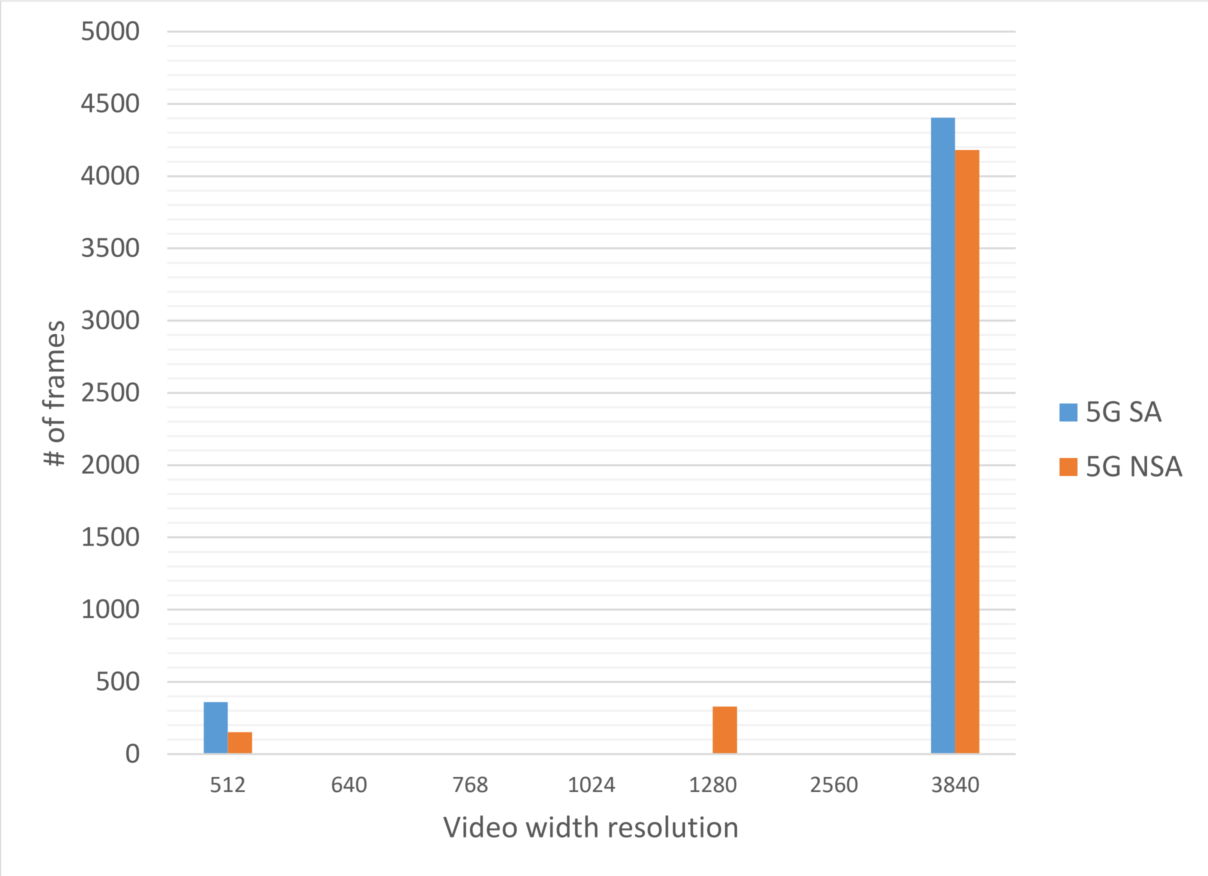

Figure 4-11 5G SA vs 5G NSA Video Resolution

4.3.2. Trial: 5G video mobile command center

The University of Málaga in collaboration with the Malaga Local Police and Telefónica I+D, deployed a 5G mobile command center to detect risk situations in mass events. The event took place on the 30th of October 2020 during the Magna procession that gathered several thousand people in the streets of Malaga.

The 5G deployment available at the city center was used to carry out this pilot. In the city there is a network of fixed cameras connected via fiber connection to a command center located at the outskirts of the city. Furthermore, for this trial the fixed cameras were connected to a 5G mobile command center deployed in a hotel located in one of the main streets of the city. Additionally, six 5G cameras were deployed to reach areas where there are not fixed cameras, while several local police agents carried 5G mobiles. The video recorded by the fixed and mobile cameras were sent to the 5G mobile command center.

For comparison purpose a command center connected by fiber was also deployed as shown in Figure 4-12.The frame per seconds of the 5G mobile cameras and the fixed cameras were the same as shown in Figure 4-13 5G mobile cameras (right screen) provide the same video frame rate with the fixed cameras (left screen).

5G radio parameters were also monitored during the trial as shown in Figure 4-14.Figure 4-13.

Figure 4-12 Command center with access to fix cameras via fiber connection (left screen), 5G command center with access to the fixed and the mobile cameras via 5G connection (right screen).

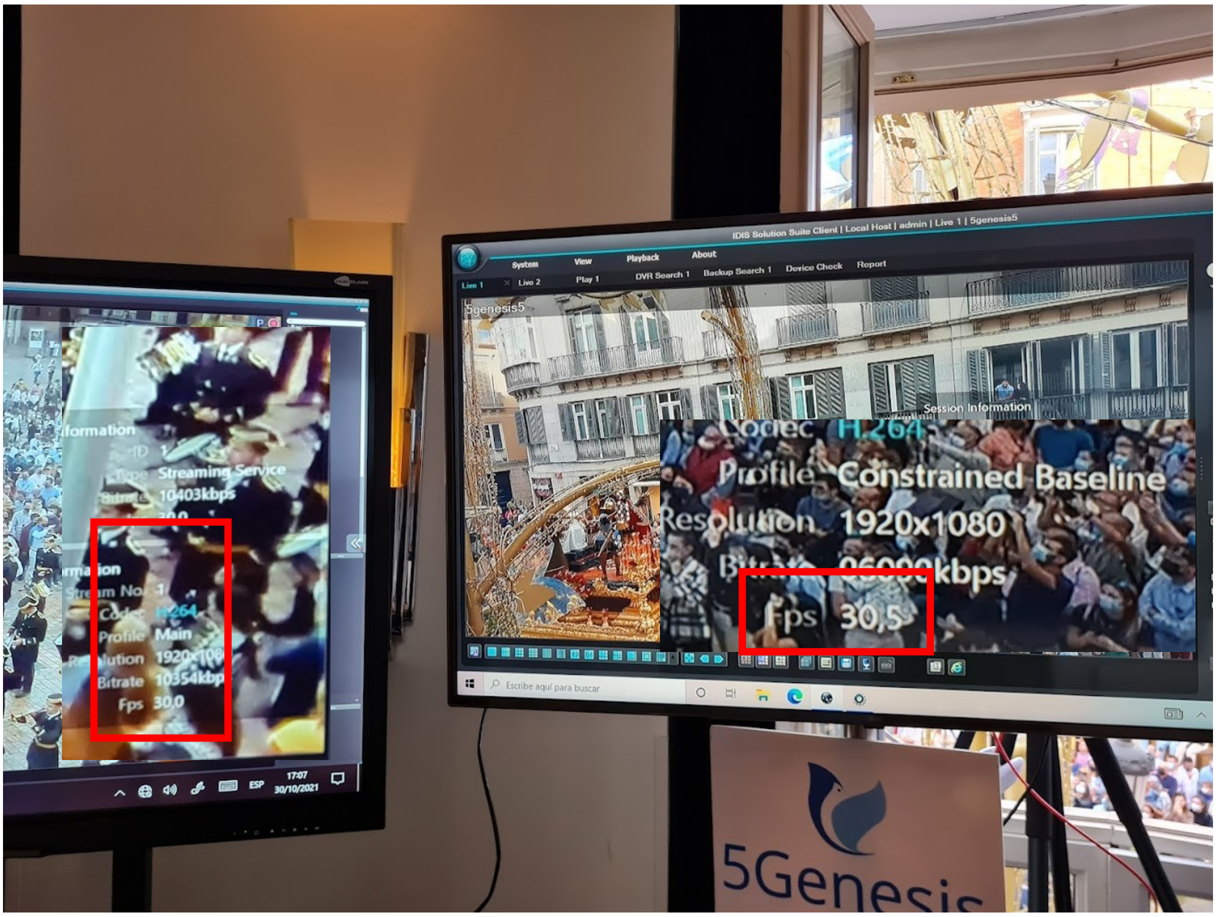

Figure 4-13 5G mobile cameras (right screen) provide the same video frame rate with the fixed cameras (left screen).

5G radio parameters were also monitored during the trial as shown in Figure 4-14.CQI and SNR values correspond with good quality conditions. For the worst radio conditions, the BLER is 8%, which is below the 10% value adopted as the facto threshold for the dimensioning of mobile networks.

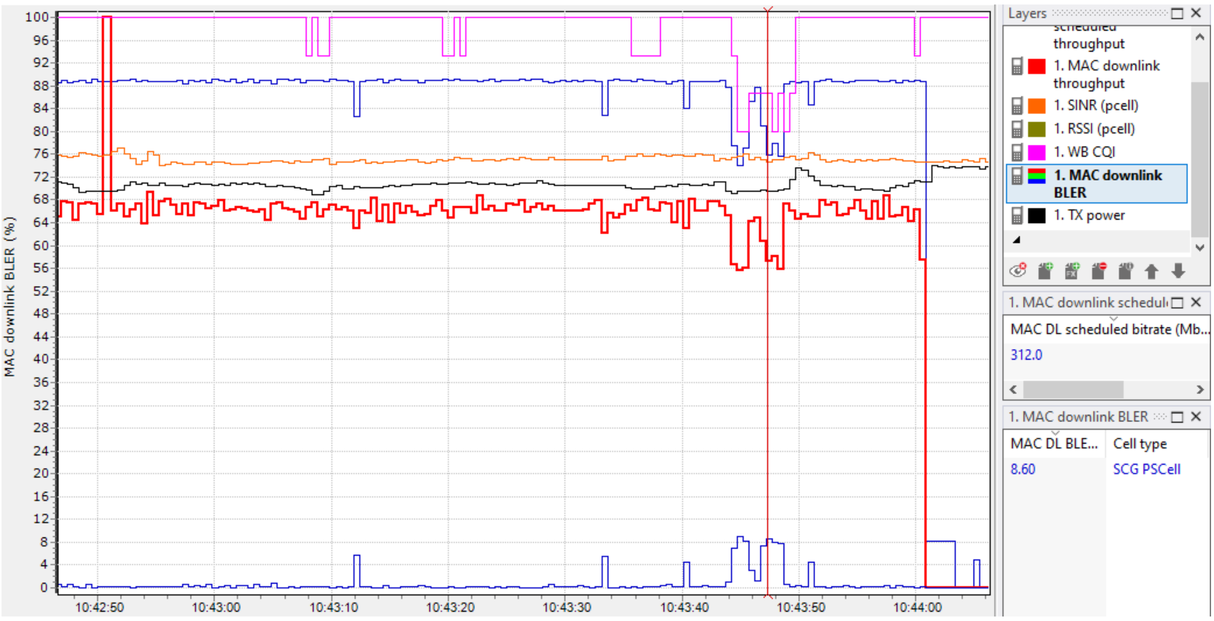

Figure 4-14 Radio measurements monitoring

Figure 4-15 shows traffic measurements captured during the reception of the video in the 5G devices. In particular, we analyze the round trip time using the acknowledgment packages exchanged during the TCP connections established for streaming the video from the mobile and fixed cameras. As depicted in Figure 4-15, the average for the round trip time is below 10 ms which is lower that the valued obtained in LTE and lower that the 30-100 ms that typically have current live streams.

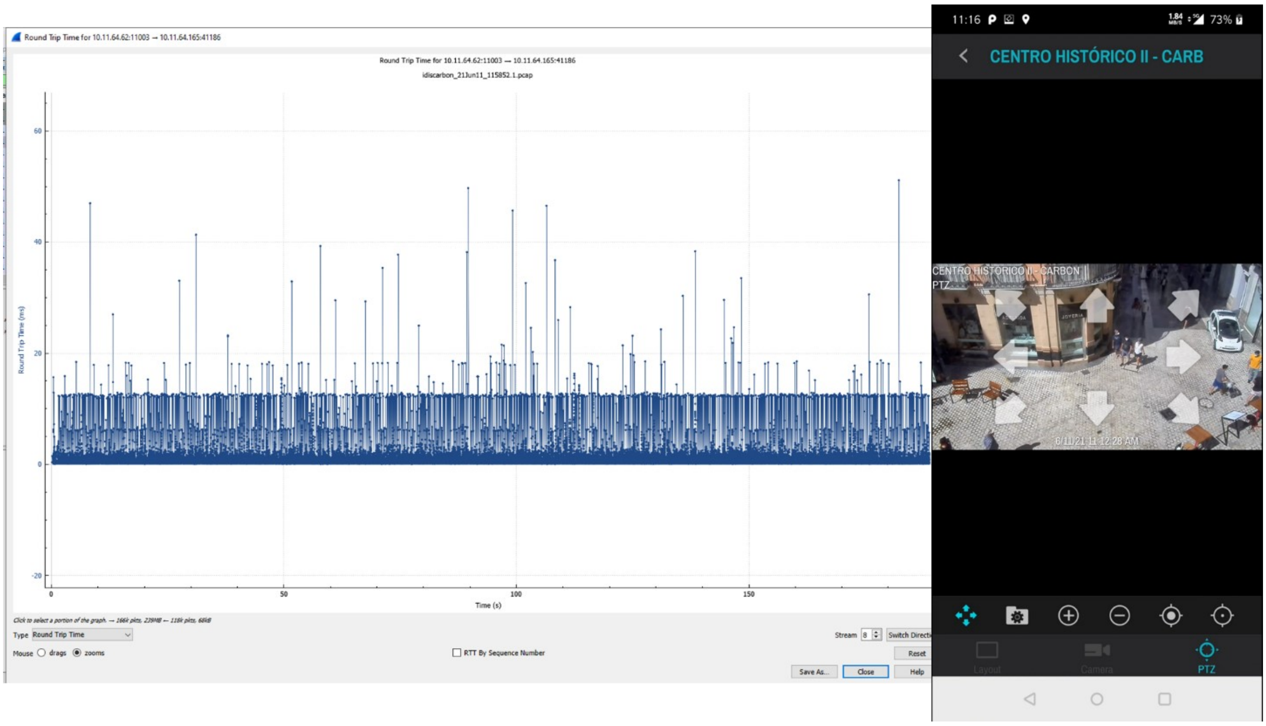

Figure 4-15 Round-trip-time measurements based on ACKs during video reception

4.4. UC#2: Multimedia Mission Critical Services – Measurements and results

In this third cycle, the MCPTT experiments have been performed using the final setup of the Málaga Platform, which makes use of 5G SA. This way the MCPTT experiments executed in this cycle serve to compare the performance of 5G NSA and the recently deployed 5G SA. The 5G SA setup used for the experiments includes the recent Athonet's 5GCore SA and UMA's Nokia Airscale gNB with the corresponding software version supporting SA function.

For Málaga Platform's use case 2 the KPI measured has been the same as in the previous cycle for the Airbus MCS service, MCPTT access time. MCPTT End-toEnd access time has not been measured for this use case due to the limited capability of the MCS service to provide such measurement. We have measured an average value for the MCPTT access time KPI of 27,775 ms, a value slightly inferior to that measured with 5G NSA in the second cycle (29,01 ms for 5G NSA with Athonet EPC, and 28,10 ms for 5G NSA with Polaris NetTest EPC).

The improvement in the delay is not significant, but it is important to consider that this is the first iteration of 5G SA equipment in the platform, both for the core and the RAN sides, and its configuration has not been tailored to minimize delay. In the future, the possibility of using slices for the URLLC use cases will provide a configuration and performance oriented to minimizing this kind of KPIs related to the delay.

4.4.1. Trial: Mission Critical services showcase

This trial, performed on 21st and 22nd November, focused on the two mission critical services integrated in the Málaga Platform. Both MCS services were demonstrated both at the Ada Byron Research building and Málaga city center in collaboration with Málaga Police, Nemergent and Airbus. The most recent version of the MCS services were shown to the Police to gather their comments and feedback through the replication of a real-life scenario. MCPTT private and group calls, as well as MCVideo calls were demonstrated successfully.

An additional functionality shown using Nemergent's MCS service was MCX-DMR interoperability, an important last-minute addition to this MCS service integrated in the platform.

4.5. UC#3: Edge-based Mission Critical Services – Measurements and results

For use case 3 the situation is similar to that described previously for use case 2. The MCPTT experiments for this use case have been executed using the 5G SA setup as well. In this case the KPIs measured are also the same as for last cycle, including MCPTT access time and MCPTT End-to-End access time.

For MCPTT access time the average value measured in this experiment has been of 23,509 ms, while for the last cycle we obtained 17,68 ms (with Athonet EPC) and 16,72 ms (with Polaris NetTest EPC).

For MCPTT End-to-End access time the average value measured in this experiment has been of 145,288 ms, in comparison with the value obtained last cycle of 138,15 ms (with Athonet EPC) and 128,24 (with Polaris NetTest EPC).

An important remark is that, although the latency values have been measured for different core solutions, the variations of small milliseconds are not related with the specific core solution used in every experiment, but with the normal variations of the network segments related to the RAN and the phone performance while running the corresponding MCS client app. The core solutions introduce a delay in the whole network of approximately 50 us, confirmed after massive campaigns of measurements, and thus are not the reason of the slight latency variations, as opposed of what could be understood while reading the results gathered from the related experimentation reported in D6.2 and this document D6.3.

Both KPIs have clearly increased by some milliseconds, which is not an important difference but makes clear that the 5G SA setup does not imply an improvement in the delay as it is now. Even with this slight increase, the values measured are still way under the ones measured last cycle for 4G, and hence they fit perfectly under the defined thresholds for the MCPTT service correct operation. As said in the previous section for use case 2, the current setup will need to progress to integrate new versions of the software including new features such as the URLLC slice configuration and other enhancements for 5G SA, which currently is in its early state and may not provide a huge difference in comparison to 5G NSA for some metrics.

4.5.1. Local Breakout setup at Edge Data Center

For this last testing cycle we have also prepared a Local Breakout setup at the Málaga Platform, which consists on the deployment of the SGW and PGW (dataplane components) of the EPC at the Edge Data Center, while the rest of the components remain deployed in the Main Data Center. For this setup, we have used the Polaris NetTest Rel. 15 EPC of the Málaga Platform.

The KPIs measured are the same as explained before for UC#3: MCPTT access time and MCPTT End-to-End access time.

For MCPTT access time the average value measured in this experiment has been of 26,304 ms, while for MCPTT End-to-End access time the average value measured in this experiment has been of 151,680 ms.

There hasn’t been an improvement with the use of this local breakout scenario using the Edge platform. The reason that explains this behaviour is that currently in the Málaga Platform the Edge Data Center and the Main Data Center are both equally close to the end user, and hence the only variation in the results is introduced by the processing times of the different Data Centers and the virtualization technologies used at each deployment for the EPC. In a real-life scenario, the Edge platform would be closer to the end user and hence the delay would be reduced since the dataplane messaging would not need to travel to the central deployment at the Main Data Center, located further away from the users.

4.6. Millimeter wave testing

The millimeter wave testing presented in this report were conducted at the setup shown in Figure 4-16 which is based on the 5G network emulator from Keysight. When operating in the mmW frequency range (FR2) 5G NR mobile devices are tested over the air, which requires a very fine-grained control of the channel for reliable experimentation. The Multi Probe Anechoic Chamber (MPAC) shown on the left in the figure provides the required control and accuracy, as it enables precise positioning. The mmW radio head is connected to polarized antenna horns inside the MPAC chamber.

Figure 4-16 Millimeter wave setup

Figure 4-17 shows the interface offered by the 5G emulator to configure the 5G emulator. At the top of the figure, we can see that there are four 5G mmWs antennas configured, which allows the aggregation of four 5G carriers. Each 5G carrier uses 100 MHz of associated channel bandwidth, accounting for a total of 400MHz of 5G NR spectrum. A 2x2 MIMO with 64QAM modulation, in non-impaired conditions, enables the selected scheduling configuration to reach 550 Mbps per carrier. In total, the maximum theoretical 5G throughput with this exceeds 2200 Mbps at physical layer. The iPerf tests performed within this configuration enabled us to reach 2 Gbps at the application level, as shown in upper Figure 4-18. The bottom of the figure shows the results of an ICMP test executed in the same scenario to quantify the end to end latency of this configuration. An average ICMP Round Trip Time (RTT) of 4.5 ms is obtained, with fluctuations between 3 and 6 ms. Comparing this value with the values obtained in the sub-6GHz scenarios, we reach a latency decrease of more that 50%.

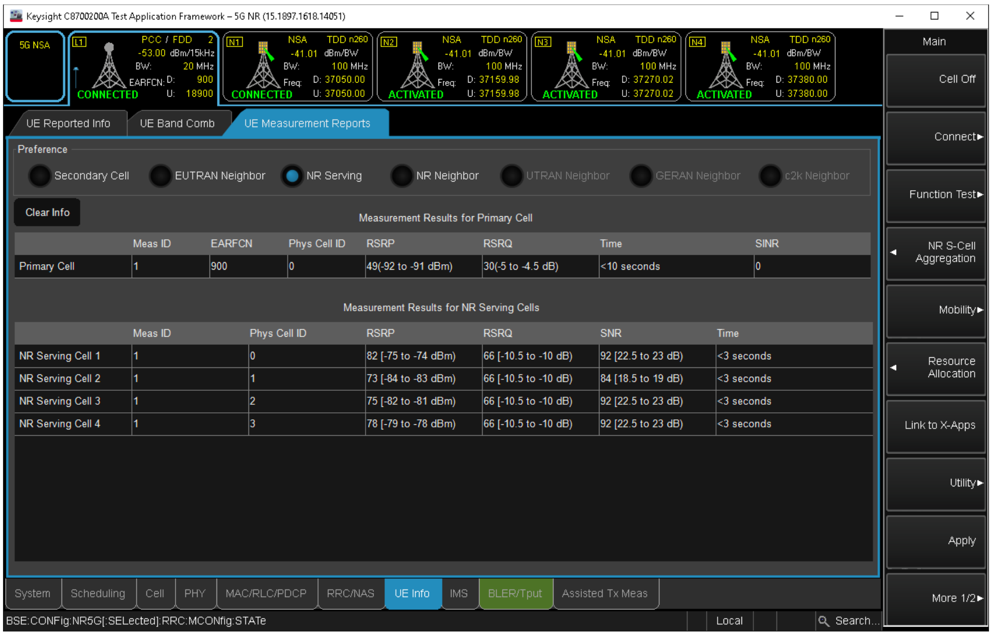

Figure 4-17 5G emulator test application

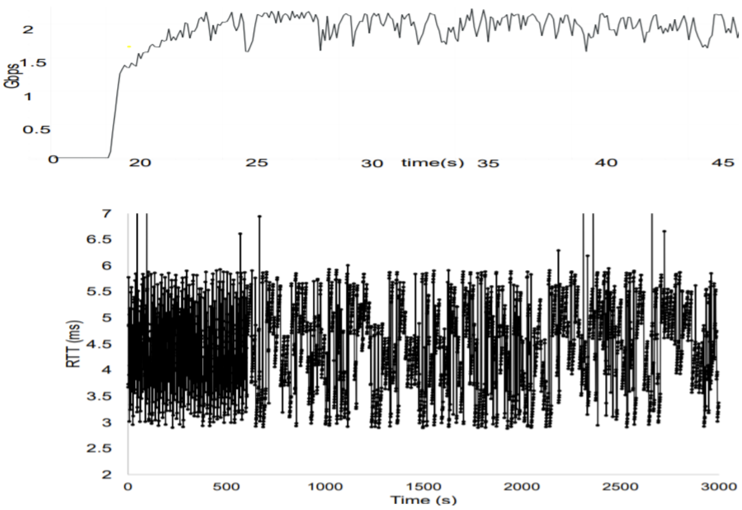

Figure 4-18 IP and RTT tests results in the mmW setup.

The test cases defined in the project could not be applied in this set-up because the mmWave mobile devices did not have enough energy autonomy to execute the 25 iterations specified in the test cases. The duration of the test cases is around one hour and half per KPI.

4.7. Summary and conclusions

At the Malaga platform we have validated the 5G NSA and 5G SA deployments done as part of the project. Both deployments are based on components and features belonging to 3GPP Release 14 and Release 15. These two releases are focused on Mobile Broadband (MBB) use cases. For each one of the scenarios applied we have reached throughput values that were closed to the theorical values.

Regarding latency values they are in the order of the 10 ms for FR1 frequencies and around 5 ms for FR2. To improve these values 3GPP features specified in Release 16, which is focuses on ultra-reliable low latency communication (URLLC) use case, should be applied. Commercial Release 16 equipment are expected by the year 2022.

At this moment there are not commercial equipment for offering location services in 5G networks, we have done the exercise to calculate location accuracy in 4G networks.

Reliability, capacity and speed have been also tested and the values obtained are aligned with the one expected for the configurations applied.

Finally, we have demonstrated that the performance of MCPTT and video surveillance services are improved when using 5G networks.

5. LIMASSOL PLATFORM EXPERIMENTS

5.1. Overview

The goal of the third phase of experiments in Limassol platform has been to:

- Test 5G SA (Stand Alone) scenarios
- Verify the functionality of the open-source Open5GS core
- Test full 5G SA satellite backhauling
- Assess the value of 5GC local break-out and edge computing in a converged 5G-satcom setup
- Implement the two use cases (Maritime communications and rural applications) and perform measurements in the field.

Table 5-1 lists the KPIs evaluated in the three trial phases and summarizes the kind of evaluation measurements conducted. The focus has not been to expand to new KPIs (most of them were already evaluated during Phase 2) but to investigate alternative network configurations and new features.

Table 5-1 5G KPIs evaluated at the Limassol Platform

KPI to be evaluated at the Limassol Platform according to DoA	Evaluated in phase	Comment
Ubiquity	Phase 2	(see comment below)
Latency	Phase 1, 2, 3	Round-trip time (RTT)
Reliability	Phase 2	(see comment below)
Service Creation Time	Phase 2	
Additional 5G KPIs evaluated at the Limassol Platform		
RTT	Phase 1, 2, 3	
Throughput	Phase 1, 2, 3	

Comment: The ubiquity and reliability enhancements achieved by the dual-backhaul mechanism of the Limassol platform were shown in D6.2, although these KPIs were not directly measured in the strict sense.

Figure 5-1 depicts the physical topology of the Limassol platform, as it has been implemented for the Phase 3 experimentation campaign. This essentially corresponds to the configuration fully described in Deliverable D4.9[5]. The satellite edge + RAN segment of the infrastructure was detached and integrated in a mobile 5G hotspot used for the lab and field trials.

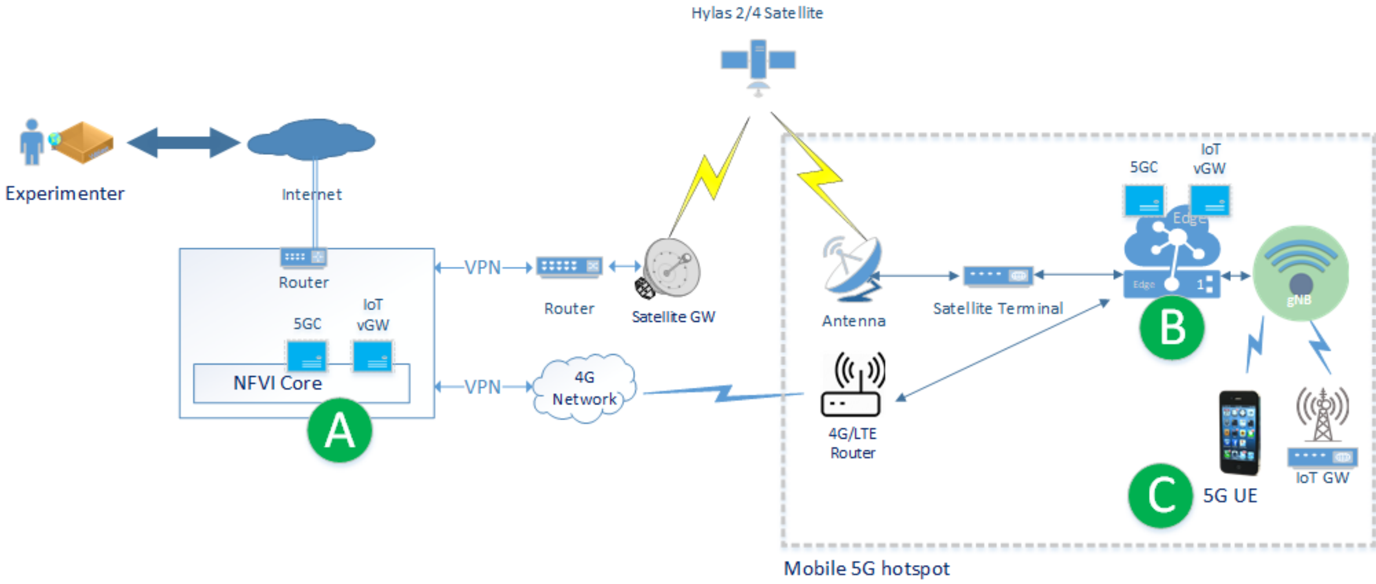

Figure 5-1. Physical topology of Limassol platform implemented for Phase 3 experimentation and measurement points

Figure 5-1 also displays the main reference points used for the measurements. All tests were carried out between (or at) these points.

- Reference point A: At the platform core compute infrastructure.
- Reference point B: At the satellite edge compute infrastructure of the mobile 5G hotspot.
- Reference point C: At the 5G UE.

5.2. Generic tests – Measurements and results

The lab tests focused on the 5G SA configuration using as core the open-source Open5GS platform (implementing most functions of a Rel.16 5GC) and the Amarisoft Callbox Classic as RAN. The field tests (Sec. 5.3-5.4) used the Amarisoft 5G Core instead.

Open5GS was deployed on Ubuntu 20.04 LTS in a VM with 2 vCPUs and 4GB of RAM.

Table 5-2 shows the NR configuration used in the tests.

Table 5-2 5G NR configuration in the Limassol platform used for Phase 3 measurements

Parameter	Value
3GPP release	Rel. 15
Bandwidth	50 MHz

Downlink MIMO config	2x2
Duplex mode	TDD
Network mode	Stand-alone (Option 2)
Band	n78 (3490 MHz)
Service type	eMBB

For this set of tests, we used as UE a Raspberry Pi 4 with a Waveshare SIM8200EA-M2 5G Hat, running on the Qualcomm Snapdragon X55 chipset.

All measurements were done using the tools provided by the Open5GENESIS suite (portal, ELCM, monitoring, analytics) and followed the test cases defined in the test case companion document.

The detailed results of the experiments described below are to be found in Annex, Sec. 7.3.1.

5.2.1. 5G SA setup with Open5GS core co-located with the RAN

The first configuration includes the co-location of 5GC and RAN (i.e. the 5GC is deployed as a whole in the edge computing infrastructure). Measurements are done between the edge node (point B) and UE (point C).

5.2.1.1. Downlink throughput

The throughput is measured using iperf3 agents at the edge node and the UE. The DL throughput measurement is on average 111.74 Mbps.

There is indeed some significant deviation from the maximum DL throughput theoretically expected with the given RAN configuration (see table above, ~300 Mbps). This deviation is attributed to limitation of the Open5GS core, especially running in a VM. Increasing the VM resources did not yield an observable increase in performance.

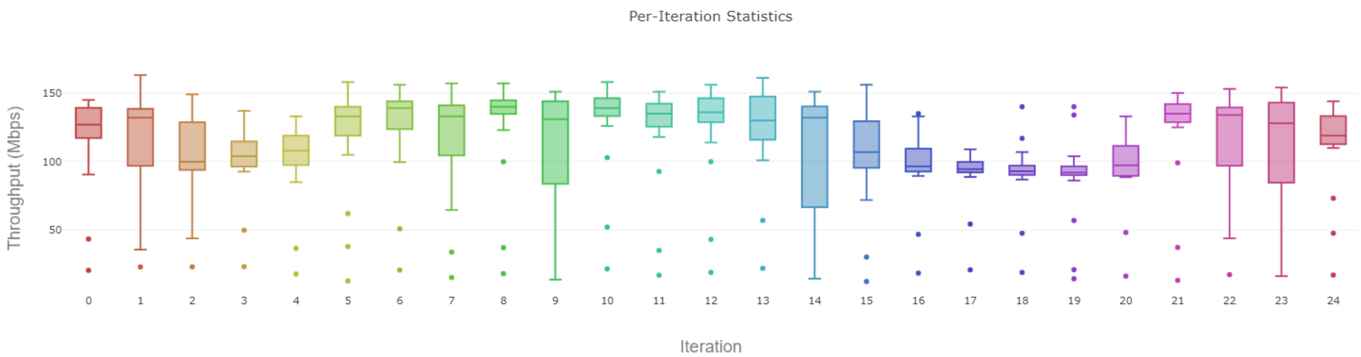

Figure 5-2 Local tests - DL throughput

5.2.1.2. Uplink throughput

The throughput is measured using iperf3 agents at the edge node and the UE. The UL throughput measurement is on average 35.94 Mbps.

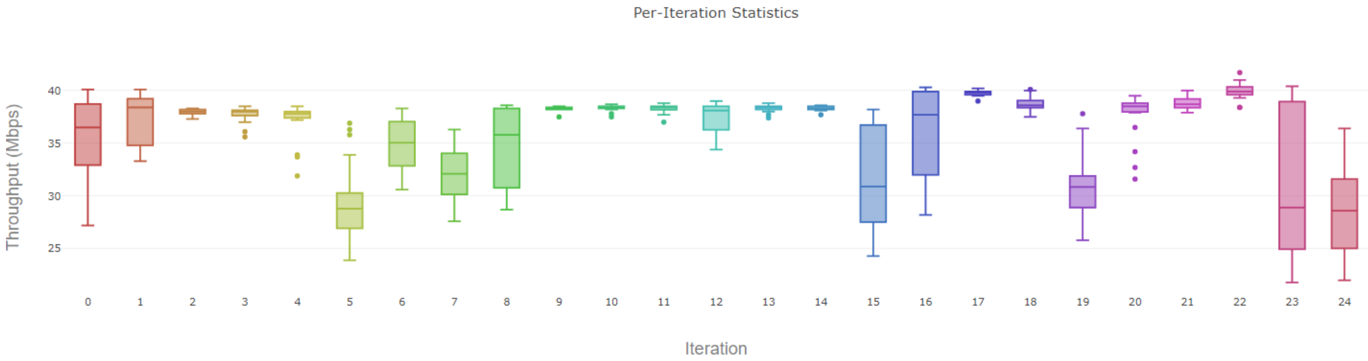

Figure 5-3 Local tests – UL throughput

5.2.1.3. RTT

The RTT is measured using ping agent at the edge node. The RTT mean value measured was 13.7 msec on average, with a median of 13.24. This is significantly lower than the RTT observed in both 4G and 5G NSA during the previous phases (around 30msec).

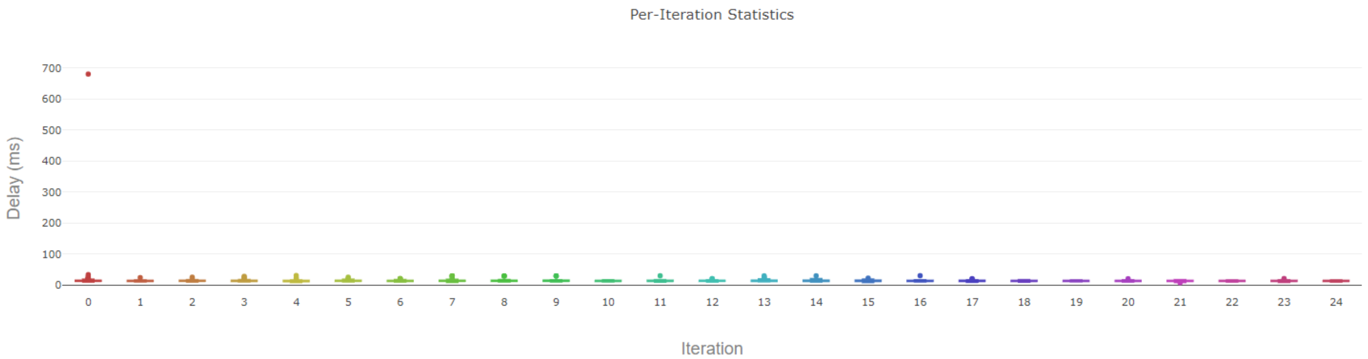

Figure 5-4 Local tests - RTT

5.2.2. 5G SA with Open5GS core fully backhauled over satellite

The second configuration tested involves the 5GC fully backhauled over the satellite link, i.e., with all 5GC functions deployed at the platform core (Limassol data centre) and the RAN at the satellite edge (mobile hot spot). This implies that the satellite link carries N1, N2 and N3 traffic. All user data also traverse the satellite link. This configuration is depicted in Figure 5-5, which shows the distribution of the 5G components in the different network segments.

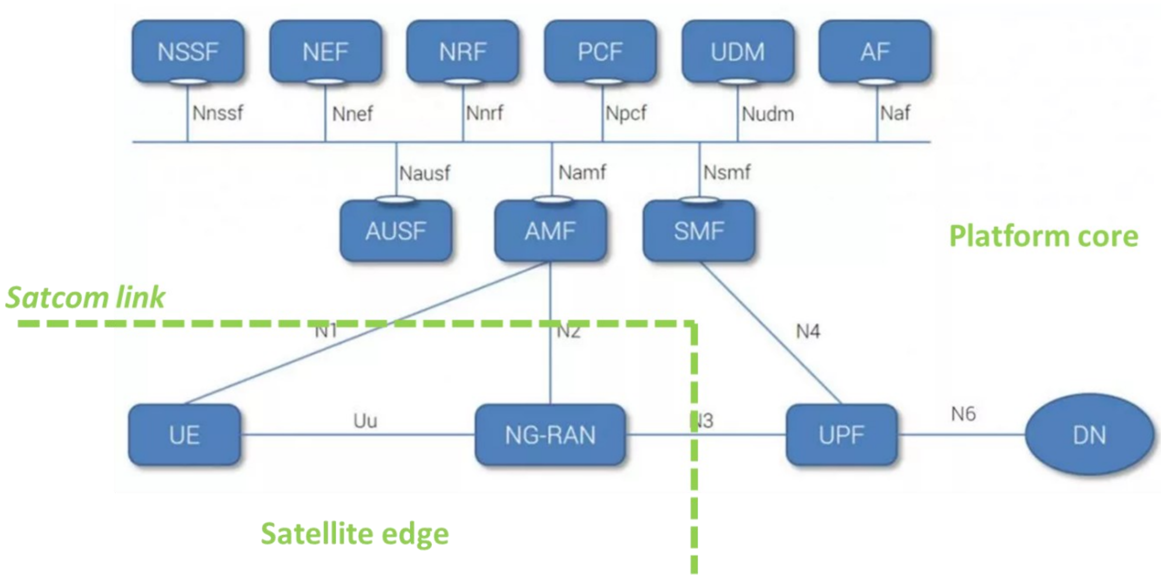

Figure 5-5 Distribution of 5GC functions in the full backhauling configuration

The measurements are done between the platform core (point A) and the UE (point C).

5.2.2.1. Downlink throughput

The throughput is measured using iperf3 agents at the core node and the UE. The DL throughput measurement was on average 2.22 Mbps, and its standard deviation was 1.11 Mbps. This is due to the satellite backhaul being limited to 5 Mbps as well as the overhead due to the double tunnelling over satellite (i.e. GTP over GRE, necessary for establishing reachability between network segments). It is expected that, moving to a carrier-grade setup with a satellite capacity of 100Mbps or more and a L2 service eliminating the need for GRE tunnelling, the performance would be adequate to match the throughput requirements of a gNB.

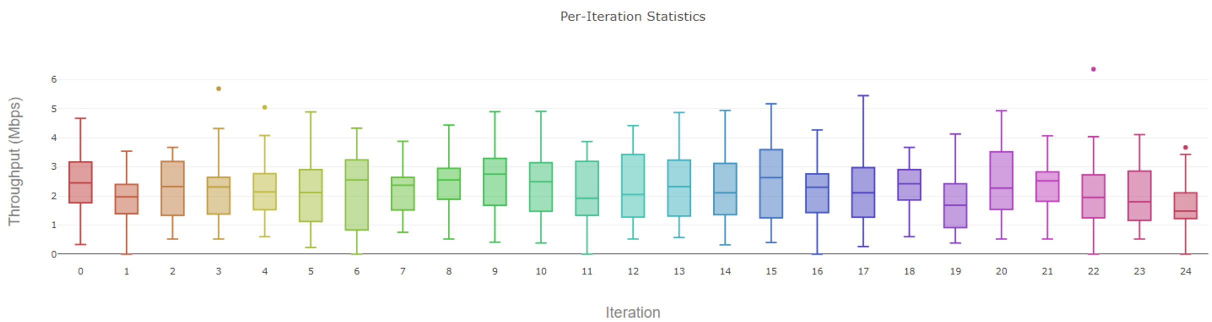

Figure 5-6 Full backhauling configuration – DL throughput

5.2.2.2. RTT

The RTT is measured using ping agents at the edge node and the UE. The RTT mean value was on average 838.34 msec, with a standard deviation of 38.48 msec, values which typically map to the latency introduced by the satellite link.

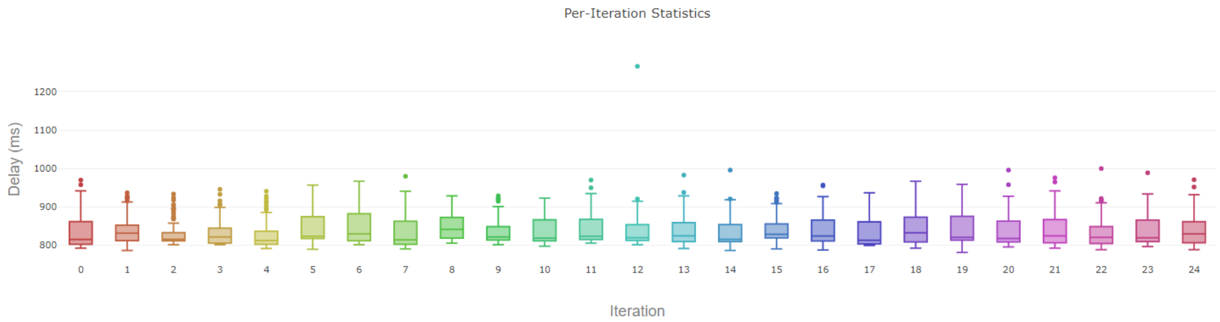

Figure 5-7 Full backhauling configuration - RTT

5.2.3. 5G SA with Open5GS core, satellite backhaul and local break-out (LBO)

The next step has been to implement a local break-out configuration. The UPF function was off-loaded to the satellite edge, while the rest of the 5GC functions (AMF, SMF etc.) remained behind the satellite link at the core of the platform. This implies that the satellite link carries N1, N3 and N4 control traffic, whereas use plane traffic (N2) is handled locally at the satellite edge and is routed (N6) either to the local server or via the satellite link for external services. This configuration is depicted in Figure 5-8, which shows the distribution of the 5G components in the different network segments.

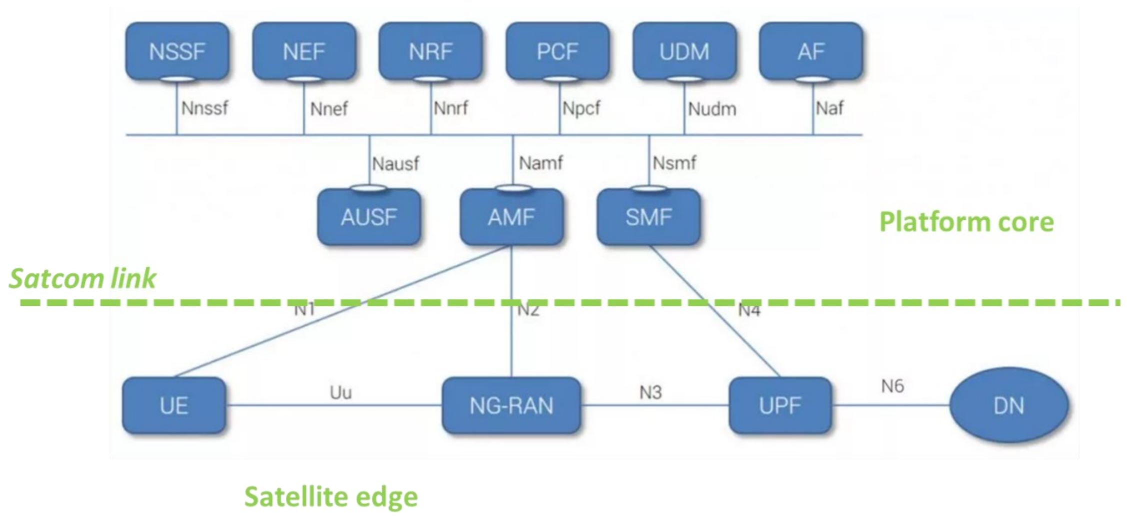

Figure 5-8 Distribution of 5GC functions in the satellite local break-out configuration

5.2.3.1. Downlink throughput LBO

The throughput is measured using iperf3 agents at the edge node (point B) and the UE (point C). As expected, the traffic destined for the local node is routed locally, yielding a considerable increase in throughput, whose mean increases from 2.2 to 30.3 Mbps.

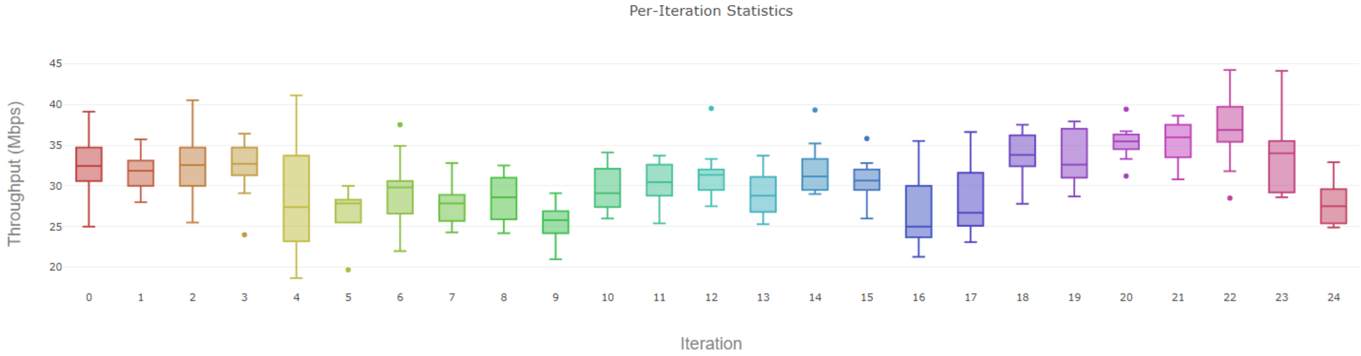

Figure 5-9 LBO at the satellite edge – DL throughput

5.2.3.2. UE RTT LBO

The RTT is measured using ping agent at the edge node. In the local break-out scenario, it drops from 848 msec in the full backhauling setup (sec. 5.2.2.2.) to approx. 25 msec.

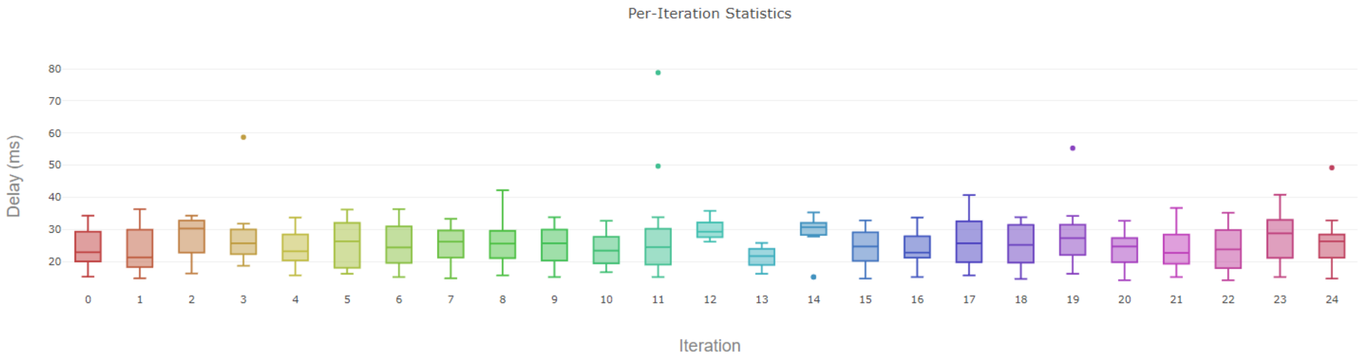

Figure 5-10 LBO at the satellite edge – RTT

Concluding, 5GENESIS has implemented one of the first globally known network configurations involving 5G SA with satellite local break-out and demonstrates the significant benefits derived from such an approach, namely the increased throughput and lower latency for edge-hosted services.

5.3. UC#1: 5G Maritime Communications – Measurements and results

The maritime communications field trials took place on 3/11/2021 on the MARAN GAS KALYMNOS LNG carrier, provided by MARAN (Figure 5-11). During the trials, the vessel was anchored in the Saronic gulf, close to Aegina island.

Figure 5-11 MARAN GAS KALYMNOS anchored, on the day of the field trials

The mobile 5G hotspot, enclosed in a flight rack case, was on-boarded and deployed in one of the control rooms of the ship (Figure 5-12). It was connected to the satellite backhaul, through which a VPN to the core of the platform in Limassol was initiated, enabling end-to-end measurements.

Figure 5-12 Installation of 5G test equipment on board

The detailed results of the experiments described below are to be found in Annex, Sec. 9.3.2.

5.3.1. Local measurements

The first set of measurements were based on local 5G SA deployment, i.e., using the Amarisoft core and RAN. The UE used was the Realme 7 5G.

The purpose was to establish a baseline and also to assess coverage within the ship. Each measurement corresponds to a different location of the UE.

Figure 5-13 Indoor coverage measurements

5.3.1.1. Throughput

The DL throughput is measured using iperf3 agents at the edge node and the UE, in this case the presented measurements are gathered from various point inside the vessel. In the locations close to the gNB, the throughput reaches the theoretical maximum of 270 Mbps, which gradually degrades as the UE <> gNB distance decreases. The last measurement is taken in the vessels' central staircase, three floors above the gNB location.

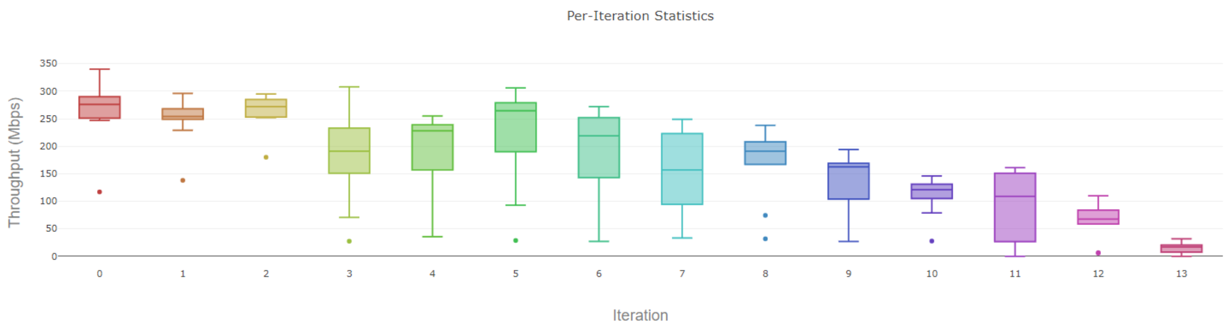

Figure 5-14 Indoor coverage measurements – DL throughput

5.3.2. End-to-end measurements

End-to-end measurements involve the satellite connection and are measured between the platform core (at the Limassol data centre) and the UE on board.

5.3.2.1. Core DC - UE downlink throughput

As seen in Figure 5-15, end-to-end DL throughput is restricted to ~1 Mbps, mostly limited by the satellite connection (nominal capacity 2Mbps), as well the overhead caused by the VPN (necessary to connect to the core of the platform over the Internet).

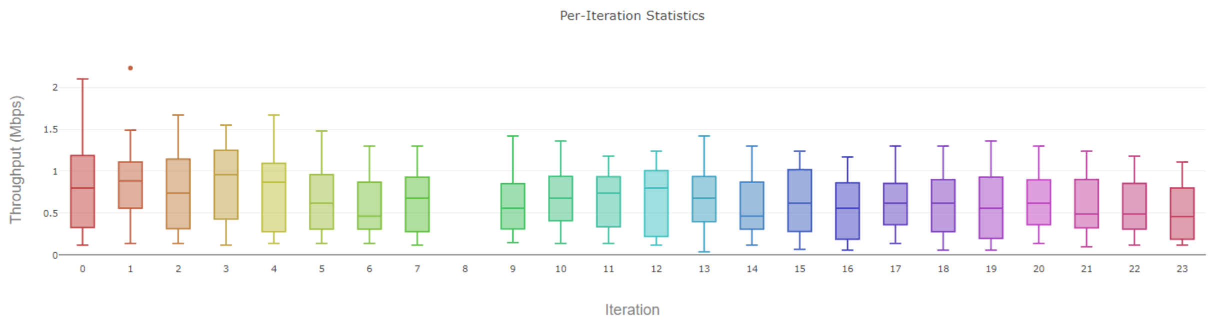

Figure 5-15 CoreDC – UE DL throughput

5.3.2.2. Core DC - UE RTT

Under the same configuration, the RTT is measured around the mean of 687 msec, attributed to the delay of the satellite link (which generally introduces ~600msec delay).

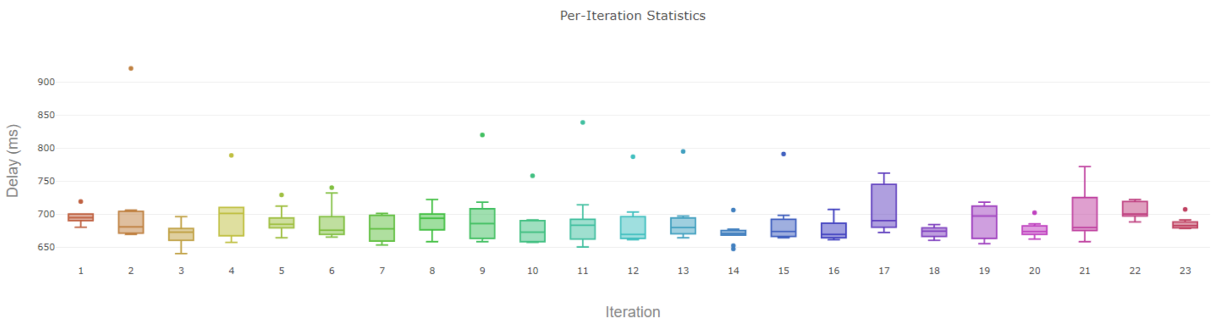

Figure 5-16 CoreDC -UE RTT

5.3.3. AR-enabled maintenance service

The AR maintenance service (see D4.9) is deployed at the edge and implements an algorithm for rust detection. The images are captured in real time by the AR glasses, sent over 5G to the edge service, which returns the image to be projected in the glasses with a bounding box around the rusty area. The application was demonstrated on board, involving the crew of the vessel.

The application delay measured -by processing the application logs- was about 0.4 sec, depending not that much on the network delay (which is less than 20 msec, as measured), but rather on the processing delay on the compute node, which resources were rather constrained in the specific configuration used.

Figure 5-17 Testing and assessment of the 5G AR service on board

5.4. UC#2: 5G Rural applications – Measurements and results

The rural use case demonstrated is based on a smart agriculture bespoke application. The application uses computer vision and Deep Learning-based analysis of drone-captured crop images for detection of unwanted weed in rural fields. More details about the app can be found in Deliverable D4.9[5].

The AI model of the application runs on Tensorflow and benefits from hardware acceleration capabilities (Neural Processing Unit – NPU) at the edge. The images are captured by a 5G UE (Raspberry Pi with 5G modem attached) mounted on a drone scanning the crop and are sent for analysis to the back-end application. To demonstrate the value brought by satellite edge computing, we consider two alternative scenarios (see Figure 5-18):

- In Scenario A (without edge computing), all 5G Core functions are deployed at the core data center in Limassol, behind the satellite link. The smart agriculture application is also deployed at the core. Traffic handled by the application has to traverse the satellite link.
- In Scenario B (with edge computing enabled), the data handling functions are deployed at the edge (mobile hotspot, in the field) to enable local break-out functionality. The smart agriculture application is also deployed at the edge.

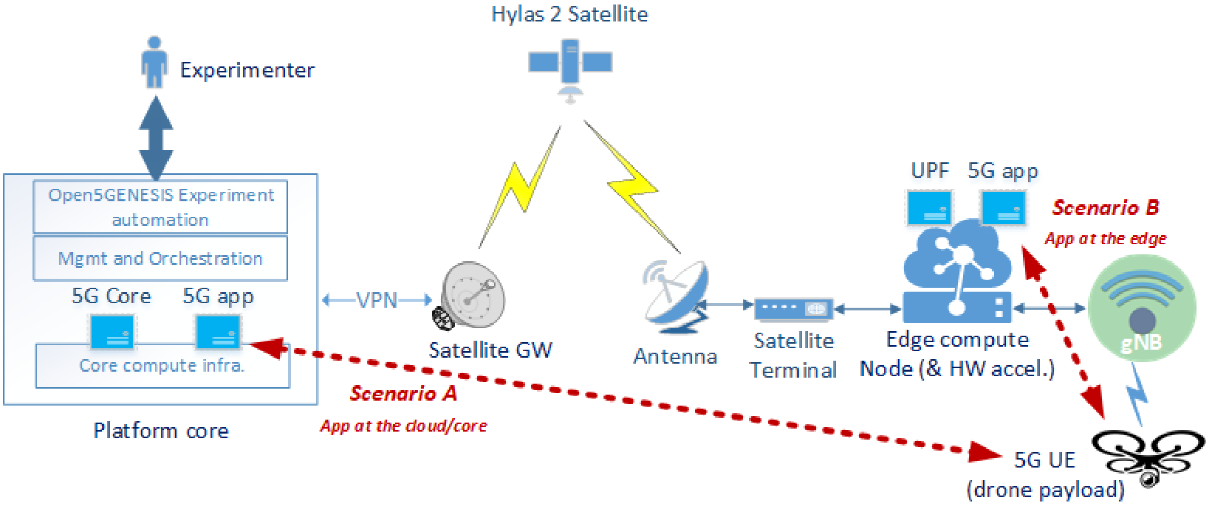

Figure 5-18 UC#2 demo setup topology

The trials were executed in the field, using the 5G mobile hotspot. For practical reasons, the application was trained to detect daisy flowers instead of weed. The 5G UE (Huawei P40 lite) was mounted on a DJI Phantom 2 drone.

Figure 5-19 5G gNB with edge compute node for the field trials

Figure 5-20 Drone with mounted UE

A second UE was used to visualize the detection results in a web interface, in real time (Figure 5-21).

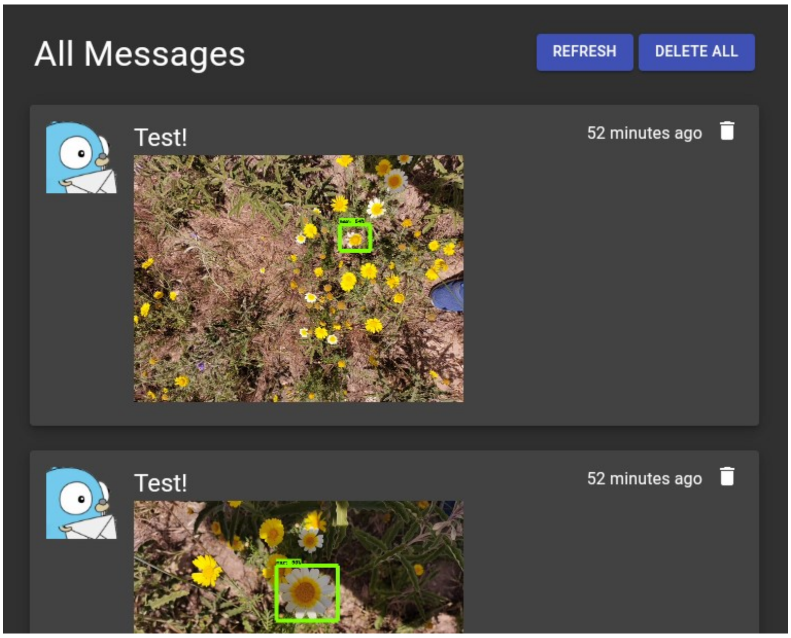

Figure 5-21 Web interface to display detection results

The main KPI to be measured is the application response time measured at the UE, i.e., the time interval from the submission of the captured image by the UE until the reception of the outcome of the image analysis. This is drastically reduced in the edge deployment scenario. Apart from the application response time, the image processing time (inference time needed for the AI process) is also captured.

For the planning and automation of experiments, as well as results analysis using the Open5GENESIS suite, a bespoke measurement agent at the UE was implemented. The agent, in turn, interfaces with the smart agriculture application using a REST API to collect application-level metrics.

The purpose of the experiment has been to show the added-value of edge computing in 5G infrastructures backhauled by satellite. The graph depicts the reduction of the delay when the application is deployed at the edge node. While processing delay remains the same around 0.1 sec, end-to-end delay can be 3-4 sec higher when data from the UE are sent towards the centralized processor at the platform core.

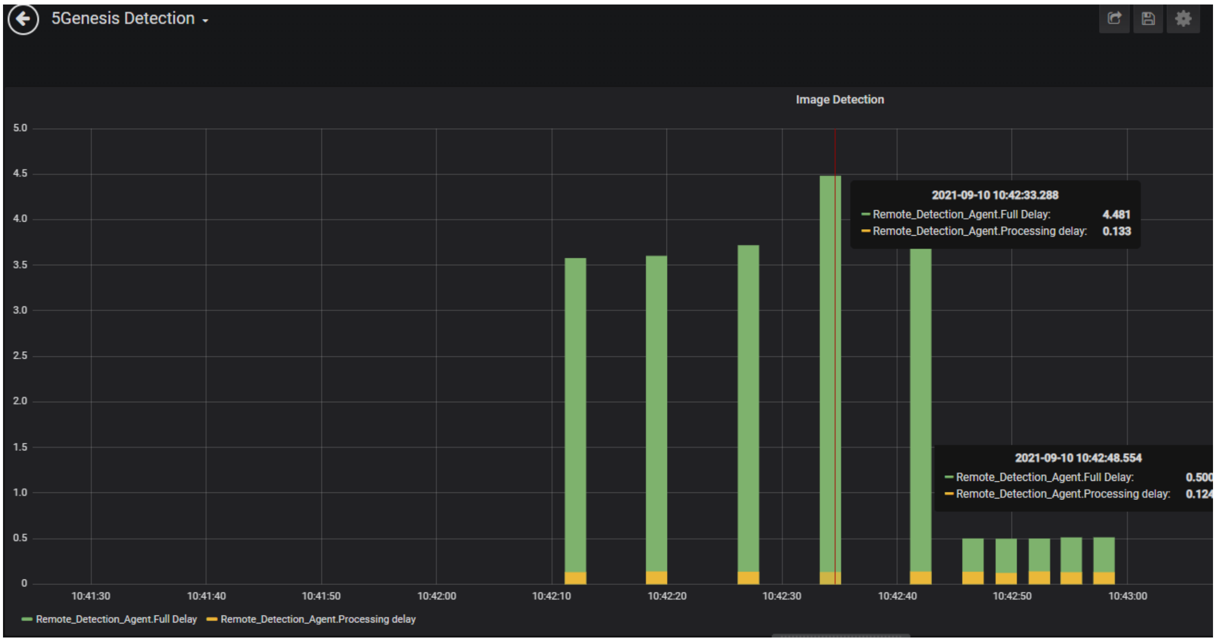

Figure 5-22 Image detection delay reduced using edge computing (Measurements 1-5: Scenario A, measurements 6-10: Scenario B)

5.5. Summary and conclusions

During Phase 3, the tests in the Limassol platform focused on various configurations related to satellite backhauling, as well as the various vertical applications.

5GENESIS implemented one of the first (to our knowledge) experimental setups with satellite backhauling fully integrated in a 5G stand-alone network configuration. The operation of the 5G Core with the N1, N2 and N3 interfaces operating over satellite, was verified. Also, the 5G SA LBO configuration was evaluated, with the UPF off-loaded to the satellite edge; this should be also considered among the key innovations of the project. 5G SA LBO configuration exhibited a >10x improvement in throughput and latency for edge applications.

The applications (maritime and rural), tailored for underserved areas, were well suited to showcase the capabilities of an integrated satcom/5G setup. Once again, the feasibility and value of edge computing to reduce latency and improve throughput -particularly for bandwidth-demanding applications such as real time video analytics- was demonstrated.

6. Surrey Platform Experiments

6.1. Overview

In the final Phase of the project, the Surrey Platform focused on the testing and evaluation of its main IoT use case based on a multi-RAT (5G and WiFi) environment with the use of concurrent slicing. As reported in [1], the main Surrey Platform use case was planned to be a public large-scale event taking place on the University of Surrey Stag Hill Campus in Guildford. The University of Surrey normally held regular large events, including graduation ceremonies and open days during which some 2000-4000 visitors were on campus as well. However, due to restrictions imposed during the COVID-19 pandemic, the University did not organize any such large-scale event for reasons of protection of health and safety. Therefore, the main use case of the Surrey Platform was moved indoors, and the initial plan to equip a subset of people on campus with devices that can access and make use of the Surrey platform was replaced by the deployment of IoT sensors inside the 5G Innovation Centre (5GIC) building.

The other use cases evaluated the 5G coverage in the outdoor sites of the UNIS campus, as well as the performance of the WiFi Slice Management Platform (WSMP) platform of the WiFi Slice Controller deployed by FON in the Surrey Platform [6]. The physical MONROE nodes that KAU deployed in the UNIS premises were also evaluated in terms of RTT and ping, while the operation of the integrated APEX policy engine is demonstrated through a use case that dynamically spins off slices based on the CPU usage of the UEs.

The KPIs evaluated in the experiments conducted within the context of the Surrey Platform are summarized in Table 6-1.

Table 6-1. KPIs evaluated in the Surrey platform experiments

KPI to be evaluated at the Surrey Platform according to DoA	Evaluated in Phase	Comment
Ubiquity (Coverage)	Phase 3	Through MCL-Based Coverage Test Case
Speed	Phase 3	For the main Surrey use case only stationary boards were used to conduct the trials. The MCL-based coverage measurements also included throughput measurements for UEs moving at a pedestrian speed, in-line with the expected CPE capabilities.
Latency (Packet Delay)	Phase 2, Phase 3	
Reliability (Packet Loss)	Phase 2, Phase 3	
User Density	Phase 3 (but with a lower number of IoT boards)	KPI fully evaluated in WSMP use case. Due to COVID-19 Restrictions, the IoT use case was moved to an indoor environment, with no participation of people. Trials considered 1 board/m^2 to depict performance in dense environments.

Energy Efficiency		Cost of required equipment required did not allow the evaluation of this KPI.
Additional 5G KPIs evaluated at the Surrey Platform		
PDCP-level Throughput	Phase 3	MCL-Based Coverage Test Case
Service Creation Time	Phase 3	WSMP use case
CPU Usage	Phase 3	APEX Policy Engine

6.2. Generic tests – Measurements and results

No generic tests were executed during Phase 3, as these were performed in Phase 2 and reported in D6.2 [8].

6.3. UC#1: Multi-RAT Support for Sensor Measurements – Measurements and results

Within the context of this use case, 30 Pycom/Pysense boards were deployed in the 5GIC building. The experiment consisted of reading sensor values and sending them –in JSON format– relying on various protocols, namely HTTP, MQTT and CoAP, through WiFi and 5G, with the use of a CPE device, to an IoT-vGW, developed by INFOLYSiS. For the purpose of this use case, two slices were spinned-off by the 5GENESIS Slice Manager and at the same time the 5GIC NOC system created these two slices at the 5GC using two separate APNs to support the end-to-end service. An architectural view of this use case is shown in Figure 6-1.

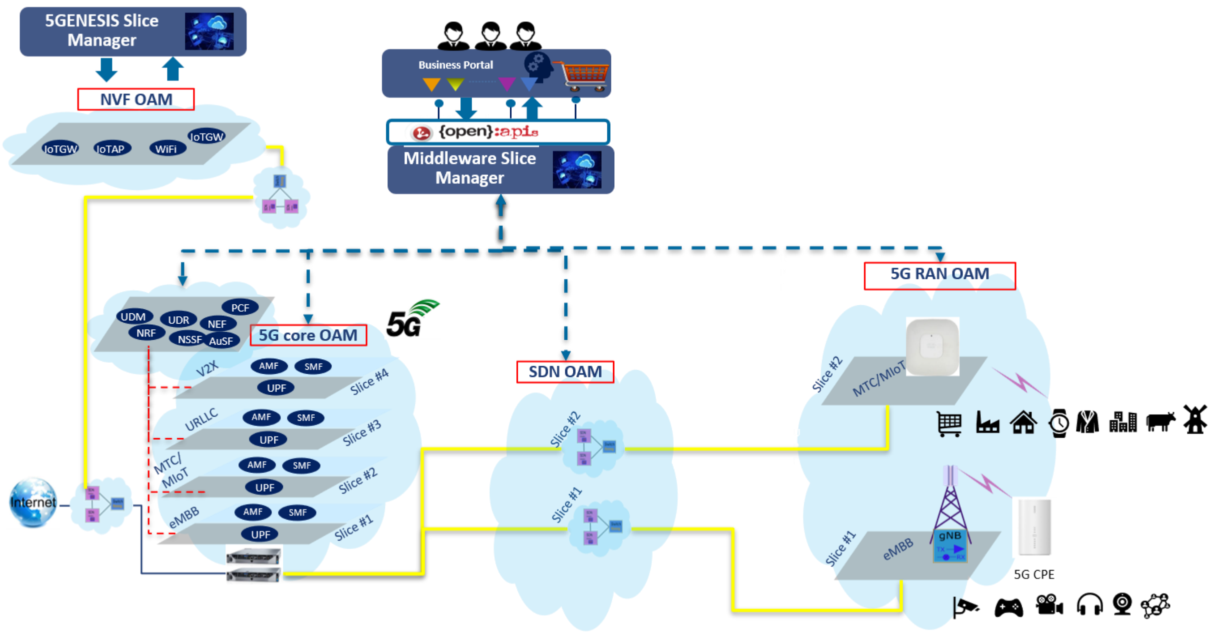

Figure 6-1. Main IoT Use Case in the Surrey Platform

For the purpose of conducting the IoT testing in the Surrey Platform, the network setup consists of the following components:

- Core: Rel.16 5G Core SA (5GIC developed in the UK)
- Control Plane: 4G RAN indoor and outdoor (Huawei commercial NSA support)
- User Plane: 5G indoor RAN (Huawei Commercial NSA support)
- WiFi: COTS indoor APs
- UE: 5G Huawei CPE Pro 2 (NSA-capable)
- Sensors: IoT Pycom/Pysense Boards

The tests conducted were used in order to evaluate the overall performance of the various servers dealing with incoming IoT data. Those servers have been implemented by INF at the Surrey Platform 5G test-bed side.

In those tests several protocols are tested at the same time over one or several radios, using single as well as multiple slicing. The general plan was the following:

1. HTTP POST, MQTT and CoAP, 10 boards each over WiFi (no dedicated slice), (TC_IoT_PacketDelayFULL_WIFI_SUR)
2. HTTP POST, MQTT and CoAP, 10 boards each over 5G (no dedicated slice); (TC_IoT_PacketDelayFULL_5G_SUR)
3. MQTT/5G and CoAP/WIFI, 15 boards each (two instances of the IoT-vGW, one for WiFi and one for 5G); (TC_IoT_PacketDelay_WIFI/5G_2SLICES_SUR)

In the following we call test probe the piece of code sitting at the server side that makes the various measurements and ultimately stores the data into an SQL database.

With those tests we are focusing on two aspects:

- Packet delay (Latency): the calculation is made comparing the time stamps at packet emission with time when the packet is received by the platform. Clocks at the board side are synchronized with google network time;
- Packet loss: As a complementary measure of Packet Delay, we count the number of received packets versus the number of packets sent by the IoT boards.

In order to facilitate the test we also include to all JSON packets a unique number (unique for the board) as meta data (starting with 1 when the 1st packet is sent and then incremented by one every time a new packet is sent).

The Slice configurations (used in Test Case#3) are shown in Table 6-2.

Table 6-2. IoT Use Case Slicing Configurations

Slice #	Protocol	Radio	IoTGW CPU	IoTGW RAM	IoTGW IP
1	CoAP	WiFi	2 x 2.8GHz	3.3 GB	10.5.31.95
2	MQTT	5G	2 x 2.8GHz	3.3 GB	10.5.31.97

The slices used the default slice descriptor provided by UNIS, which was configured to reference the INF Network Service descriptor, that was used to deploy IoT-GW instances with all the services running and ready to receive IoT data.

6.3.2. HTTP POST/MQTT/CoAP over WiFi (no slicing)

This set of measurements is the result of the test conducted with 30 boards in total, 3 protocols (CoAP, HTTP POST, MQTT) with 10 boards per protocol over WiFi.

The overall packet loss results for the protocols show 0% packet loss for HTTP and CoAP and roughly 4% for MQTT. Therefore, the reliability achieved is 100% for HTTP and CoAP and 96% for MQTT, respectively. It is worth noting here that for all tests performed, either during rehearsal or during the actual test campaign, with the 30 boards scattered all over the 1st floor of the UNIS 5GIC building, boards 14, 17 and 31 have been problematic, probably due to some hardware instability and/or power connector issues.

Delay-wise, MQTT and HTTP Post give similar results (respectively 142.3ms and 140.6ms), while CoAP shows 116.5ms delay. These results are in accordance with the respective results of the unitary tests executed in Phase 2 and reported in D6.2 [8], where the maximum average delay of the MQTT and HTTP Post protocols was 143ms and 140ms, respectively, while in the case of CoAP, the average delay was 103ms. It has to be noted that in Phase 2 the tests executed involved a single IoT board at a time, therefore the delay performance of the Phase 3 tests where 30 boards were used) was expected to be higher or at least equal compared to Phase 2.

6.3.3. HTTP POST/MQTT/CoAP over 5G (no slicing)

This second set of measurements is the result of the test conducted with 30 boards in total, 3 protocols (CoAP, HTTP POST, MQTT) with 10 boards per protocol over 5G.

The same test conducted over 5G shows slightly higher delays for all protocols. MQTT and HTTP Post show, once more, similar delays (respectively 155.8ms and 163.8ms) while CoAP again shows approximately 20% lower delay, at 130.1ms.

Figure 6-2 depicts the packet delay performance of all the three IoT protocols over WiFi and 5G without the use of slicing. As it can be seen, the overall packet delay is slightly lower using WiFi compared to using 5G. This is a result of the fact that, at the time of the experimentation, the UNIS 5G system was concurrently used by a number of 5G research projects, therefore, the IoT traffic of the 5GENESIS trial was facing congestion conditions. However, the absence of slicing in this experiment did not allow any guarantee regarding the provision of the necessary resources for this service from the 5G network. On the other hand, due to limited staff attendance in the 5GIC building (as a result of the work-from-home guidance for protection of the University staff from COVID-19), the University WiFi network was not experiencing high loads, allowing improved performance for the IoT traffic.

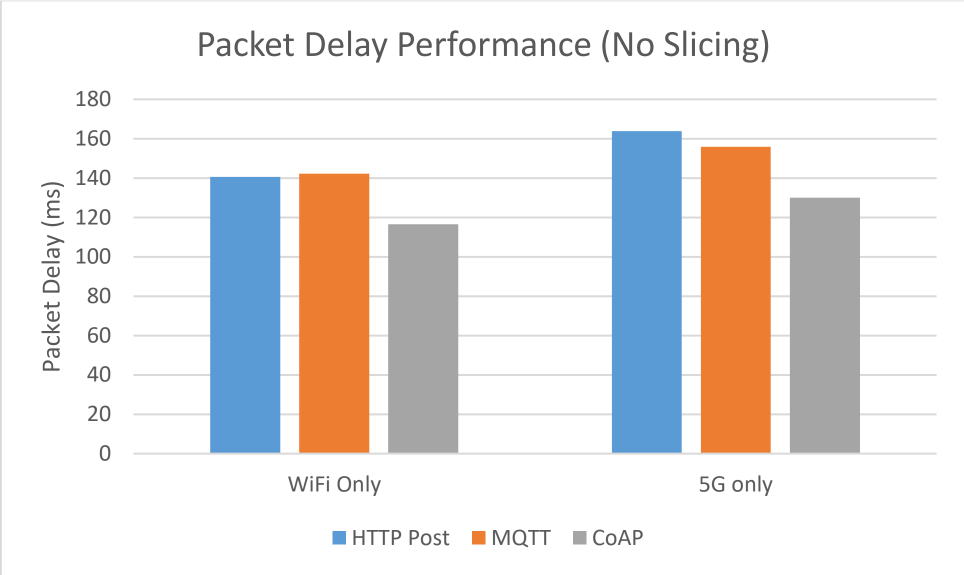

Figure 6-2. Packet delay performance in the case of no slicing

Packet loss-wise the same scenario appears with boards 17 and 31 (as a result of potential hardware issues), while all remaining boards are at 0% loss. Those two first tests show that, despite just a few glitches with 2 or 3 boards, the packet loss remains below 1%.

6.3.4. MQTT/CoAP over WiFi and 5G radios (2 slices, one per RAT)

The third set of measurements is the outcome of the test conducted with 30 boards in total, 2 protocols (CoAP, MQTT) with 14 boards per protocol relying on 2 slices: 1) CoAP over WiFi and 2) MQTT over 5G.

This latest test shows 0% packet loss for both slices (WiFi and 5G).

With slicing, we can see that the delay is considerably higher with WiFi (almost double) than with 5G (respectively CoAP and MQTT). This is the case even though CoAP always outperformed MQTT and WiFi outperformed 5G (in the no slice configuration). With a double-slice configuration, it appears that the “winner” couple WiFi/CoAP is then outperformed by 5G/MQTT. Figure 6-3 depicts the packet delay performance of the IoT protocols in the case of concurrent slicing. From this figure we can see the considerable performance improvement that can be achieved with the use of 5G slicing. Contrary to the case of no slicing (Figure 6-2) we can see that with the use of 5G slicing, the 5G system secured the necessary resources in order to accommodate the IoT traffic, despite the congestion from other research activities using the University 5G network, and achieved a significant reduction of the IoT packet delay compared to the traffic being served by the WiFi slice.

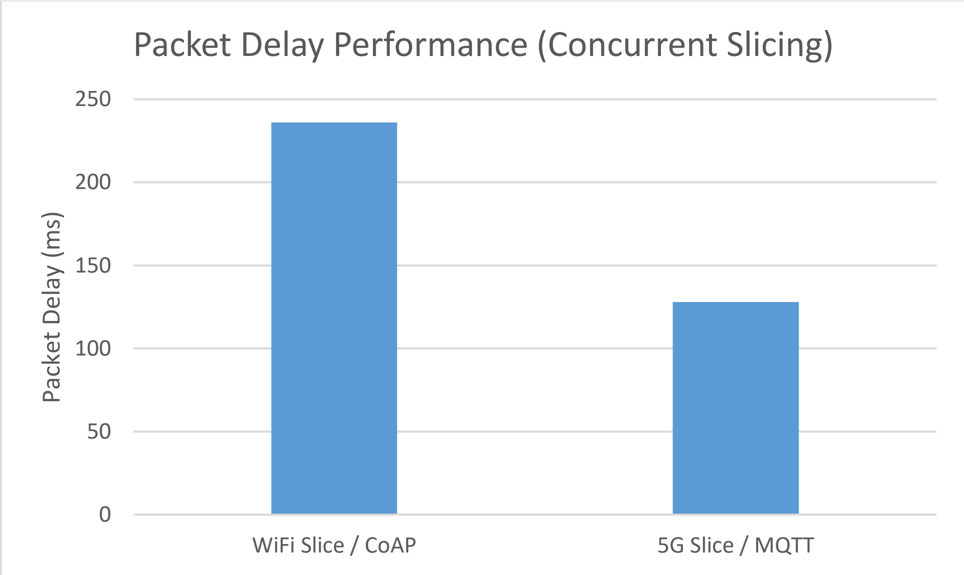

Figure 6-3. Packet delay performance in the case of concurrent slicing.

As the 30 boards were deployed over an area of approximately 30m^2, this resulted in the user density of 1 board per m^2, corresponding to the original user density KPI get of 1 million users per km^2.

6.4. UC#2: Coverage Evaluation – Measurements and results

In order to evaluate the downlink 5G coverage in the University of Surrey campus (home site of the Surrey Platform), a measurement campaign was conducted, during which the received Signal to Interference plus Noise Ratio (SINR) values of a 5G UE were measured in different locations (i.e., distances of the gNBs) in 5 outdoor sites. These SINR measurements were used in order to evaluate the coverage with the use of the Maximum Coupling Loss (MCL) KPI, as per 3GPP TR 38913. The coupling loss is defined as the total long-term channel loss over the link between the UE antenna ports and the gNB antenna ports, and includes in practice antenna gains, path loss, shadowing, body loss, etc. The MaxCL is the limit value of the coupling loss at which the service can be delivered, and therefore defines the coverage of the service.

For this test, the following RAN parameters and configuration were considered and 50 measurements were performed per site, see Table 6-3.

Table 6-3. Coverage Test RAN parameters and configuration

Carrier frequency	2.6 GHz
Bandwidth (BW)	100 MHz
SCS (Subcarrier spacing)	15 kHz
Number of gNB TX chains	4
Number of gNB RX chains	4
Number of UE TX chains	1
Number of UE RX chains	4

UE velocity	3 km/h
Traffic type	UDP
UE type used (Commercial/CPE)	CPE

For an eMBB service the target value of MCL is 140dB. As shown in Figure 6-4, all 5 outdoor sites in the University campus demonstrated MCL at or above the 140dB threshold in multiple gNB sites. The fact that a part of the measurements still fall below the threshold indicates the need for further optimisation by the vendor, taking also into consideration inter-cell interference constraints. The presented coverage measurements do not contain information on the distance to the gNB/RRU.

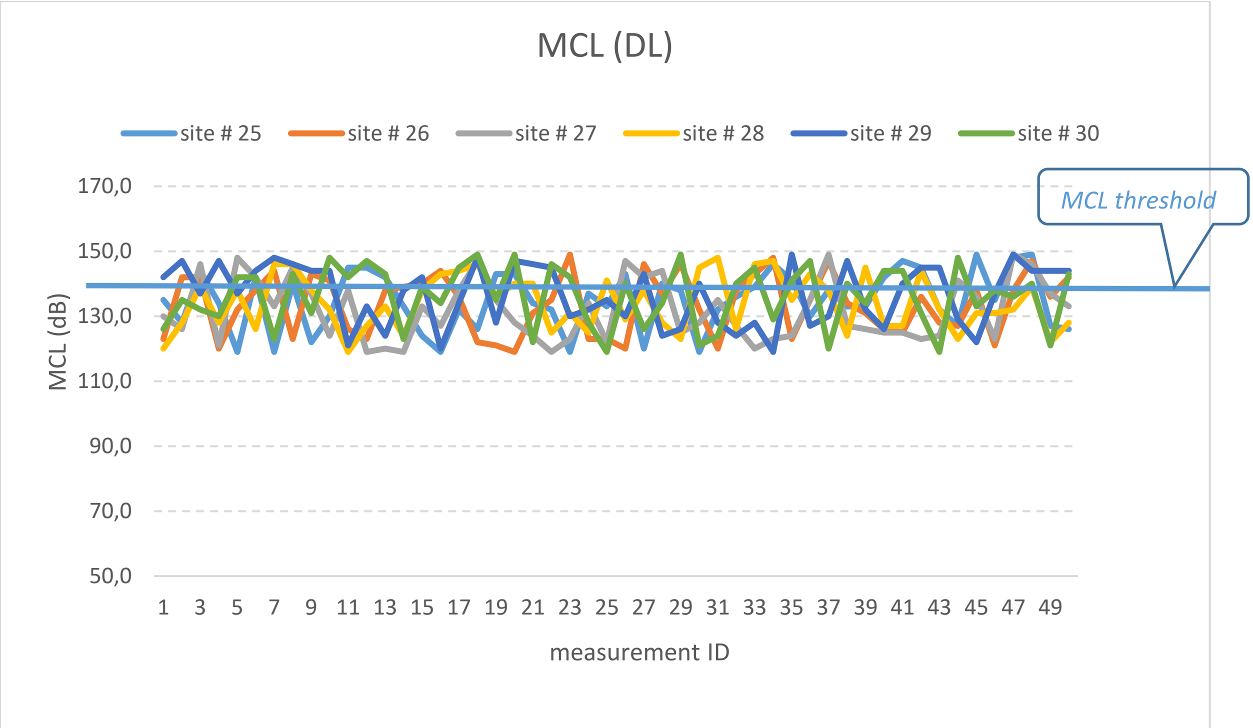

Figure 6-4. Downlink Maximum Coupling Loss in the University of Surrey Campus

As a complementary KPI of this test, the PDCP-level throughput in the downlink was measured as well. The results are depicted in Figure 6-5. The average throughput measured was 799.5

Mb/s, which was in accordance with the expected throughput of the 1st generation CPE used during these trials.

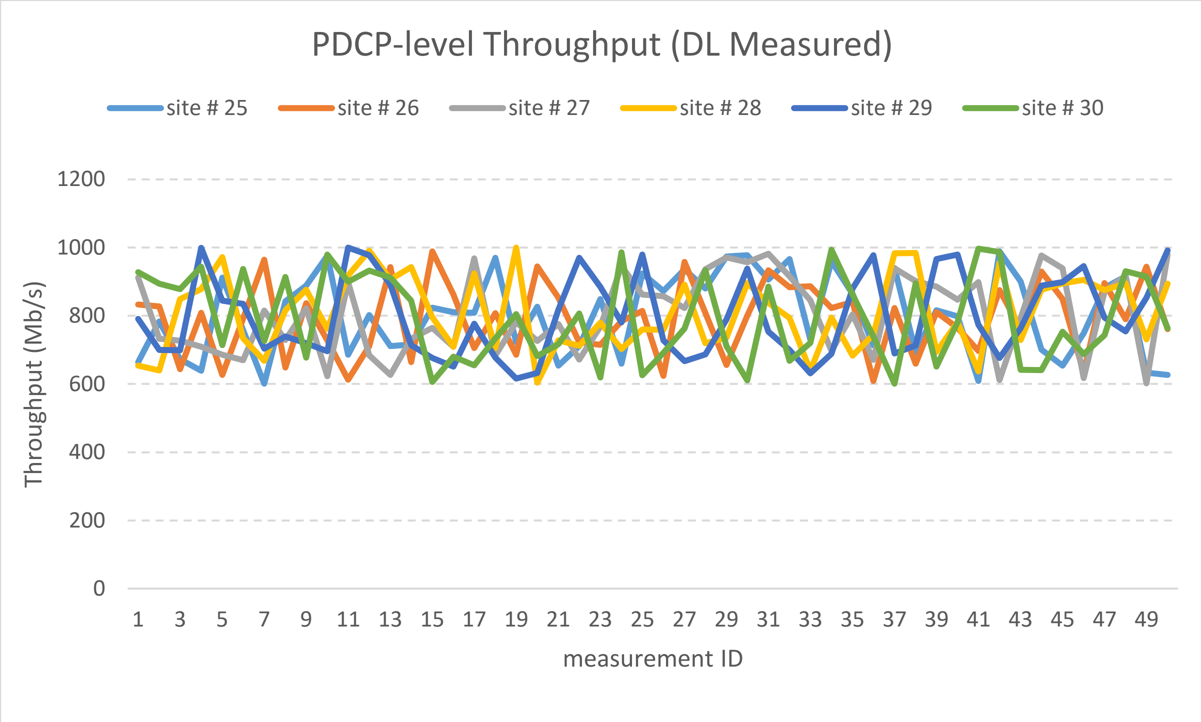

Figure 6-5. PDCP-level throughput measured on the downlink direction

6.5. UC#3: WSMP – Measurements and results

The developed WiFi Slice Controller (WSC) integrated in the WiFi Slice Management Platform (WSMP) receives instructions from the 5G slice manager through an API interface for creating, modifying, reactivating, deleting and monitoring the status of WiFi slices. As defined in Deliverable 4.12, it consists of three entities, namely the WiFi Slice Manager, the AP Manager and the Tunnelling Manager, located in three respective virtual machines in UNIS premises.

Figure 6-6. FON VM Server in UNIS Premises including WSC

Note that the model that has integrated WSMP has been adapted to the final deployment at UNIS premises as the existing WiFi Access Point is not mature enough to integrate with N3IWF that was developed at the 5GIC testbed. Secondly, a proper 5G SIM card (and a 5G SIM programmer), as well as 5G SA-capable UEs that allow the configuration in the device of a specific N3IWF endpoint for the non-3GPP RAT (WiFi) are not available on the market. However, to overcome these issues, 3GPP introduced a trusted WLAN Interworking Function (TWIF), which would support Authentication for devices that do not support 5GC NAS over WLAN access. However, this optional feature has not been implemented and therefore is not available for the Surrey Platform. This situation affects solely the initial authentication using AAA through legacy WiFi APs and has been overcome by setting up a separate radius server acknowledged by both APs and UEs and a translation service that enables the identification of the UEs in the 5GC when accessing via WiFi.

In this scenario, we conducted several trials to specifically evaluate the performance and operation of the WiFi RAT connected to the 5GC at Surrey Platform.

6.5.1. WiFi Service Creation Time Measurements and Results

The **Service Creation Time (SCT),** as defined in Deliverable 6.1, is the time required for the provision, deployment, configuration and activation of a full E2E communication service over a network slice. Translated to the WiFi segment, this is the elapsed time between a request to establish a new WiFi Slice is received, and the moment the service is completely up and running:

$$SCT = t_{up} - t_{req}$$

We can evaluate the SCT iterating the WiFi Slice creation requests from non-existing WiFi Slices in any AP, to reaching full capacity int at least one AP. The resulting values can offer a reference of how much the existence of other WiFi Slices in the APs affect the creation of new ones.

The request time (t_{req}) is registered upon reception of a WiFi slice creation request from the Surrey premises. The WiFi Slice Controller (WSC) collects the request timestamp and notifies it to the WiFi Slice Analytics Monitor (WSAM).

The deployment time (t_{up}) is obtained in the AP Manager, considering that at least one AP is up and running. When the new WiFi slice is deployed in that AP, it sends a trap notification to the WSAM with the corresponding timestamp.

Figure 6-7. Ruckus APs used in the deployment

After several iterations of the test case, we obtained the results depicted in the next figure. Note that these values are obtained for concrete AP models (2 Ruckus R550 & 2 Ruckus R650) and a concrete APM version (Ruckus vSZ 5.2.0). During the tests, the APs started to reboot and behave erratically when more than 10 WiFi Slices were deployed, hence the limit to 10 WiFi Slices per test round. Values should differ for other APs with different capabilities.

	SCT (s)				
WS #	ROUND 1	ROUND 2	ROUND 3		
1	13.834	17.737	12.764	MEAN	13.006
2	11.477	6.677	7.602	MEDIAN	12.764
3	19.626	15.382	11.958	ST DEV	4.018
4	10.817	9.57	7.679	MEAN	10.506
5	3.914	16.925	10.762	MEDIAN	10.790
6	12.923	11.336	13.312	ST DEV	4.058
7	6.103	5.507	17.228		
8	25.316	13.159	20.909	MEAN	23.260
9	27.611	16.314	23.276	MEDIAN	24.502
10	24.502	29.709	29.447	ST DEV	5.378
MEAN	15.612	14.232	15.494		
MEDIAN	13.379	14.271	13.038		
ST DEV	7.817	6.551	6.768		

Figure 6-8. Collected results for the Wi-Fi SCT test case trials

In the figure we observe three iterations of the tests, from which certain patterns of operation can be identified. If we divide the WiFi Slice deployment in three phases (Cold start, medium load and heavy load) and analyse all rounds at a time for each of these phases we get interesting results:

- Cold start phase: When the APs have no WiFi Slice deployed, the APs seem to last a bit more than expected in setting up the WiFi Slices. The average SCT is about 13 seconds, with a standard deviation of 4.
- Medium load phase: APs are already operating with increasing load as we deploy new WiFi Slices, but the behaviour is better than in the previous phase: The SCT decreases to about 10 seconds, maintaining the deviation.
- Heavy load phase: APs are reaching full capacity and deployment times grow. The internal management within each AP grows, which implies longer deployment times. The average SCT doubles the value of previous phases (about 23 seconds) with a standard deviation of 5.

If SCT value is a limiting KPI for a concrete deployment, and it is required to be kept low, the immediate conclusion is that APs should deploy a maximum of 60-70% of the capable WiFi Slices, to avoid experiencing considerable time increases.

6.5.2. WiFi RAN Measurements and Results

The **Throughput**, and **Reliability**, as well as the **Density of Users (DoU)** are evaluated in this test case. Specifically for the WiFi Segment, the DoU identifies the maximum number of UEs that can be connected throughout a WiFi Slice either at each AP or in the whole set of APs of a zone.

These three KPIs can be evaluated under different demanding conditions as a mean to model the WiFi Radio Access Network. For each iteration of the test, different configurations can be considered (varying the number of UEs, APs, and even the distance of the UEs to the APs). This results in an extensive collection of measurements that are very interesting in real deployments, and that require a more detailed analysis of all parameters involved.

For the matter of these trials, we adopted a simplified scenario in which to ratify the capability of measuring the achievable throughput during the test, the user count, and basic reliability parameters (packet loss, disconnections, AP rebooting), with the following considerations:

- Devices will be placed in a known and fixed location in close range to a single serving AP, so the RSSI values registered should remain mostly unaltered.
- The devices should be capable of gaining access to the 5GC throughout several WiFi Slices, as well as data offloading to other WiFi Slices, keeping any active session they might have (downloading data, video streaming), without expecting high data loss.
- As soon as the devices gain access to the 5GC for the first time, they will start downloading high volumes of data.
- WiFi Slices might differ in the deployment conditions (i.e., limited/unlimited maximum throughput). Also, one or many of them will be undeployed temporarily, and redeployed with different conditions (limited downlink throughput).
- AAA will be provided separately and won't be considered for the WiFi RAN modelling.

For this test, we connected two 5G capable UEs to the Surrey 5GC through a WiFi RAN with a single AP and two starting WiFi Slices (referred as WS01 & WS02 respectively from now on). Later in the test, WS02 will be undeployed and redeployed with different throughput capabilities (renamed as WS02'), and a third WiFi Slice will also be created (WS03).

The UEs gain access to the 5GC through the selected WiFi Slice using EAP-AKA', authenticating themselves towards the AAA server within the 5GC, and start exchanging high volumes of data (streaming video from a well-known video platform, and establishing a video call among them).

The sequence of events of the presented measurement test is the following:

Table 6-4. WiFi RAN Modelling Test Case – Sample Sequence

16:00	Test start. Both WS01 (SSID: 5GENESIS_EXPERIMENTS_01) and WS02 (SSID: 5GENESIS_EXPERIMENTS_02) are up and running with unlimited (unrestricted) bandwidth.
+5'	UE1 connects to WS01 and starts streaming a film.
+10'	UE2 connects to WS01 and establishes a video call to UE1 (UE1 keeps both tasks at once)
+17'	Both UEs are requested (manually) to connect to WS02. UE1 connects to WS02 at 16:17. UE2 connects to WS02 at 16.18. Video streaming and video call keep sessions and no freezing effect is observable by the operator.
+21'	WS02 gets intentionally undeployed. As expected, both UEs automatically return to WS01, keeping the sessions: UE1 keeps streaming the film, while maintaining the video call with UE2. The film keeps playing as expected (already had cached data to keep on) whereas the call freezes for a while, before resuming apparently unaltered.
+23'	WS02 gets deployed again with the same SSID, but capabilities have changed: Now the maximum downlink throughput is set to 5Mbps. (Hence the renaming to WS02').
+24'	UE02 voluntarily offloads to WS02' by decision of its internal network selection algorithm (detected with better signal or ranked above WS01 in the internal list of the smartphone). It is a new WiFi Slice but, as WS02' keeps all network access configurations of WS02, it is detected as the same slice.

+28'	WS01 gets undeployed. UE01 is forced to connect to WS02'. Same behaviour observed as in 16:21: The call freezes for a while, before resuming. The session remains.
+31'	WS03 is deployed (SSID: 5GENESIS_EXPERIMENTS_03) with downlink throughput set to a maximum of 40Mbps.
+32'	UE01 is forced (manually) to connect to WS03. And several speed tests are conducted in both UEs. UE01 shows 32.98 Mbps, and UE02 shows 4.63 Mbps, (as expected).

The following graphs depict the collected measurements described in the sequence above. Packet loss & retransmission metrics are not included in the graphs of this example as no loss or retransmission errors have been reported to the WSAM.

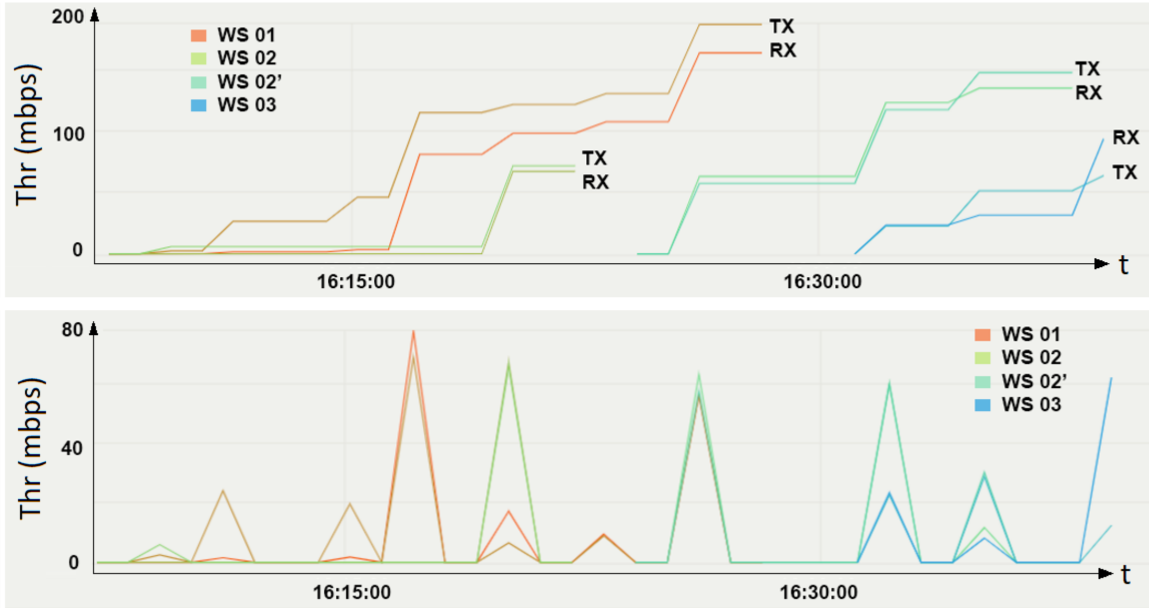

Figure 6-9. Above: Aggregate Tx & Rx throughput per WiFi Slice over time. Below: Instant Tx throughput per WiFi Slice over time.

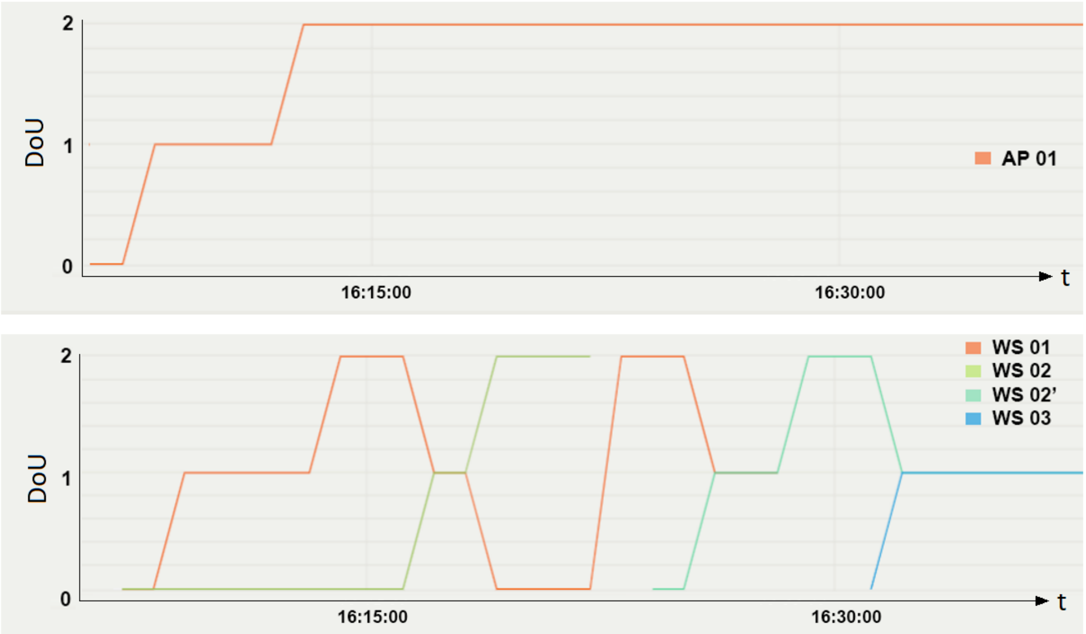

Figure 6-10. Above: Aggregate Density of Users per AP over time. Below: Density of Users per WiFi Slice over time.

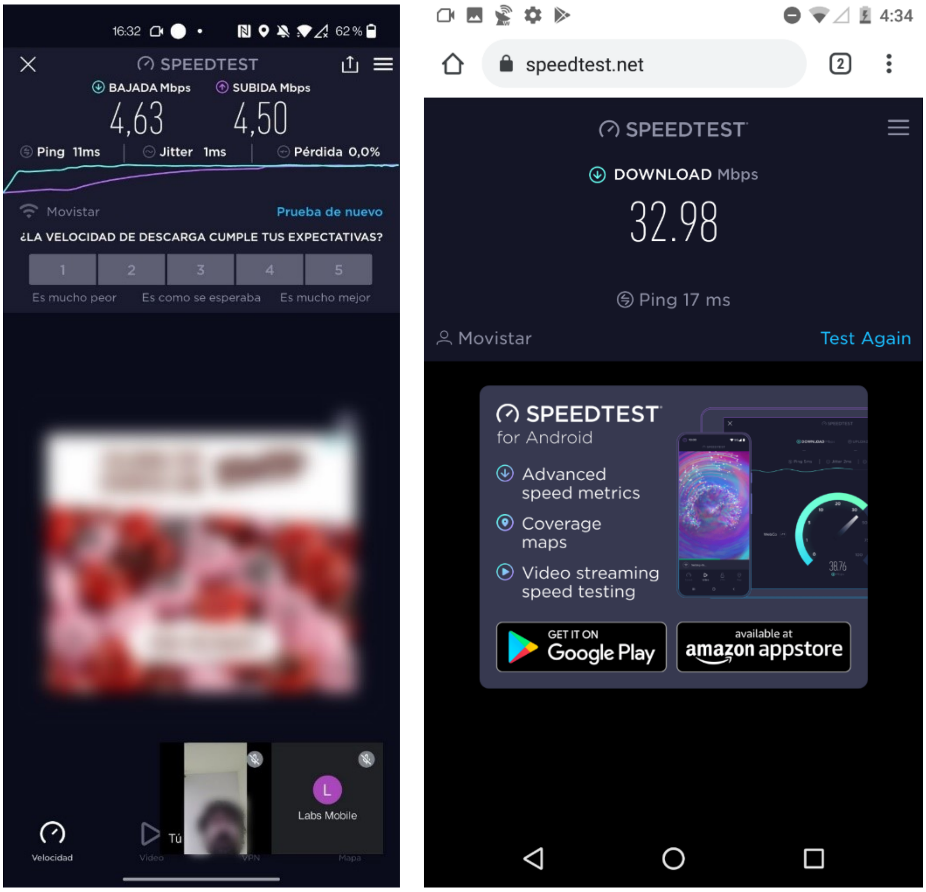

Figure 6-11. Captures of speed tests conducted in both UEs at the end of the test.

As demonstrated in the previous paragraphs, the WSMP and its entities (WSC, WSAM, Tunnelling Manager) are sufficiently capable of offering conclusive results on the modelling and contribution of the WiFi segment to the calculation of KPIs in the establishment of a commercial 5G network with support for non-3GPP RATs.

6.6. UC#4: CoAP over LTE and 5G radios – Measurements and results

The experiment evaluates CoAP performance over 5G and LTE, using the round trip time (as defined in D6.1) under different workloads as the main KPI. The CoAP server is run at Karlstad University, Sweden. The CoAP client runs at the MONROE nodes placed within the Surrey Platform testbed. The client uses 5G (CPE AP) and LTE eNB respectively. The server is an echo server that echoes any message received from a client. In each run, the CoAP client sends a sequence of four messages to the server. Three different message sizes (100, 200 and 400 Bytes) and three different inter arrival time (0.5, 1 and 2 s) are used. The experiments are repeated 30 times, for a total of 270 experiment runs for each access technology. Supplementary ping experiments are also performed to validate the CoAP results. See also the test case template TC_RTT_COAP_SUR in the 5GENESIS Test Cases Companion.

Results obtained from the tests are shown in Figure 6-12 and Figure 6-13. We can see that the RTT performance between LTE and 5G is very similar, with the performance of LTE being slightly better. Figure 6-12 shows that 5G is faster (mean and median wise) for message sizes up to 200 bytes and message inter arrival time up to 1 s. LTE has both lower mean and median RTT as compared to 5G for the other cases. Figure 2 shows the results from the ping experiment. Similarly to the case of the CoAP experiment, the figure also shows that the performance of LTE and 5G is similar. Please see also the accompanying test case report for detailed numbers.

At this point it should be noted that the similarity in the RTT and ping performance between LTE and 5G is justified by the fact that 5G access to the MONROE node was provided with the use of a CPE UE (Huawei 5G CPE Pro 2), which was only capable of operating in NSA mode. The node was connected to the CPE via an Ethernet connection. Considerably improved 5G results would be expected if the node was equipped with a 5G modem and had SA capability.

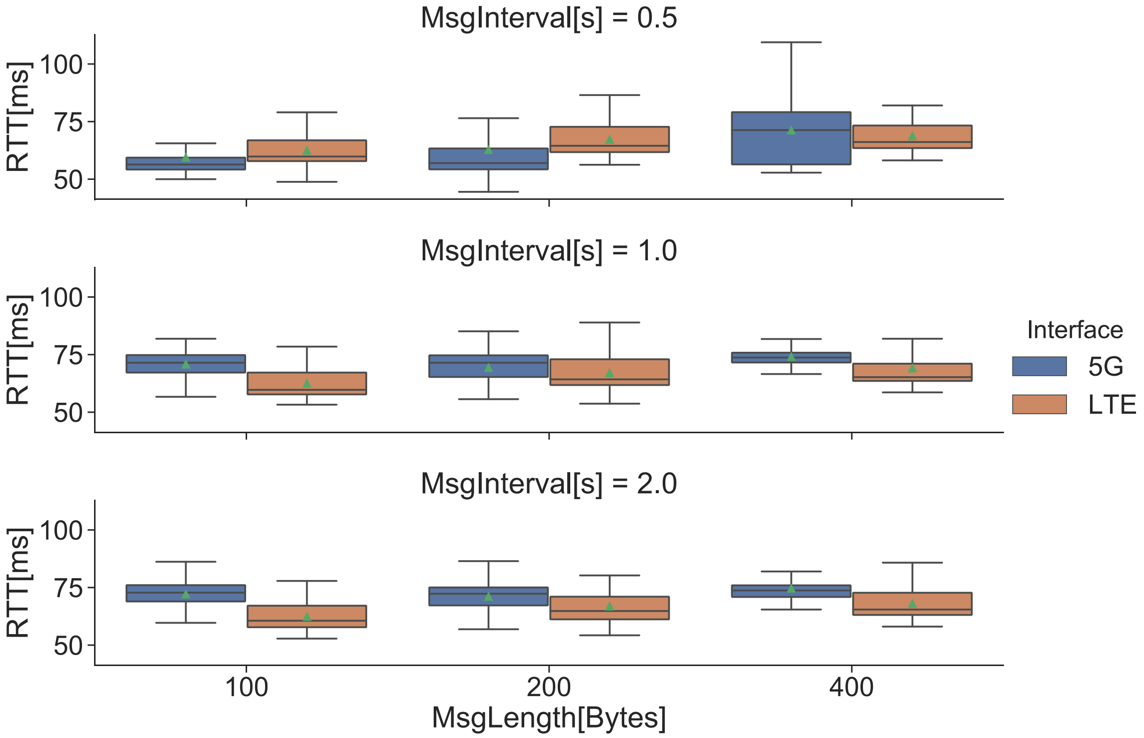

Figure 6-12 CoAP RTT for different MsgLength and MsgInterval.

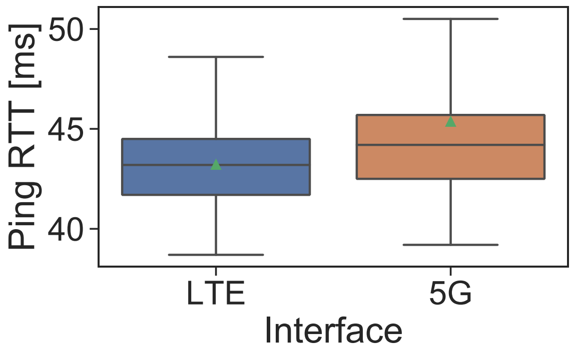

Figure 6-13 Ping RTT

6.7. UC#5: APEX Integration – Measurements and Results

Adaptive Policy Execution (APEX), previously described in D4.11, is a lightweight engine for the execution of APEX policies. APEX policies can be designed for the straightforward execution of a single task or expanded into a larger complex model of multiple tasks and states. APEX policies can be designed to self-adapt through the dynamic selection of tasks influenced through the onboard context information of the policy driven entity. In the context of the Surrey Platform, APEX was integrated and tested in a dynamic slicing use case. More specifically, slices are dynamically spinned-off based on the values of specific performance metrics, such as CPU and RAM use.

The architectural setup of the APEX integration use case is shown in Figure 6-14. The *POSTMAN* is the entity that is responsible for sending requests to the Use Case Data Manager in order to initiate or terminate i) event feeding to the APEX policy engine, ii) collection of metrics of interest (such as CPU usage), and iii) workload generation. The *Use Case Data Manager* is responsible for receiving the respective requests and acting accordingly. Specifically, it sends metric collection requests to *Prometheus*, who is responsible for metric monitoring. Following the receipt of this information, it sends events to the *APEX policy engine*, who is responsible of processing this information and deciding on whether an adjustment of the workload carrier replicas is necessary. The Use Case Data Manager, when receiving POSTMAN requests to start or stop workload generation, also notifies the *Use Case Workload Carrier*, who acts accordingly. *Grafana* is used for the visualization of the respective metrics.

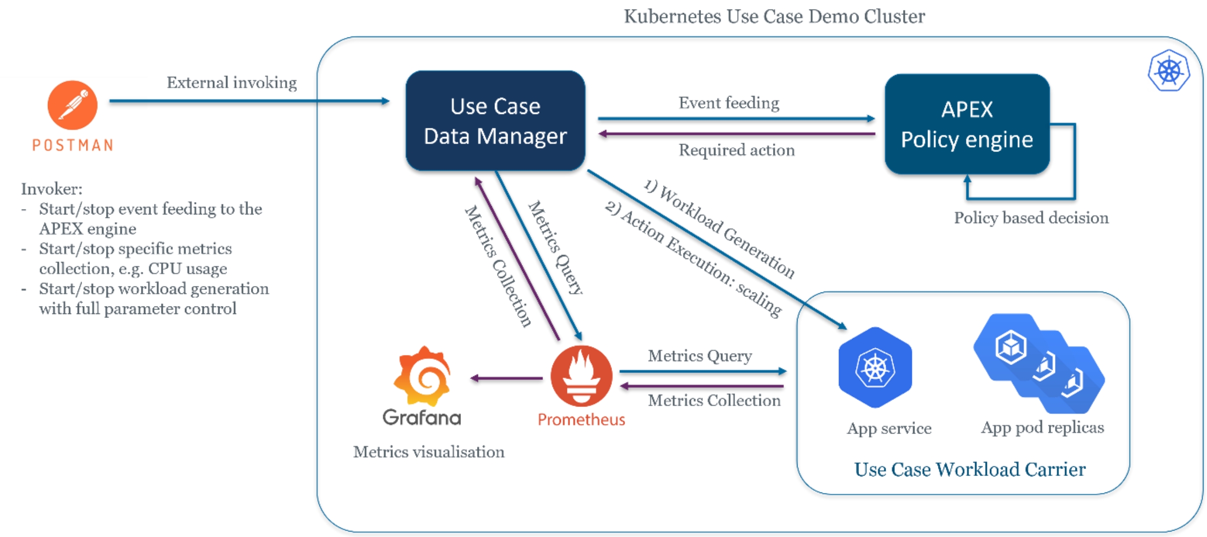

Figure 6-14. APEX Integration Use Case Setup

The test scenario included a number of steps, each of which was used in order to validate the correct operation of the different components and their efficient communication. The detailed steps followed are shown in Table 6-5.

Table 6-5. APEX integration test scenario

Steps	Expected Actions	Test Results
Send POST request from POSTMAN to the Use Case Data Manager – “apex-feeder-5genesis”, with parameters of: action(“start”), metric types that will be collected, metrics threshold for APEX decision making, microservice which produces the metrics, and metrics collection interval.	- The Use Case Data Manager receives the request and correctly processes the parameters. - - The Use Case Data Manager starts to collect metrics of the data type requested in the parameters, and the collection follows the received collecting interval.	Passed
	- The Use Case Data Manager starts to send policy events to the APEX engine in JSON body. - The APEX engine correctly processes the event data and makes decision based on the defined policy logic. - The Use Case Data Manager receives the action from the APEX engine and takes the action accordingly.	Passed

Send POST request from POSTMAN to the Use Case Data Manager – “apex-feeder-5genesis”, with parameters of: computing factor and loops for the Workload Carrier – “workload-5genesis” to increase its workload, an interval timer for the Use Case Data Manager to periodically send requests to the Workload Carrier to trigger the computation task.	- The Use Case Data Manager receives the request and correctly processes the parameters. - The Use Case Data Manager starts to send requests to the Workload Carrier based in time ticker (interval). - The Workload Carrier receives the requests with the parameters and executes the computation task accordingly.	Passed
	- The workload rises up and is shown in Grafana. - The decision of increasing the number of Workload Carrier replicas is made by the APEX engine when the required workload limitation is exceeded. - The average workload of the Workload Carrier replicas is dropped alone with the increment of replicas, and shown in Grafana. - More Workload Carrier replicas keep being created until the workload drops below the threshold.	Passed
Send POST request from POSTMAN to the Use Case Data Manager – “apex-feeder-5genesis”, with parameters of: action(“stop”) to stop generating workload	- The Use Case Data Manager receives the request and stops sending task requests to the Workload Carrier. - The Workload Carrier metrics (CPU usage) value drops below the 10% of the threshold, which triggers the APEX engine with the decision of reducing the number of Workload Carrier replicas. - The Use Case Data Manager receives and executes the action. - The number of replicas is reduced and shown in Grafana.	Passed

A detailed description of the test is provided in Annex 3.

6.8. Summary and conclusions

The main use case of the Surrey Platform in Phase 3 focused on demonstrating the support of multi-RAT communication for IoT services with the use of concurrent slicing. The most important outcome of this work was the capability to offer the IoT-vGW as a service, with a dedicated slice per RAT, and the ability of more than one slices to be spinned at the same time. Packet loss and packet delay of different IoT protocols (HTTP Post, MQTT and CoAP) over WiFi and 5G slices were evaluated, showing that the 5G slicing resulted in considerable reduction of the packet delay.

The MCL-based coverage measurement campaign in the University campus showed multiple sites that exceed the eMBB specific MCL threshold, with a PDCP-level Throughput that was in accordance with the expected values (in accordance to the capabilities of the CPEs used).

The experiments including the WSMP and its main components, i.e., WSC, WSAM and Tunnelling Manager, were used to evaluate the preferable load of WiFi Slices in order to achieve a reasonable WiFi service creation time (in the range of 10.79 to 24.5s). Moreover, Throughput and Reliability KPIs were measured, demonstrating that the deployed WiFi Slices provide efficient support of the UEs performing data streaming.

Finally, the successful integration of the APEX policy engine is a key asset of the Surrey Platform, as it has allowed for the performance driven adaptation of the available workload carrier resources (see Annex 3).

7. BERLIN PLATFORM EXPERIMENTS

7.1. Overview

Phase 3 of the Berlin Platform experiments and trials focused on two main targets in terms of KPI assessment. First, given the major upgrades that took place at the Berlin Platform infrastructure in this last phase, there was a need to re-assess the performance of the underlying compute, storage and network infrastructure of the Platform. The second target relates to the re-execution of the 360-degree camera use case employing the nomadic 5G node, which allows the involvement of many (project-external) users in a field trial (c.f. 7.1.1).

With the end of Phase 3, the assessment of all KPIs to be evaluated by the Berlin Platform has been completed. Final measurement results and the executed trail show that:

- The performance of the underlying compute & storage and network infrastructure does not impose any limitation to any 5G system deployed onto, as it features sustained throughput of at least 20Gbps staying within the virtualization environment, and of at least 9Gbps leaving the virtualization environment towards external, bare-metal components or between different compute nodes. Means of observed RTTs range, depending on the set-up, between 2.5 and 3.5ms.
- The performance of the nomadic 5G node suits larger field trials as demonstrated with 360-degree video applications.
- The 5G Core (Open5GCore) and RAN deployed in the Berlin Platform provides 350Mbps (50Mbps) throughput in the down- (up-) link and an RTT in the order of 28ms, using COTS mobile phones. During the field trial, using 4x4 MIMO CPEs, sustained data rates of 575Mbps in the downlink and 80Mbps in the uplink at an average RTT of 15ms were measured.

In addition, the measurement results show that the evaluation methodology and tools leveraged throughout the experiments are well capable of revealing the influence of, e.g., different packet sizes, on some KPIs such as the RTT. This makes the 5Genesis Berlin Platform a well suitable environment for the evaluation of future, use case specific applications, having specific requirements on the 5G network performance based on the characteristic of induces traffic.

7.1.1. Outline of Phase-3 Measurements

Since the execution of phase 1 and phase 2 experiments, the Berlin Platform infrastructure for experimentation was completely renewed – especially inside the FOKUS facility. The renewal entailed a full replacement of the underlying compute & storage and switching infrastructure as well as the DWDM-based interconnect to additional remote facilities connecting to the Berlin platform. Additionally, components of the infrastructure that had been based on prototypes or hardware running pre-commercial prereleases of firmware were now replaced with commercial off-the-shelf versions. This replacement, in particular, applied to the 5G SA RAN of the Berlin Platform. The details of the new setup are given in [7]. The specific details of the components that are located especially at the FhG FOKUS facility are given in Section 2.1.2 of [7].

Due to the changed platform infrastructure, it was required to re-execute thorough baseline measurements that allow verifying whether execution requirements for the final trial's measurements are holding the required properties. To that end, the methodology employed in phase 1 and 2 was followed. Therefore, the experiments described in section 7.2 ensure that the newly deployed infrastructure will not be a bottleneck – or have similar negative interference effects – for the actual phase 3 use case measurements.

A 360° camera use case was executed during phase 3 trials at the IHP facility, likewise it was carried out at the Berlin Festival of Lights (as described in [8]). For the final trial event at IHP, several experiments were executed during the preparation. Their results are described and analyzed in detail in section 7.3.

Table 7-1 Primary 5G KPIs evaluated at the Berlin Platform in the third trial phase

KPI to be evaluated at the Berlin Platform according to DoA	Evaluated in Phase	Comment
Density of Users	Phase 1	Based on number of users attached to the Open5GCore (using a UE emulator)
Service Creation Time	Phase 1, Phase 2	Based on deployment of VMs in Phase 1 Based on deployment of a full testbed tenant in Phase 2
Speed	Phase 3	Experiments assessing RTT and Throughput were executed at pedestrian-speed of the UE and for stationary UE.[6]
Reliability (RTT)	Phase 3	Trials at IHP event & additional measurements at Berlin 5G infrastructure
Additional 5G KPIs evaluated at the Berlin Platform	**Evaluated in Phase**	**Comment**
RTT	Phase 1, Phase 2, Phase 3	Preliminary infrastructure and 5G SA RAN prototypes (phases 1 & 2) New infrastructure and COTS 5G SA RAN (phase 3)
Throughput	Phase 1, Phase 2, Phase 3	Preliminary infrastructure and 5G SA RAN prototypes (phases 1 & 2) New infrastructure and COTS 5G SA RAN (phase 3)

[6] As pedestrian velocity vs. stationary UEs had no influence on the KPI measurements, measurement results are given in the following for stationary UEs only.

Additional experiments were executed to assess the 5G infrastructure at the Berlin facility and are detailed in section 7.4.

The table above presents the overall status of all KPIs evaluated by the Berlin Platform as of the end of phase 3.

7.1.2. Note on Covid-19-related influences

The final trial event at IHP facility and all directly related preparations were executed during week 45 to week 47 of 2021. The experiments associated with the trial focused on the evaluation of the deployed 5G system "in the field", as well as on involving users in the trial. To minimize the risk due to Covid-19 and due to healthcare restrictions, additional experiments not requiring a larger number of external trial users were executed in the weeks prior to the event.

The IHP facility is in the *Bundesland* (state) "Brandenburg", whereas the FOKUS facility is located in Berlin. Both regional administrations act mainly according to the Covid-19 state of play of their respective regions. Simultaneously to Phase 3 Trials, the case numbers of Covid-19 infected people increased significantly. Especially in Brandenburg, the numbers went significantly up, fast. This resulted in new Covid-19 restrictions that were in effect during the very last day of the trials, planned for the involvement of many project-external users. Those new restrictions prevented most employees of IHP from coming to work, as they were sent into home office the day before. This fact impacted the actual trial event, since only a much-reduced number of users could be engaged.

While the actual execution of measurement experiments (KPI evaluations) for the IHP facility infrastructure was also affected by the overall Covid-19 situation, the effects were not as harsh as for the actual trial event. Though the execution of experiments took significantly longer due to the limitations to access the laboratories at FOKUS, all planned tests could eventually be conducted. Only the number of project-external users involved in the final trial was reduced.

In summary, Covid-19 did not affect the planned evaluation of the KPIs at the Berlin Platform except for the fact that the final trial did not involve a very large number of project-external users. However, this was compensated by the involvement of the public during the first field trial during the Festival of Lights in 2019.

7.2. Generic tests – Measurements and results

The executed baseline measurements were focusing on the endpoint-to-endpoint interconnect between two possible measurement nodes. The objective of those measurements is to assess whether the underlying infrastructure imposes performance limitations once it is used to instantiate the Berlin Platform testbed. To this end, the main properties that needed to be analyzed in detail were: (a) the latency of a given interconnect[7], and (b) its throughput capacity.

To provide measurements that indicate the latency of a given interconnect, round-trip time (RTT) measurements are taken at the network layer. To analyze the throughput capacity of a given interconnect TCP-based throughput measurements were executed. It should be noted that only the payload throughput is used for the analysis, instead of using the full-frame at the physical layer.

The expectation for the final KPIs of an end-to-end 5G system are in the order of O(1 Gbps) throughput and O(10ms) RTT. Thus, in order to be validated as a suitable underlying compute & storage and network infrastructure, generic measurements should reveal a performance at least in this order of magnitude or, preferably, one magnitude larger for throughput and lower for RTT observations.

At the FhG FOKUS facility, the newly set-up infrastructure also involves a data center, based on CISCO Unified Computing System – UCS-mini and VMware vSphere (see: Section 2.3.2.1. "Main Data Center" of [7]). The deployed CISCO UCS mini blade server consists of six computing nodes. Furthermore, the new computing infrastructure provides different ways of interconnecting measurement nodes, and two main classes of interconnects were analyzed:

1. Intra-Compute: Both measurement endpoints are running on dedicated, separate virtual machines (VM) but both VMs run on the same compute node. The interconnect between two VMs is handled by the compute node, internally only.
2. Inter-Compute: Each of the two measurement endpoints is run on a dedicated, separate VM and each VM is executed on a separate compute node. The interconnect between two VMs spans from one compute node to the other but also via the networking backplane of the whole computing infrastructure.

7.2.1. A note on "auto-migration"

As mentioned in the introduction of this section, the CISCO UCS mini blade server deployed in the FhG FOKUS data center comprises six computing nodes – C101 to C106. As noted in the upcoming measurement subsections, during some of the measurements, VMs acting as measurement endpoints were auto migrated by load balancing procedures of the system. The load balancing could not be disabled in time for the measurement campaign.

During early test measurement sessions, the following issue surfaced first. Especially when executing throughput measurement experiments, the cluster control software tried to load

[7] Please see the term "interconnect" as a tryout to circumvent misunderstandings when using terms "connection" or "link" instead. Interconnect refers to anything that connects two (logical or physical) endpoints – directly or indirectly. To some readers "link" refers to physical connections only. To other readers, the term "connection" will imply only stateful data transmission.

balance between the different cluster nodes. For example, assuming that an Intra-Compute throughput experiment for C101 was started. That means the VMs that hold the measurement endpoints are to be run only on the C101 compute node. For some experiments it could be observed that the load balancing “kicked in” and the cluster control software automatically migrated (in the following called “auto-migration”) one of the measurement endpoint VMs to another (e.g. C102) compute node.

For yet to be established reasons, we did not manage to prevent this behavior during the trials – even with configurations suggested by the manufacturer. Therefore, whenever the problem showed up, we tried to re-run the experiments as the time allowed. Detailed descriptions to this problem are provided in the Annex.

7.2.2. Intra-Compute Throughput (TCP)

The initial tests of the new virtualization platform focus on assessing the performance of a single compute unit (blade) within the virtualization cluster. This way the sources of possible (negative) interferences were narrowed down to the bare minimum.

For Intra-Compute Throughput (TCP) measurements, two VMs are running on the same compute node and TCP-based data is sent in a unidirectional way. The connection initialization will be from a Client endpoint to a Server endpoint for all tests. Data traffic is always sent from the Client towards the Server. In case the throughput of the other direction needs to be assessed, the roles of the two endpoints have to be switched, i.e., the endpoint running the Client as then to be the Server, etc.

The measurement tool (iperf) allows to have two observation points reporting the recorded throughput. If the Server is the observation point, the reported throughput is directly based upon the traffic successfully received by the Server from the Client. If the Client is chosen as the observation point, the reported throughput is based upon the acknowledgments (TCP-ACKs) sent from the Server towards the Client, which allows to calculate at the client-side the data successfully sent towards the Server.

In order to assess the throughput in both directions between endpoints, the roles of the endpoints are switched for each experiment. The experiment ID indicates which endpoint (e.g. VM A or VM B) acts as client or server. The first mentioned endpoint is always acting as the Client and the second mentioned endpoint as the Server. Thus, it can be analyzed whether the interconnect shows symmetrical behavior regarding data transfer direction. Additionally, for some experiments the endpoint pair is varied (used VMs on the same cluster). Verifying that each of the six compute nodes of the cluster exhibits the same behavior, tests are repeated for multiple nodes.

Continuing with the results, Figure 7-1 and Figure 7-2 depict TCP Throughput experiments that were executed on separate compute nodes C104 and C106.

A rather surprising result is that each of the first executed measurements (for these figures) show an average TCP throughput of roughly 30 Gbps, while each additional measurement shows only 20 Gbps. Various additional series of experiments were executed to find a proper explanation of this behavior. As the additional experiments could not provide a reasonable explanation for the previous, single observation, the measurement results of those confirmation experiments are not discussed in detail below.

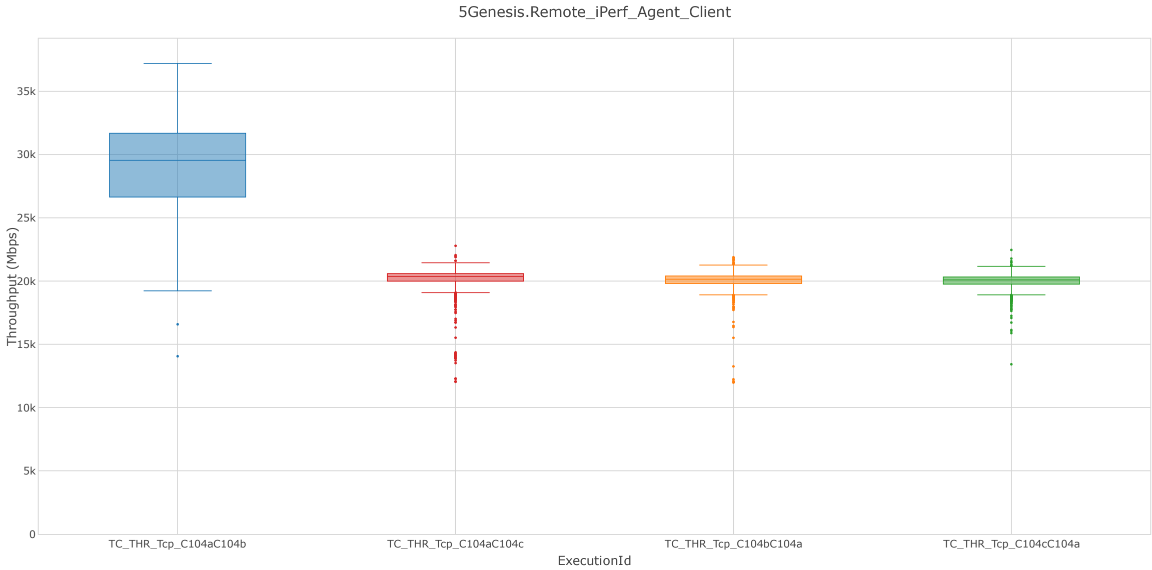

Figure 7-1 Intra-Compute Throughput for C104 (values observed at Client)

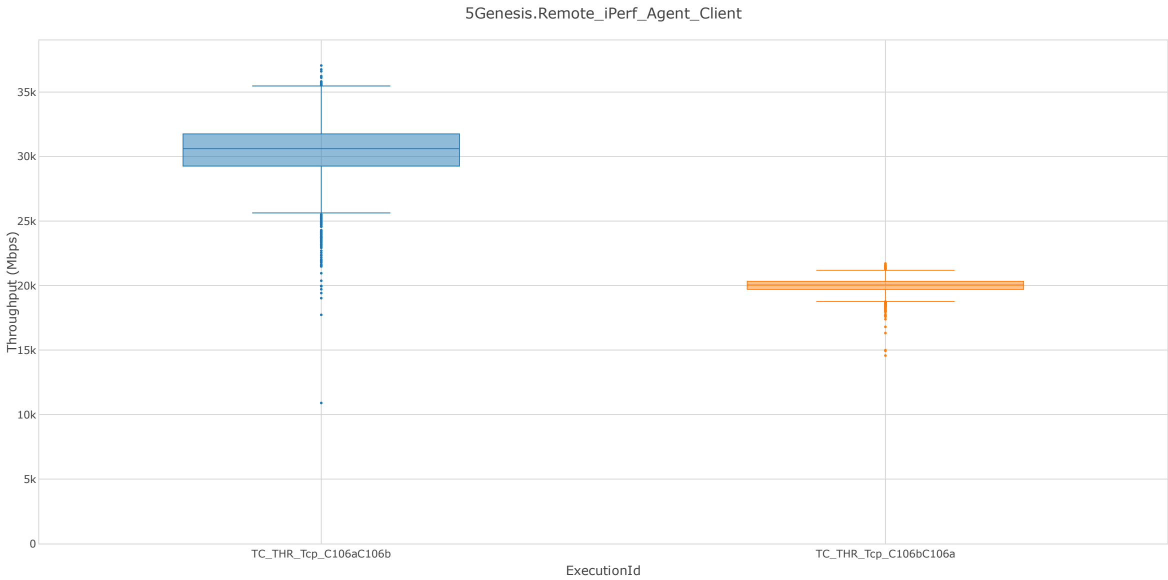

Figure 7-2 Intra-Compute Throughput for C106 (values observed at Client)

What can be excluded as explanation is that only first measurement runs will explain this behavior (in some instances it was not only observed for first but also later runs). This behavior is also not explainable by the direction of data transmission or whether the order of executions (and their direction) plays a role. Finally, this behavior is not even specific to unique compute nodes in the cluster.

Regardless of this, all four figures reveal that an average of 20 Gbps for TCP throughput can be reached in the Intra-Compute case reliably. The direction of the traffic flow between endpoints does not impact the observed throughput. Since the targeted baseline was 10 Gbps (as a minimum) the capacity of the interconnect in the Intra-Compute case is fulfilled.

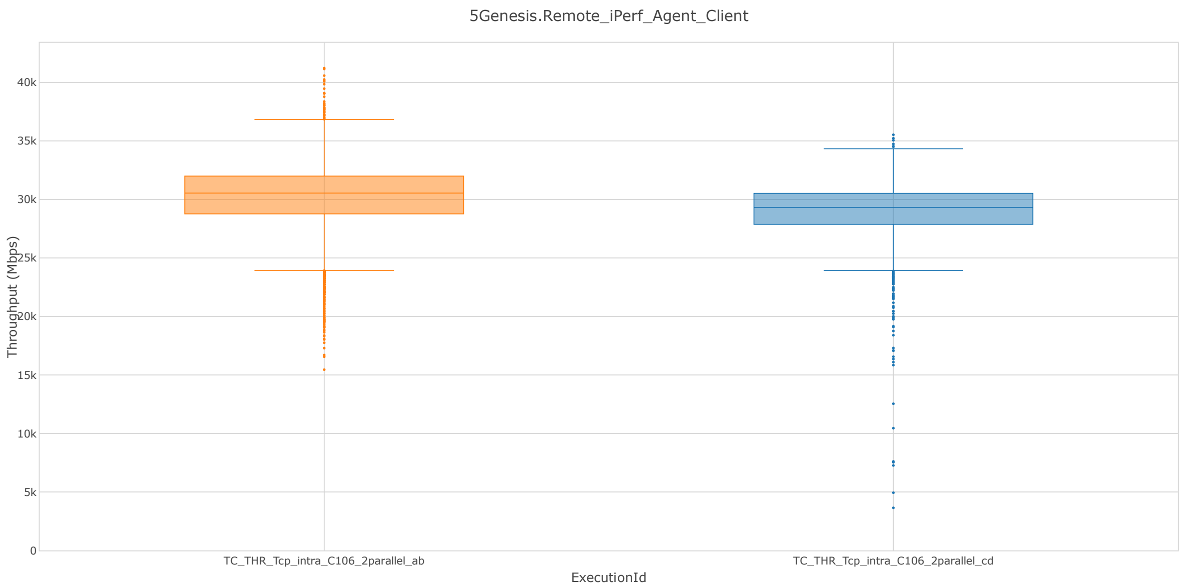

Figure 7-3 Intra-Compute Throughput for C106, two parallel streams (values observed at Client)

TCP is designed for fair resource sharing. In short: whenever multiple TCP streams that span over a path between their individual endpoints share a common path element, the throughput usage is to be shared among the TCP streams in a fair way. Therefore, running a TCP stream over a path element that is shared with other TCP streams will affect the throughput of the other TCP streams and vice versa. To be able to inspect this behavior and its effects, the Intra-Compute Throughput measurements are extended with measurements that run multiple TCP streams simultaneously. For these experiments two pairs of VMs were used, where each VM was deployed on the same cluster node (e.g. C106). The throughput measurements are executed by running a single TCP stream for each VM pair (e.g. A/B and C/D). Both TCP streams were started and stopped at (nearly) the same time. Hence, the TCP streams were run simultaneously (or in other words: in parallel). The results of these experiments are depicted in Figure 7-3 as well as in the annex.

Figure 7-3 depicts the Intra-Compute Throughput experiment with two in-parallel running TCP throughput measurements for the compute node C106 as observed at the Client. Observation at the Server are comparable and are hence included in the annex for the same two-in-parallel streams experiment. For either observation point, the median value of the data rates for stream C/D and stream A/B are roughly at 30 Gbps. In the case of A/B (red), outliers go up to over 40 Gbps, while the C/D only reaches a little over 35 Gbps, at maximum. It is also worth noting that the C/D case (blue) must have experienced quite low data rates, at times – outliers even below 5 Gbps are visible.

7.2.3. Intra-Compute Round-Trip Time (RTT)

Like the first measurements in Section 7.2.2, with the measurements presented in this section only the capabilities of a single compute node are investigated. Additional results for other compute nodes are quantitively the same and included in the Annex.

This time the main interest shifts to RTT measurements. Later, in a follow up section, measurements are presented that will span multiple computing nodes.

For Intra-Compute RTT measurements, two VMs are running on the same compute node and ICMP echo request/reply – at networking layer – are sent between the two endpoints. For the different measurements, the used compute node is varied to check whether different compute nodes behave in the same way. To investigate whether payload size influences the RTT of a given interconnect in any way; the payload size of a given ICMP echo message is varied. Here, the values of 32, 56 and 1400 bytes are used. Each of the executed experiment is repeated 25 times.

Considering the three measurements shown in the figures below, varying the payload size does not appear to influence the RTT in any significant way. In all three cases the median value is roughly 0.2ms.

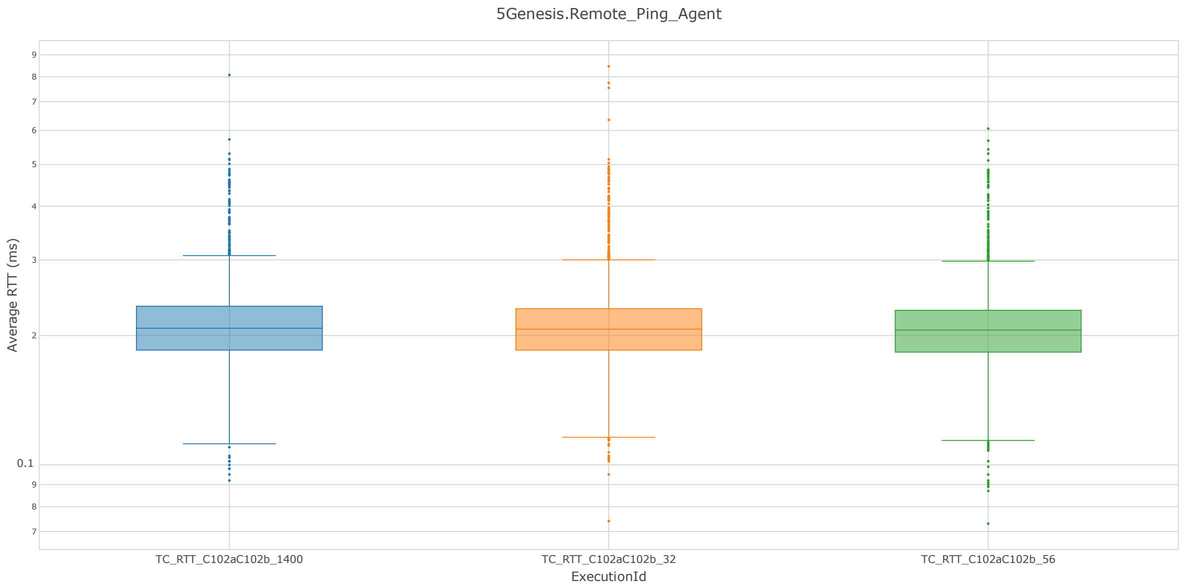

Figure 7-4 Intra-Compute RTT for C102

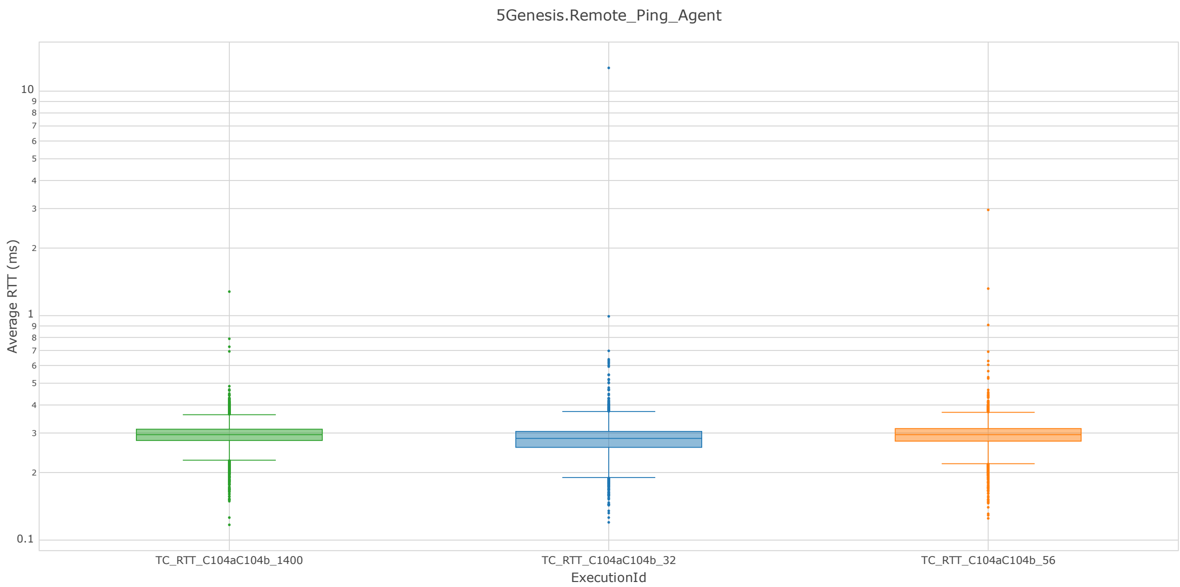

Figure 7-5 Intra-Compute RTT for C104

When comparing the results of compute node C102 with the ones for C104 (Figure 7-5) a significant change is visible. While the payload size does not seem to influence the RTT for this case, a wider variance in the measurements is visible. Also, while in the case of RTT measurements on compute node C102 the median value of all three payload sizes was at about 0.2ms, in the case of node C104 slightly higher median values are visible. Here, the median is a little below 0.3ms.

7.2.4. Inter-Compute Round-Trip Time (RTT)

The measurements provided in this subsection can be seen as an extension to the interconnect latency analysis of subsection 7.2.3. This time the connections span multiple compute nodes by leveraging the cluster's networking backplane. Like the Intra-Compute RTT measurements, ICMP echo requests/reply messages are sent from one endpoint to another. Each endpoint running inside a separate VM. Unlike the Intra-Compute RTT measurements case, however, each VM is running on a separate compute node.

In line with the intra-compute measurements, it is investigated whether the size of send data will influence in any way the RTT of the given interconnect. Here, the payload size of a given ICMP echo message varies with values of 32, 56 and 1400 bytes.

The cluster consists of six compute nodes. Therefore, theoretically, each possible pair of inter-compute interconnects could be tested. To reduce the complexity of this effort, only for one compute node (acting as an anchor point for Inter-Compute interconnect experiments), the full set of experiments is executed. Here one compute node is selected as the source anchor compute node for RTT measurements. All other compute nodes are selected as RTT measurement targets.

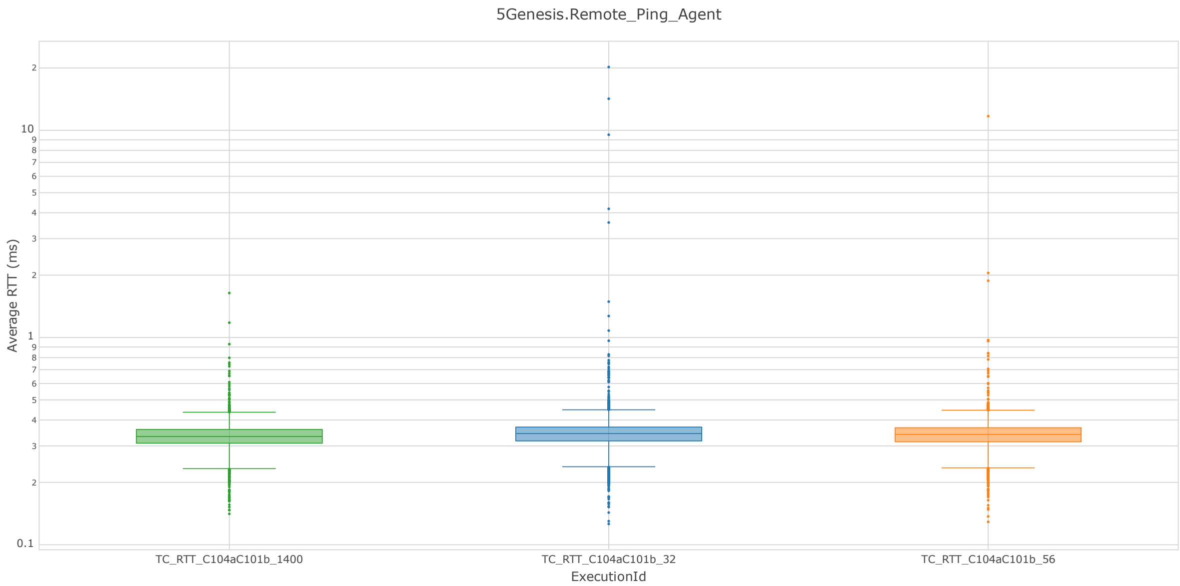

Figure 7-6 Inter-Compute RTT, 104 -> 101 with different payloads

For each of the experiments, it can be seen that varying the payload size seems to have no effect on the average RTT. To exemplify, Figure 7-6 depicts the RTT results between nodes 104 and 101. The results for other connections are presented in the Annex.

In Figure 7-7 one can observe the results of experiments using compute node C104 as source and a 56 Byte payload. Similarly, the results for source node C102 are depicted in Figure 7-8.

With results rarely surpassing 1ms, varying the target compute node did not have a significant impact on the average RTT.

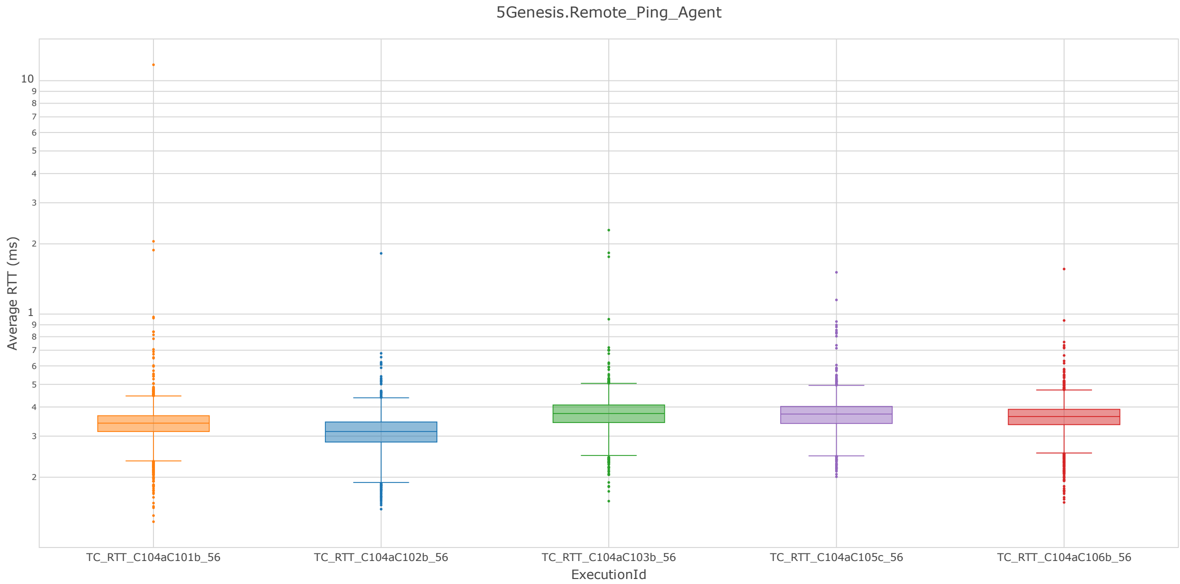

Figure 7-7 Inter-Compute RTT, 104 with 56 Byte payload

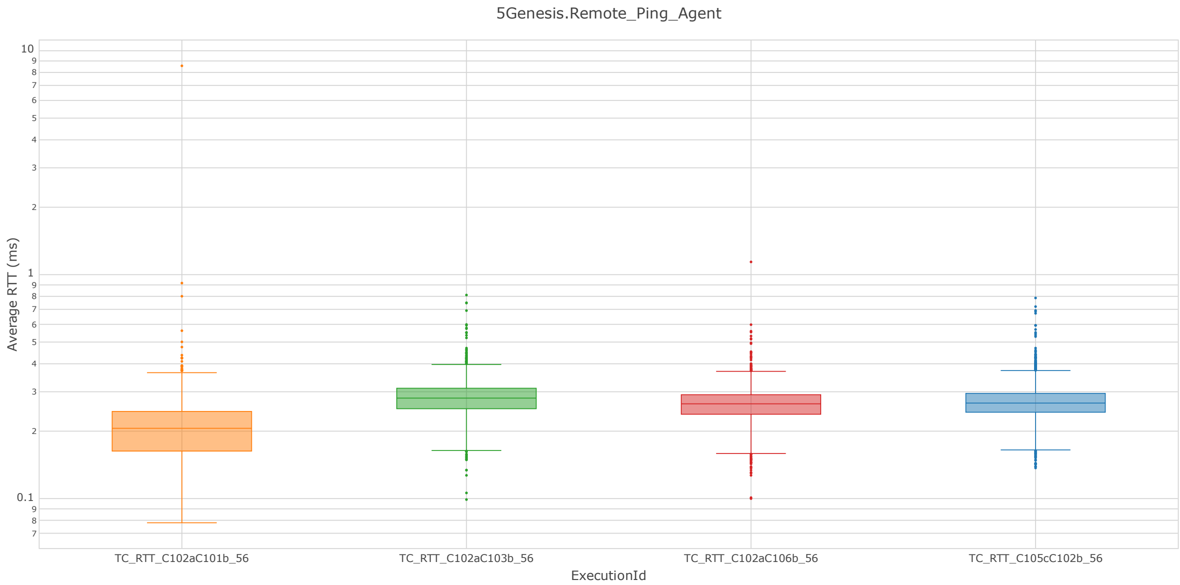

Figure 7-8 Inter-Compute RTT, 102 with 56 Byte Payload

7.2.5. Inter-Compute RTT Reliability

The reliability of the compute node interconnection with regards to RTT was assessed with a series of RTT measurements running over a longer time. For each measurement, RTT was recorded for two hours. Each experiment was repeated at least twelve times, resulting in 24 hours overall. The links evaluated include those between compute nodes C101 and C104, C102 and C104, C103 and C105 as well as C106 and C102. Figure 7-9 shows measurement results as a Cumulative Distribution Function (CDF). The plot was truncated to focus on measurement values below 1ms. To allow an additional visual investigation of outlier values, Figure 7-10 is provided.

One can observe that all links provided a sub millisecond RTT over 99% of the time. Therefore, it can be concluded that the inter-compute connection is highly reliable, in regard to RTT, and capable of supporting various 5G use cases, especially those that require low latencies.

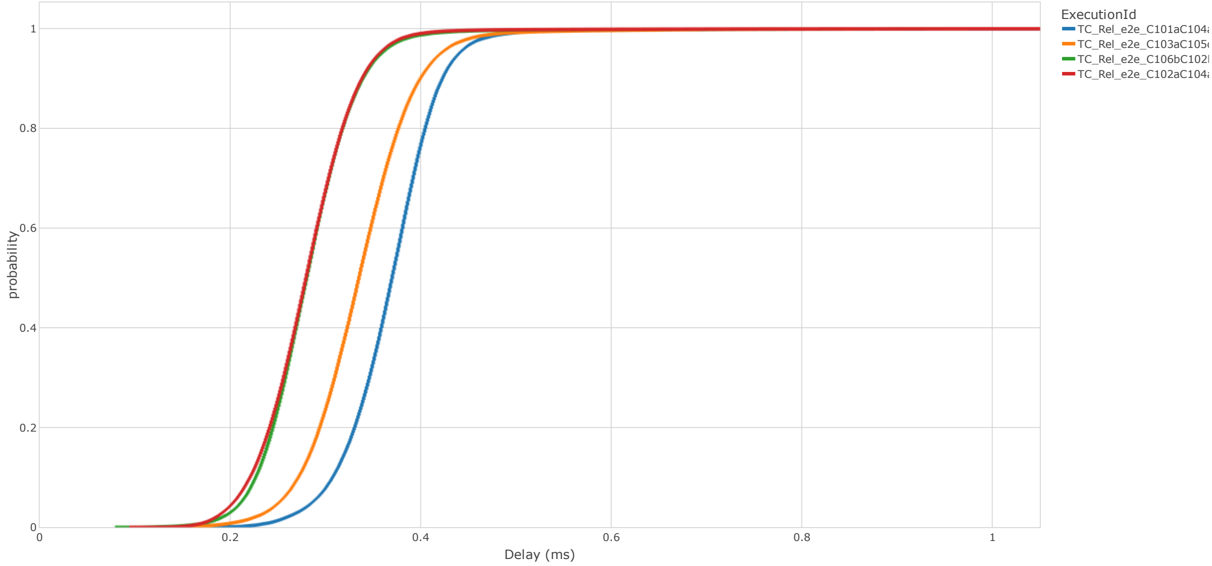

Figure 7-9 RTT Reliability for the C101-C104 case as CDF (limited to 1ms)

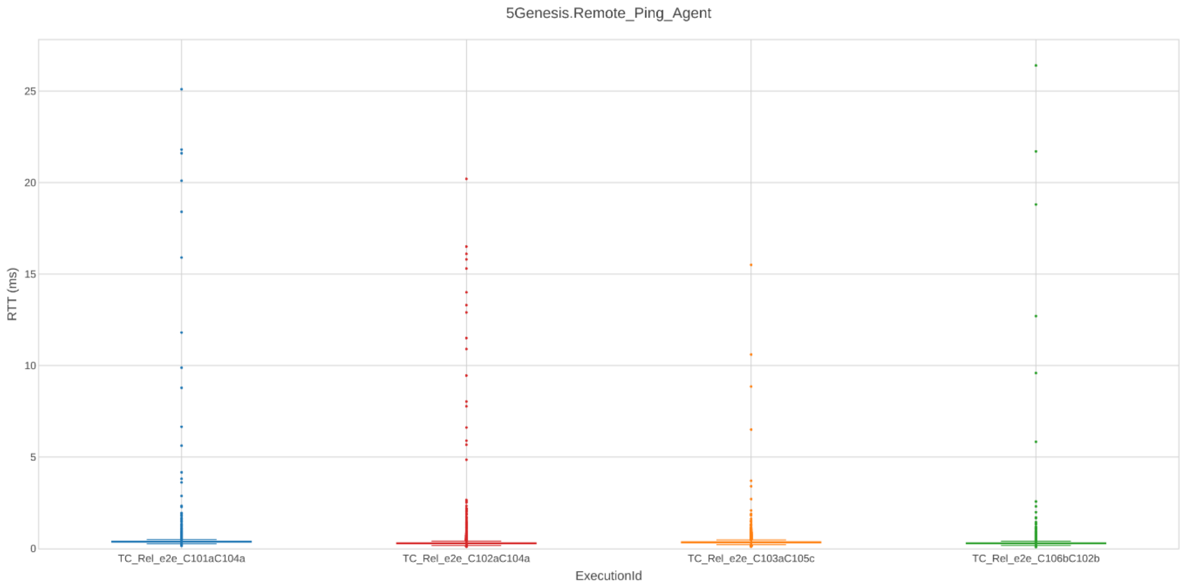

Figure 7-10 Comparison of outlier values for all RTT Reliability cases

7.2.6. Inter-Compute Throughput (TCP)

Additional experiments have been executed to analyze the throughput capacity of connections that span multiple compute nodes. For those Inter-Compute throughput experiments, a pair of VMs running on separate compute nodes are created. Thus, the compute cluster's backplane is used to provide the networking connectivity between VMs.

For the first set of experiments one VM will be placed on compute node C104. It that acts as a source anchor point. The target compute node is moved between nodes C101 to C106, excluding C104. The second part of executed Inter-Compute experiments vary the number of executed measurements that are run in parallel: two in parallel running measurement streams (Figure 7-12) and three in parallel running streams (Figure 7-13). Each experiment was repeated 25 times.

The following figures depict the measurement results. Please note that the y-axis starts at origin 0, for a consistent visual comparison of average throughput capacity results depicted throughout this document. Details regarding variance of results can be found in the appendix of this document.

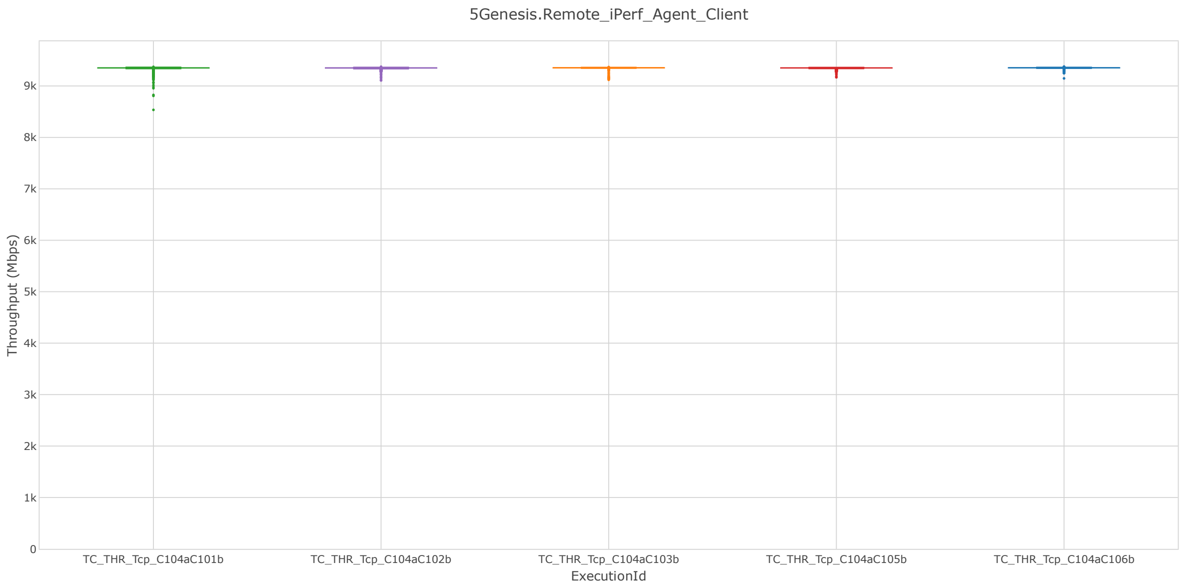

Figure 7-11 Inter-Compute Throughput non-parallel, C104>{C101 … C106}

Figure 7-11 provides the details for the first set of executed experiments for a single TCP stream. The results show no significant difference in average throughput capacity. The average value is in the range of 9.3 Gbps (mean value). The values show only slight differences in variance.

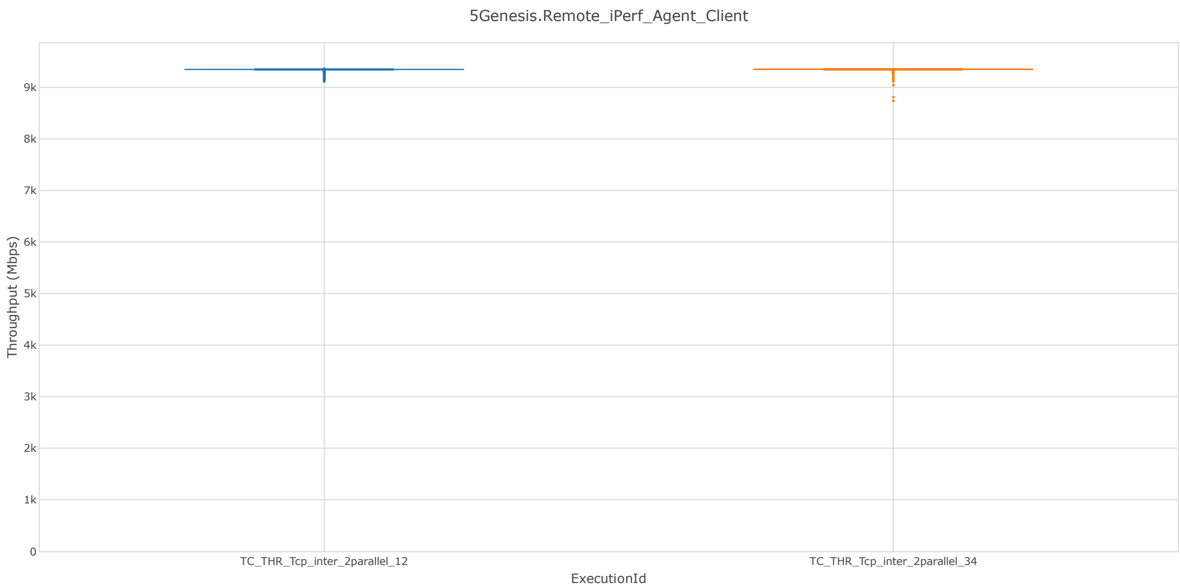

Figure 7-12 Inter-Compute Throughput, two parallel streams, C101>C102; C103>C104

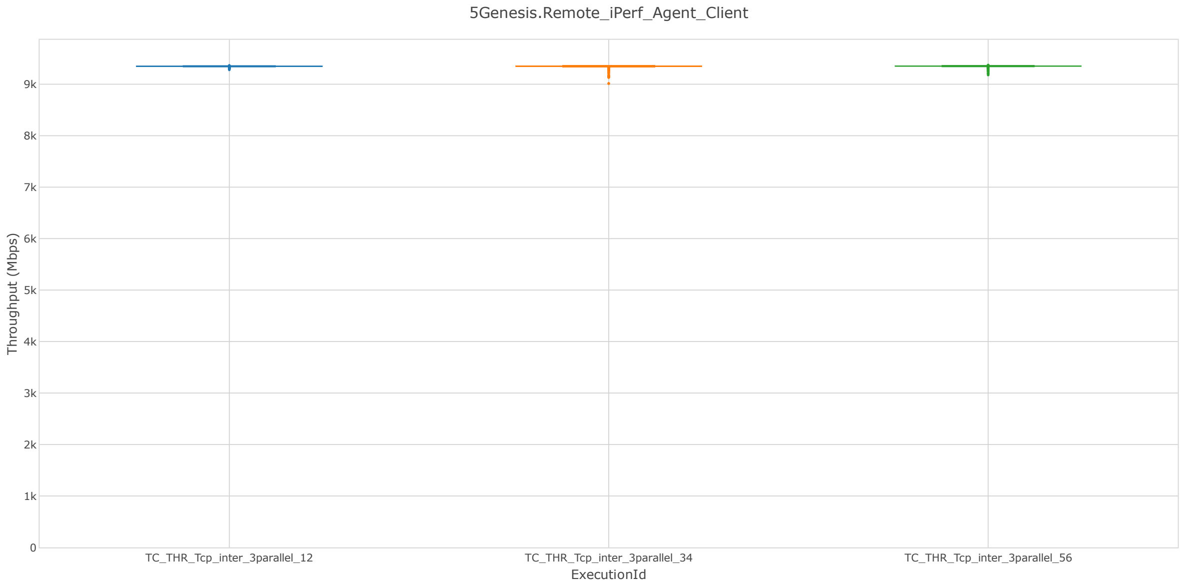

Figure 7-13 Inter-Compute Throughput, three parallel streams, C101>C102; C103>C104; C105>C106

For the case of two in-parallel running streams (Figure 7-12), roughly 9 Gbps are observed for both measurement streams. Similar values are observed for the case of three in-parallel running streams (Figure 7-13). It should be noted that, for each parallel measurement, each endpoint VM runs on a separate compute node. Therefore, the only shared component is the compute cluster's networking backplane.

Combining the results of all Inter-Compute throughput experiments, it can be stated that the backplane is capable to provide roughly 9 Gbps. In addition, if multiple experiments run in parallel, the backplane can sustain multiple parallel connections and provide each one with roughly 9 Gbps throughput capacity.

7.2.7. Conclusion on Generic tests

The generic measurements primarily targeted the validation of the underlying compute, storage and networking infrastructure. They assessed in infrastructure's suitability as a 5G and beyond testbed allowing the evaluation of end-to-end 5G network KPIs.

It can be concluded that the throughput and RTT performance exceeds that expected of an end-to-end 5G SA network. Thus, the underlying infrastructure does not impose any bottlenecks when instantiating the 5Genesis framework and when executing the 5G KPI evaluations.

In detail, the measurements show that the performance of the system significantly increases for end-to-end connections having both endpoints on the same compute blade in the virtualization environment as compared to having two endpoints located at different compute nodes. As the placement of VMs is at least partially automatic and influenced by load-balancing (see discussion on "auto migration" in section 7.2.1), having two end-points on the same compute node is mostly a theoretical scenario. Rather, the observed performance when placing VMs on different compute nodes matters as a limiting performance factor. And here, the system is well capable of providing mean RTTs of less than 3.5ms, which translates into a latency/delay of less than 1.75ms. Throughput between two VMs is sustained at around 9 Gbps and it is not affected by several parallel VMs communicating with each other, thanks to the high capacity of the switching and compute backplane.

7.3. UC#1: 360° camera – Measurements and results

In line with the field trial during the Berlin Festival of Lights 2019 (Section 7.1.2 in [8]), a 360° camera use case was executed during Phase 3 trials at IHP premises in Frankfurt (Oder). For this trial, FOKUS deployed its nomadic node on the campus as captured in Figure 7-14. The 5G NR antenna on the rooftop is seen in Figure 7-15. Figure 7-16 provides a closer look at the nomadic node hardware prior to the deployment. The photos of the deployment shown here complement the information provided in deliverable D4.15 [7].

The trial infrastructure, as depicted in Figure 7-17, spans over two main sites: IHP (yellow/orange and blue boxes) and Fraunhofer FOKUS (green box). Both sites are interconnected via a tunnel provided by the German National Research and Education Network (DFN). For the field trial, all management components of the 5Genesis testbed, including the coordination layer and the database for test results, run at the FOKUS site. For the trial, a full 5G Core is deployed on an edge-compute node residing at the IHP side. This core controls the 5G SA RAN installed at IHP for the trial. Thus, the baseline deployment for the trial allows to assess local, nomadic edge-based deployments, as well as connections terminating at a remote location, i.e., the FOKUS site.

Figure 7-14 FOKUS Nomadic node at the IHP campus in Frankfurt Oder

Figure 7-15 5G NR at IHP's rooftop (Wing C)

Figure 7-16 Nomadic Node – Compute, Storage & UPS

Details about the IHP site deployment can be found in section 2.1.3.1. "IHP's 5G Campus Network" of deliverable D4.15 [7].

Figure 7-17 visualizes how the 5G RAN is accessed by the UEs on the left-hand side. This includes the 360° camera (via CPE), two dedicated measurement endpoints (Probes D and G) and the cell phones of some of the trial participants. In contrast to this, the cell phones of the trial participants on the right-hand side will leverage the connectivity provided by a dedicated Wi-Fi access network.

The 60GHz backhaul link is used to provide a wireless connection between two different wings of the IHP building. It is used for all network data traffic, thus allowing the evaluation of a 60GHz backhaul link. The depicted wired connection in-between buildings is only used to provide remote management access to deployed nodes in case of failures of the wireless 60GHz link. It is not involved in any performance measurements.

A full set of measurements has been executed. For each type of measurement, a different part of the overall infrastructure can be considered as the SUT. Within this subsection, only the main important results are detailed, additional results are provided in the Annex.

In addition to the trials executed at IHP, complimentary controlled experiments of 360° video performance have also been executed at Fraunhofer FOKUS, as detailed in the last subsection.

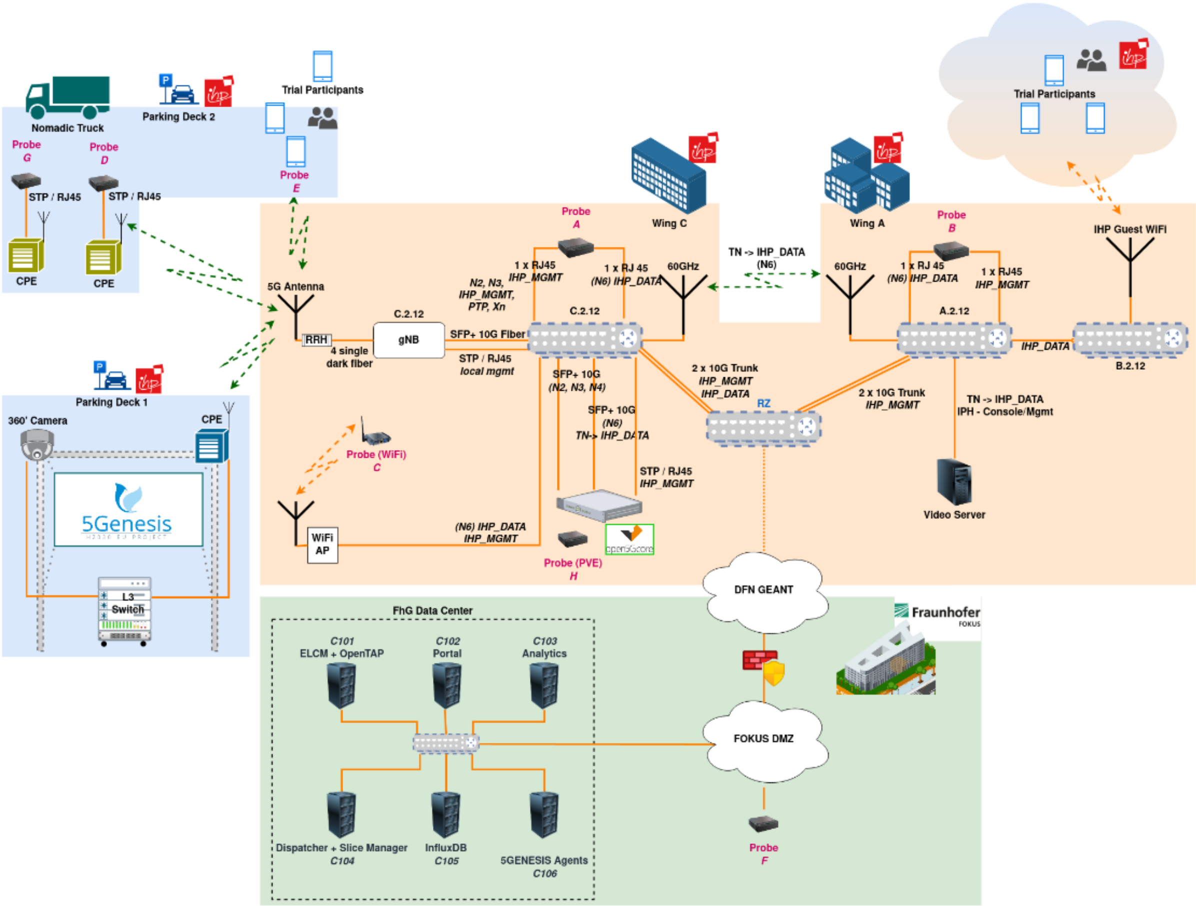

Figure 7-17 Final Trials Setup with 360° Camera

7.3.1. 60 GHz backhaul link

The 60 GHz link provides a wireless backhaul connection between the 5G core and the video server. Analyzing its capabilities in detail allows insights into possible interferences of the end-to-end connection and other measurements. The experiments described in this subsection are narrowed down to only the 60 GHz backhaul link. The KPIs analyzed here are the throughput capacity of the link and its latency.

For latency, in line with the baseline measurements in section 7.2, the RTT between two hosts is used. RTT measurements were executed from A to B and B to A, by leveraging ICMP echo/reply messages. For the measurements, the ICMP message payload is varied with the values of 32 bytes, 56 bytes and 1400 bytes. Each experiment was repeated 25 times.

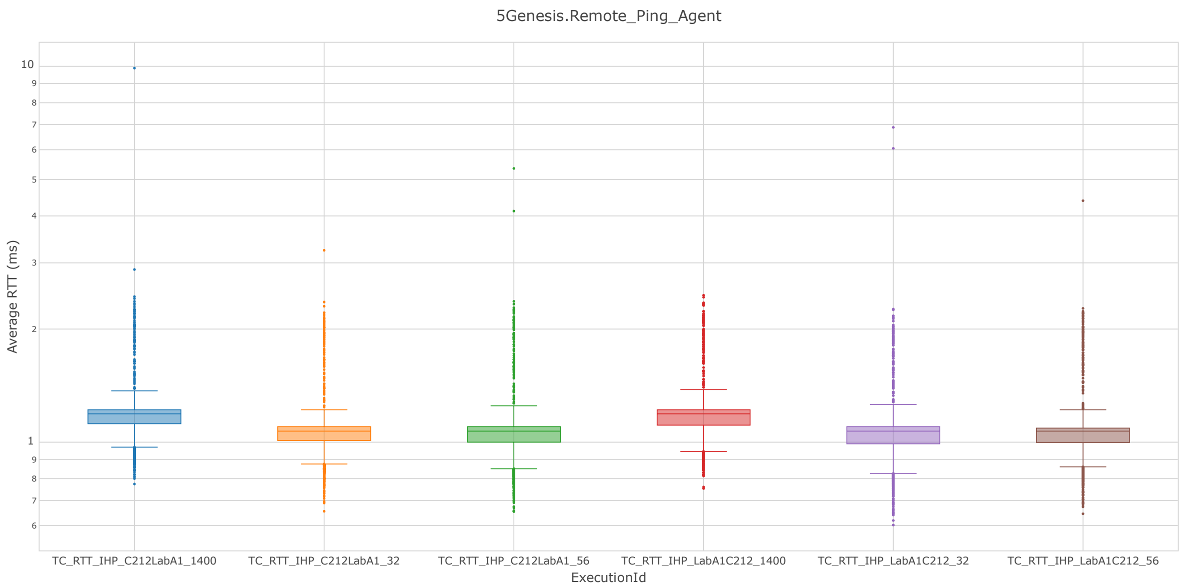

Figure 7-18 RTT experiments for 60 GHz Backhaul both directions

Figure 7-18 reveals the results of the RTT experiment for both directions and all three payload sizes. The median RTT is consistently between 1.1 and 1.2ms. Increasing the payload size to 1400 Byte causes a median increase of 0.1ms. The direction of traffic does not affect the RTT.

To complement the RTT experiments, throughput experiments of the 60 GHz backhaul link were executed by leveraging TCP streams. Each stream trying to maximize its throughput. Again, each of these experiments was repeated 25 times.

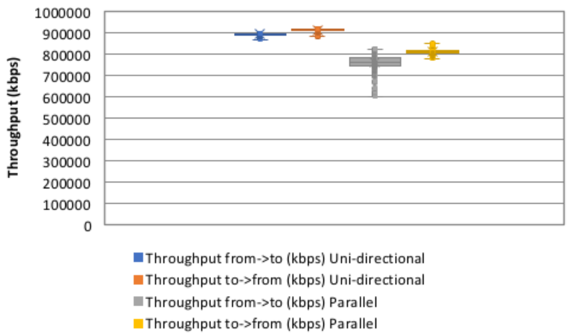

Figure 7-19 Throughput A>B, B>A (single & parallel)

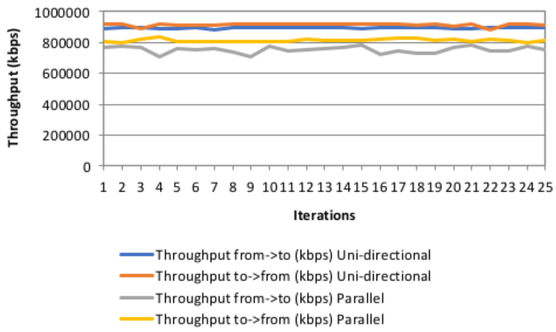

Figure 7-20 Throughput A>B, B>A (single & parallel)

Figure 7-19 and Figure 7-20 depict the results of these experiments. Here, the blue plot depicts a unidirectional stream from A to B, while the orange plot depicts a unidirectional stream from B to A. These two experiments were executed in sequence. Both directions reach roughly 900 Mbps. Interestingly, the figure reveals a difference in throughput capabilities when it comes to the direction of the data stream. However, the difference is marginal.

In addition, the grey and yellow plots in this figure depict similar unidirectional measurements, but this time both data stream experiments were executed in parallel. A reduced throughput, when compared to single measurements, is observable – but only by roughly 100 Mbps. Also, for the parallel measurements a similar difference in throughput capacity in relation to the used direction is observable. Moreover, in the case of A to B (parallel) stream execution – grey plot – a noticeable increase in variance can be observed.

7.3.2. 5G RAN and Core Network

Starting with this subsection, the experiments provided incorporate 5G network components in the analysis. Again, the main approach here is to extend the path between two measurement endpoints step-by-step. As an initial step, the experiments in this subsection are executed in a way that they entail the connectivity between a user equipment endpoint (UE) to a measurement endpoint directly deployed after the 5G network. Two different kinds of 5G devices are considered (a) a CPE and (b) a cell phone as UE. The connectivity path is shown in the figure below.

Like the last subsection, the main capabilities that are being analyzed are: throughput capacity of the selected connections, as also latency. Again, latency is described via RTT. For the RTT measurements, ICMP echo/reply messages are used and the ICMP message payload size is varied. Unlike most of the measurements in the previous subsections, only payload sizes 32 and 56 bytes are used. Throughput measurements are executed again with a main data transfer direction and switched over the measurements to cover both data transfer directions. As opposed to most of the previous subsections, no parallel throughput measurements are executed. All experiments are repeated 25 times.

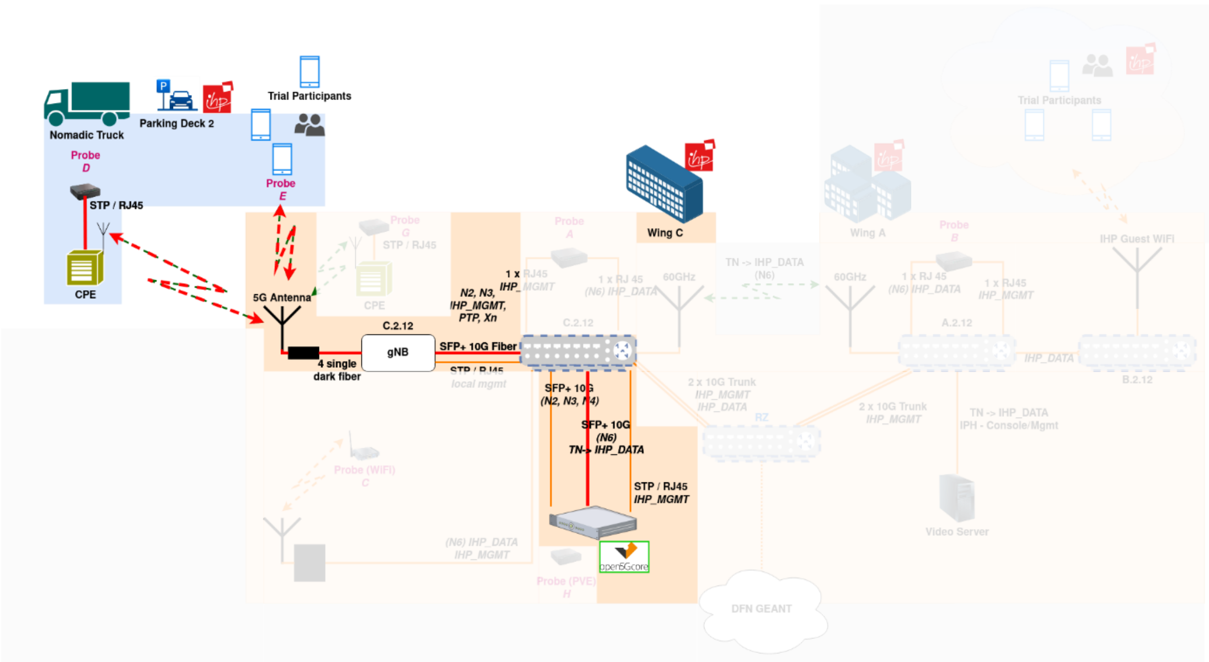

Figure 7-21 5G RAN and Core Network, at IHP

7.3.2.1. 5G CPE to Network

The first case analyzed is the CPE connection D to A. Here, no ICMP packet loss was observed in any of the executed experiments. In addition, the observed Minimum RTT values are in the range of 5 to 7ms (32 bytes payload size) and 5 to 6ms (56 payload size).

For the case of 56 bytes payload size, a Maximum RTT in the range of 35 to 115ms was observed. A stark difference for the case of 32 bytes payload size is visible in the same figure, when it comes to outliers. Here, for the 32 bytes case, the overall range is 35 to 395ms. It should be noted that the majority of observed results in the case of Maximum RTT only differ slightly. Finally, in regard to Average RTT, the 32 bytes payload size case also shows a higher deviance in values (due to the wide spread of outliers visible in the Maximum RTT) then maximum deviance of Average RTT values in the 32 bytes payload size case. In summary, the majority of values of Average RTT are all in a quite similar range of about 15 to 17ms.

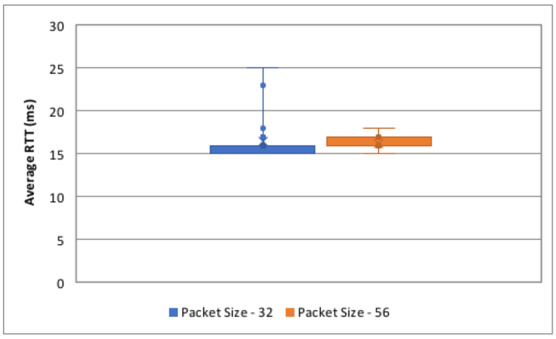

Figure 7-22 Average RTT D->A (32 and 56 byte payload)

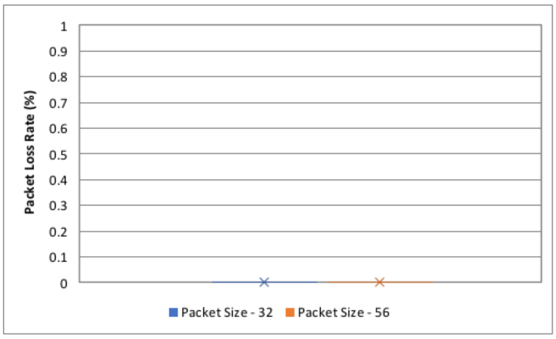

Figure 7-23 RTT Loss Rate D->A (32 and 56 byte payload)

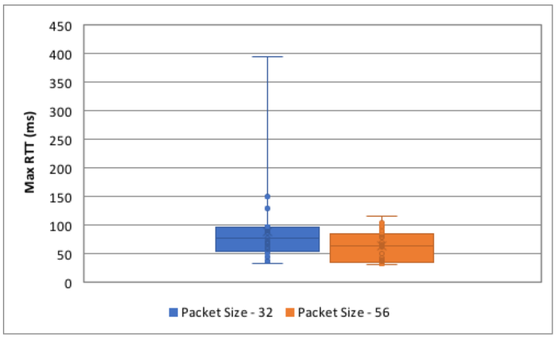

Figure 7-24 Maximum RTT D->A (32 and 56 byte payload)

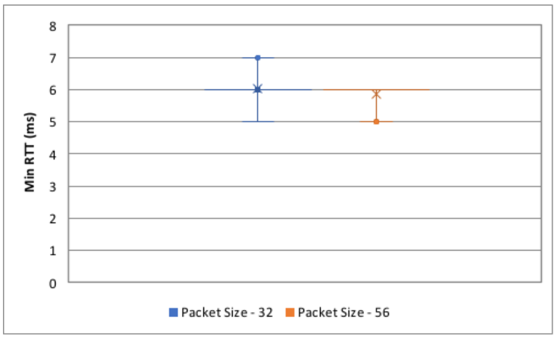

Figure 7-25 Minimum RTT D->A (32 and 56 byte payload)

7.3.2.2. 5G Smart phone to Network

To allow investigating throughput capacity, the UE connection E to A was used. The results of those throughput measurements are shown in Figure 7-26 and Figure 7-27 (line plot that holds values for each experiment iteration). The measured throughput values show an expected difference when it comes to data transfer direction. In the case of E to A, an average throughput of 80 Mbps is visible, while in the case of A to E an average throughput of about 575 Mbps could be observed. A slight difference is also visible when it comes to the variance of values for both cases. The E to A case (blue) shows only little variance, whereas the A to E case (orange) shows a reasonable variance of measurement values from about 520 to 605 Mbps. The slight differences in variance of measurement can be also observed in Figure 7-27 (as line plot).

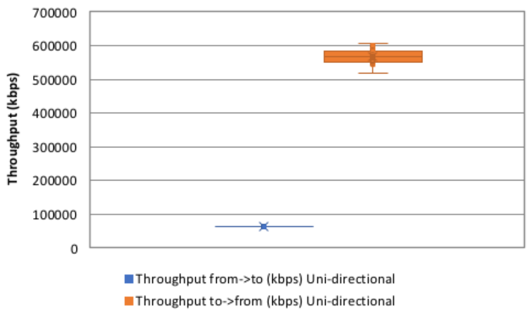

Figure 7-26 Throughput E>A, A>E (single only)

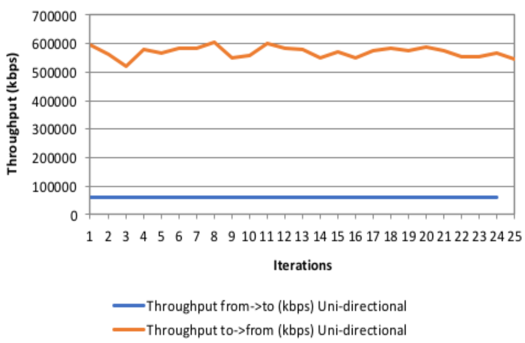

Figure 7-27 Throughput E>A, A>E (single only)

7.3.2.3. 5G UE-to-UE connectivity

The experiments shown in this subsection focus on UE-to-UE measurements. Therefore, both selected endpoints of the given experiments are connected via the 5G network, deployed at the IHP facility.

For this subsection, the main capability that is analyzed of the selected UE-to-UE connection is RTT. Again, ICMP echo/reply messages are used, and experiments are varied in two payload sizes: 32 and 56 bytes. All of the executed experiments have been repeated 25 times.

As can be seen in Figure 7-29, for the D to E connection, only in the case of 32 bytes payload size a loss rate of 1% is observed, for the case of 56 bytes payload size no loss was observed at all. For Minimum RTT (Figure 7-31) both cases of payload sizes are in a similar range of 16 to 19ms. For Maximum RTT (Figure 7-30) a stark difference in outliers can be observed. Here, in the case of 32 bytes payload size, the overall range of values is about 51 to 475ms. Whereas, the case of 56 bytes payload size range of values (with outliers) is in the range of 51 to 140ms.

In Figure 7-28 is visible that the majority of Average RTT values, for both cases of payload size, are roughly in the range of 30 to 33ms. Only the case of 32 bytes payload size shows a bigger variance in (maximum/minimum) values.

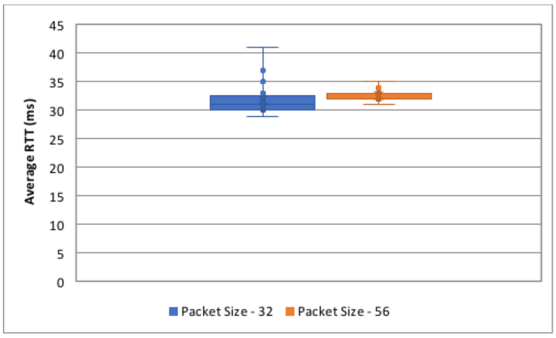

Figure 7-28 Average RTT D->E (32 and 56 byte payload)

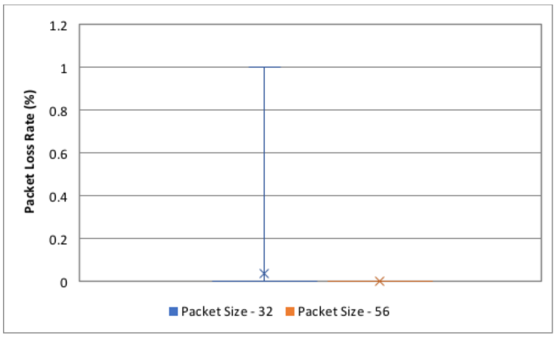

Figure 7-29 RTT Loss Rate D->E (32 and 56 byte payload)

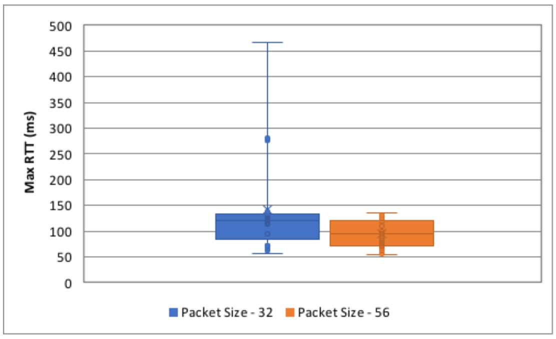

Figure 7-30 Maximum RTT D->E (32 and 56 byte payload)

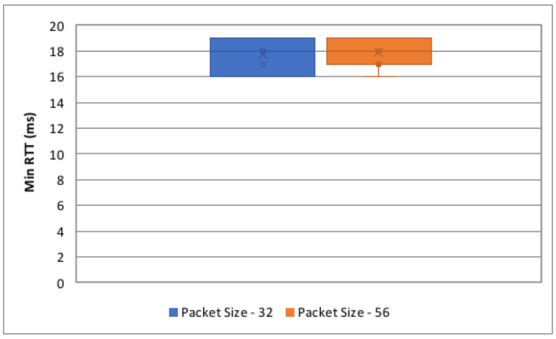

Figure 7-31 Minimum RTT D->E (32 and 56 byte payload)

7.3.3. 5G Network with 60 GHz backhaul link

In this subsection the 5G-based connection shown in section 7.3.2 is extended via the 60 GHz backhaul link. This time, again, one of the selected endpoints is some type of 5G user equipment connected via the 5G network (at IHP), but unlike subsection 7.3.2, the opposite measurement endpoint resides in the IHP network, behind the 60 GHz backhaul link. Like subsection 7.3.2, two types of 5G user equipment are considered as UE, (a) a CPE and (b) a cell phone. Note, while throughput capacity measurements were varied in the data transfer direction, the executed RTT experiments are varied in payload size. Experiments are again repeated 25 times.

7.3.3.1. 5G CPE

First a 5G CPE type end device is analyzed. Figure 7-32 depicts the topology of this experiment and shows what components are involved in the overall connection between endpoint D and B. For first connection only RTT experiments were executed, with varying payloads sizes of 32 and 56 bytes.

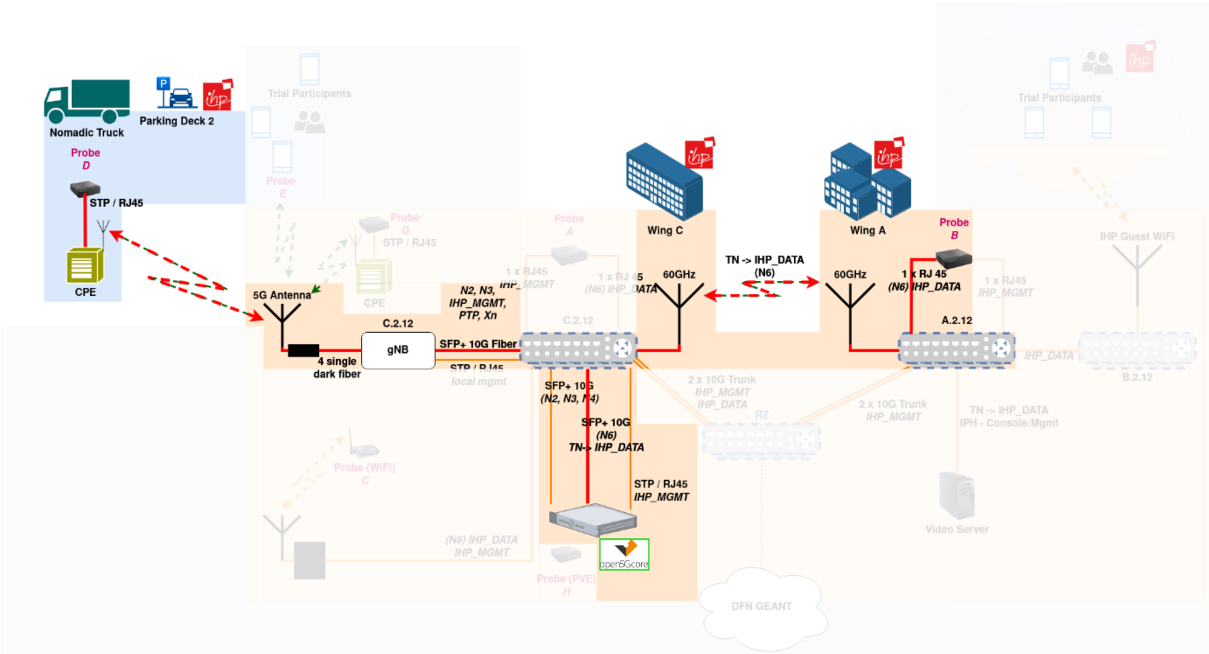

Figure 7-32 5G Network with 60 GHz backhaul link (Part 1)

Regarding average RTT (Figure 7-33) there is also a slight difference visible – for the majority of values – when the two cases are compared. Here, the case of 32 bytes shows a main range of 16 to 17ms, whereas the case of 56 bytes has a slightly higher range of majority of values, at about 17 to 18ms. Like the case of Maximum RTT, also for Average RTT the case of 32 bytes shows a higher variance of values (up to 26ms).

No ICMP packet loss was observed for both cases of payload sizes (see Figure 7-34).

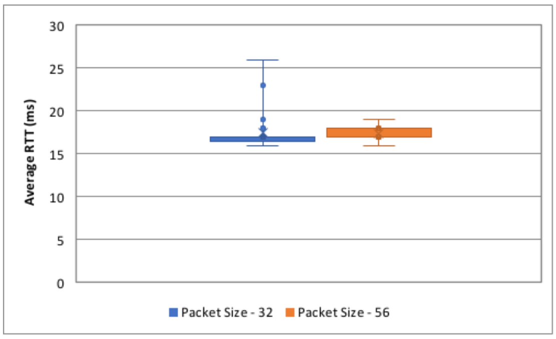

Figure 7-33 Average RTT D->B (32 and 56 byte payload)

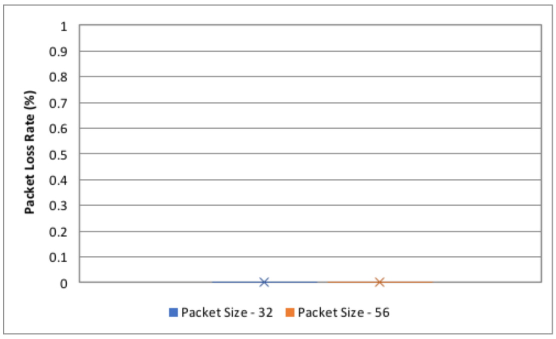

Figure 7-34 RTT Loss Rate D->B (32 and 56 byte payload)

As Figure 7-36 reveals, the observed Minimal RTT for both cases are both in the range of 6 to 7ms. For the Maximum RTT (Figure 7-35) the measured values for both cases differ. Here, the majority of observed values show only a small drop of about 10ms for the case of 56 bytes payload size (values range roughly between 49 to 90ms) when compared to the case of 32 bytes (values range roughly between 60 to 100ms). Though, there are a couple of extreme outliers especially in the case of 32 bytes payload size (up to 349ms Maximum RTT). In comparison to that the outliers for the case of 56 bytes payload size only show maximum outliers of up to 175ms.

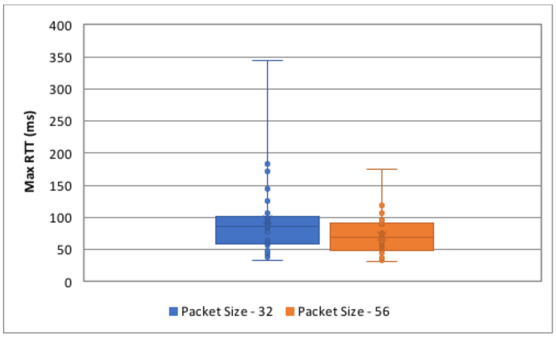

Figure 7-35 Maximum RTT D->B (32 and 56 byte payload)

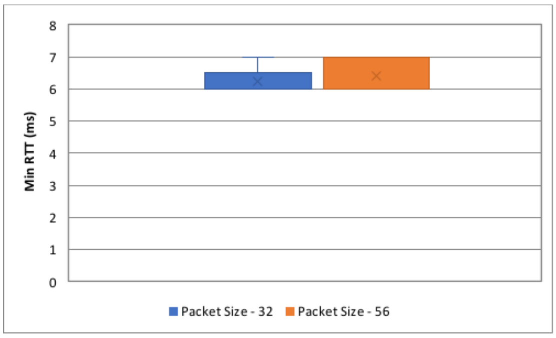

Figure 7-36 Minimum RTT D->B (32 and 56 byte payload)

7.3.3.2. 5G Smart phone

Now, for part 2 of this subsection, a 5G capable cell phone was employed as user equipment. The figure below shows the connection path between endpoint E and B (behind the 60 GHz backhaul link). For this connection, throughput capacity is analyzed. The executed experiments are varying the main data transfer direction. Please note that there are no in-parallel executed throughput measurements for this subsection.

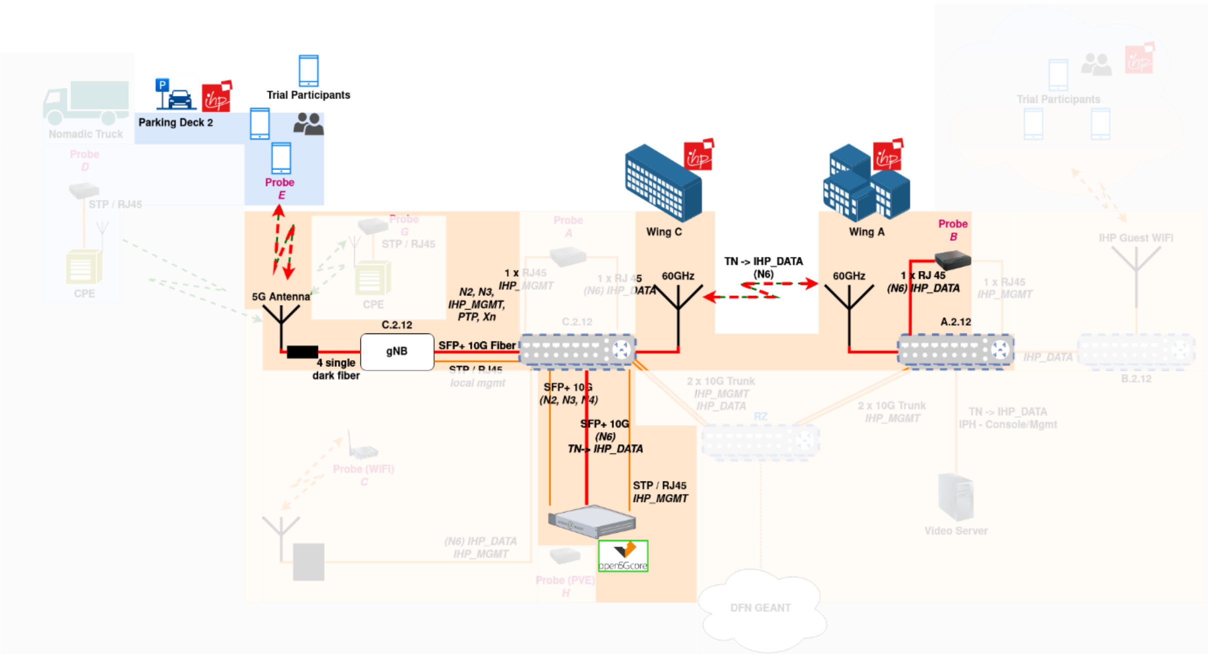

Figure 7-37 5G Network with 60 GHz backhaul link (Part 2)

For the case of throughput capacity experiments values are shown in Figure 7-38 and 7-39 (line plots that depict individual experiment iteration values). First of all, again there is a stark difference in data transfer direction visible. For the case of E to B (blue, 5G uplink direction) significant lower average values (roughly in the range of 45 to 65 Mbps) are visible, whereas in the case of B to E (5G downlink direction) the value range is 480 to 520 Mbps (as typical values). For both cases of data direction also individual data rate drops are visible. In the case of E to B the data rate dropped to nearly 0Mbps. In the case of B to E the data rate dropped to about 180Mbps.

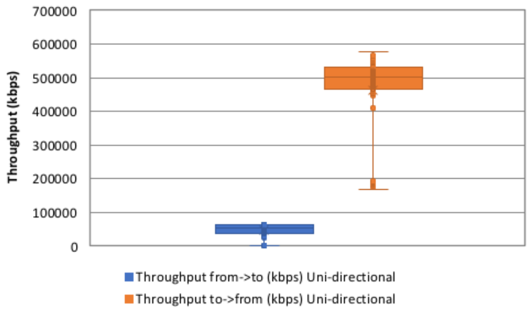

Figure 7-38 Throughput E>B, B>E (single only)

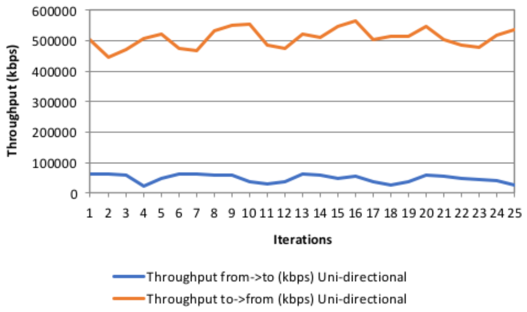

Figure 7-39 Throughput E>B, B>E (single only)

It should be noted that the huge drops of data rate (for both cases) do not show up in the single iteration line plots, because these plots only show values aggregated per iteration. The drops of data rate were only visible in the unaggregated measurement data.

7.3.3.3. Analysis and conclusions

Since the overall path of the above described E to B connection constitutes of path elements of the 5G core network as also the 60GHz backbone link, a comparison of the observed values RTT values of their analysis is advised. For executed RTT experiments of this subsection as also for the RTT experiments in the 5G baseline case (7.3.2), as also in the case of 60GHz backbone link baseline (7.3.1), no ICMP packet loss could be observed.

Now, let us have a look at whether the observed RTT of the E to B connection path, add up with the observed RTTs of its (main) constituent path elements.

Let us first investigate Average RTT. For the backbone link, the observed Average was roughly 1ms – though most values were in the sub microsecond. For the 5G baseline measurements the Average RTT was roughly in the range of 15 to 17ms – for all payload sizes. Now, the observed Average RTT for the whole E to B connection was roughly in the range of 16 to 18ms. For the case of 32 bytes a higher variance of values could be observed, this is also similar in the 5G baseline case.

As second, let us investigate Minimum RTT. Since the observed RTT values for the backbone link were low as into the sub microsecond range they can be ignored here. The observed Minimum RTT for the 5G baseline measurement experiments were in the range of 5 to 7ms (both payload sizes merged together). Again, the observed measurements in the same case for this subsection were in the range of 6 to 7ms.

As third and last, let us compare Maximum RTT. While for the case of backbone link Maximum RTT outliers of 26 ms (32 bytes) and 17ms (56 bytes) are visible in Figure 7-18, the majority of values, for both payload sizes, are roughly in the range of 1 to 3ms. Now, within the 5G baseline experiments Maximum RTT outliers of up to 395ms (32 bytes) were observed but the majority of values of 32/56 payloads were roughly in the range of 37 to 98ms.

For the E to B connection, the majority of Maximum RTT values are roughly in the range of 49 to 100ms. Here, outliers differ regarding used payload sizes.

In summary: there is no observable correlation or cross-influence of the per-segment links in the analyzed end-to-end connection. The observed values of the baseline experiments add up with the values analyzed in this subsection. The only difference to highlight is that the observed maximum outlier value (for 32 bytes) – in the case of this subsection – was lower than for the 5G baseline experiments. Overall, the E to B RTT/latency property seems mainly influenced by its 5G part. The 60 GHz backhaul link provides a stable extension to the connection with only marginally increased RTT/latency.

As was analyzed in 7.3.2, the throughput capacity of the 5G connection part was sensitive to the data transfer direction. For the upload direction (from 5G user equipment to endpoint after 5G core network), an average throughput of 80 Mbps were observed. For the download direction case, an average throughput of 575 Mbps could be observed. In comparison, the 60GHz backbone link analyzed in 7.3.1 showed no stark difference in data transfer direction. Here the observed throughput, for both directions, was in the range of 900 Mbps.

Now, we will compare the observed values of the E to B (upload) and B to E (download) throughput cases of this subsection with the above described baseline cases. As it seems, the B/E connection’s throughput capacity is mainly influenced by the 5G network part of the connection. Similar to the 5G baseline cases also the B/E connection is sensitive to the given data transfer direction. In regards to the observed throughput, the average of the 5G baseline tests (80 Mbps) is not reached but only a slightly lower value (45 to 65 Mbps). Also, in difference to the executed 5G baseline experiments, the values for the E/B connection are fluctuating higher. This can be seen best, when directly comparing Figure 7-27 with Figure 7-39 – both figures reflecting average values per iteration.

The lowered capacity performance of the E/B connection, in comparison with the 5G baseline cases, could be explained with the observed cases of throughput dropping to 0. While this would also explain the higher fluctuation in average values per iteration, the actual source of those to-zero-drops requires further investigation.

7.3.4. 360°-camera-to-video-server scenario

In this section, we summarize the panoramic live streaming field trial, as conducted at IHP on November 23 and 24, 2021. This use case targets at allowing a large number of users gaining access to a video live stream, via the 5G network that is deployed at IHP. As indicated in section 7.1.2, the majority of users that were involved in the trial were IHP employees that could connect to the 5G system using their smartphones. Additionally, the live stream was also accessed by involved researchers.

An overview of the whole setup is given in Figure 7-17. A 360° camera is located on the parking lot and provides a live stream to the media server, from where users (i.e. IHP employees) can access the video stream. The connectivity from the camera to the media server is achieved via access to the deployed 5G network. Please note that the camera is connected via Ethernet to a CPE which provides the actual 5G network access.

As mentioned above, the users can use their cell phones to access the media server and by that the video stream from the camera. Accessing the media server with cell phones is possible by two different options: (a) via the 5G network or (b) via IHPs in-house Wi-Fi network. Please note that, additionally to the user's cell phones, also tests were executed where the video live stream was accessed by CPEs, via the 5G network.

In the following, a short overview on this subsection is given: In the beginning, additional details of the experiment setup are provided, which is followed by a brief overview on the collected dataset. Afterwards, the dataset is analyzed in detail and results are discussed. The subsection is closed with a final conclusion for the provided analysis.

7.3.4.1. Experimental Setup

To carry out the field trial, two particular pieces of hardware equipment were employed: a) a media server, which played the dual role of web server and video transcoder, and b) a panoramic camera. While the media server was connected using an Ethernet cable to the local area network, the camera was connected via 5G, leveraging a CPE for the access. The media server was equipped with an Intel Core i7-7700K CPU, 32 GB DDR3-SDRAM, a 1TB NVMe SSD, and, most importantly, a Nvidia GeForce GTX 1080 graphics card. The panoramic camera, a Vivotek FE9391-EV, was equipped with a 12-megapixel CMOS sensor, which delivered video at a frame rate of 30 fps and a resolution of 2816 pixels squared. Its fisheye lens captured (minutely less than) 180 degrees horizontally and vertically and 360 degrees around the circumference, that is, a hemisphere. Requiring active PoE, specifically PoE 802.3at Class 4, the camera was connected to a PoE-enabled switch. The camera was located outdoors, mounted on a mast five meters above the ground.

To benefit from hardware acceleration, FFmpeg was compiled with the Nvidia Video ENCoding (NVENC) library. A patch was applied to the device drivers to remove a restriction particular to the consumer-grade GeForce series that allows maximum two concurrent output streams. Live transcoding requires one dedicated stream per bitrate, so having more than two streams available is a clear advantage. FFmpeg was configured to ingest the Real Time Streaming Protocol (RTSP) stream from the camera and transcode it at three different bitrate levels.

The HTTP Live Streaming (HLS) playlist, along with the video segments themselves, were stored in shared memory on the media server. Circumventing storage on disk speeds up read and write operations by at least one or two orders of magnitude and does not incur any drawbacks as long as recording is not needed. A symbolic link on disk pointed to the base directory, which contained the manifest and one subdirectory per stream. Same-bitrate segments were sorted into their respective subdirectory, containing as well the initialization segment and the stream-specific HLS playlist. Nginx was employed as the web server. A simple design for handheld devices was put together with HTML and CSS.

The web application was built with Video.js: a free and open-source HTML5-based video player. Video.js was configured with the official Virtual Reality (VR) plugin, which adds the capability of displaying panoramic video. This plugin supports two different projections: the equirectangular projection and the cube map projection. Since the camera only provides a fisheye projection, support for the camera's native projection was implemented ad hoc. This was done purely in Video.js, that is, on the client side, by reengineering the existing plugin. Viewport controls were available either through an auxiliary input device (mouse or touch screen) or motion sensors for head-mounted displays.

7.3.4.2. Data Collection

Data was collected by means of two software tools: a custom-built JavaScript/PHP web application that captured data on web sessions, geolocation, network connection at (primarily) the client side, and Bitmovin Analytics. The latter represents an API-driven video analytics system that provides insight on player performance, user behavior and other aspects of the video chain.

A large dataset consisting of viewing sessions of different users was collected. The summary of the different parameters of the dataset is presented in Table 7-2.

Table 7-2.Totals and averages pertaining to viewing sessions and users

Parameter	Value
Total number of viewing sessions	232
Total number of users	56
Total number of browsers	11
Total playback duration	22 hours, 45 minutes, and 48 seconds
Average number of viewing sessions per user	4.14
Average playback duration per viewing session	5 minutes and 53 seconds

7.3.4.3. Analysis of the Results

We first analyzed the video player states. We used Bitmovin Analytics that (as of version 2.5.4) defines nine different states, in which the video player can operate at any given moment. It records state transitions along with the duration of each intermittent state:

- Playing: The video player is displaying video.
- Paused: The video player has been paused by the user.
- Stalling: The video player has exhausted its buffer and is requesting more data.
- Setup: The video player is just starting up—not yet ready for playback.
- Startup: The video player has started, but not yet displayed its first frame.
- Seeking: The video player is seeking within its buffer to a particular position in the stream.

The **setup** state immediately precedes the **startup** state and, in the following, these two states are combined. They are denoted as the initialization state; the sum of their durations is denoted as the initialization time. Other states, notably **quality switch**, **error**, and **closed**, are recorded too, but these are momentary states and do not contribute to the overall duration—or do so only to a negligible degree. Thus, they are disregarded from this analysis.

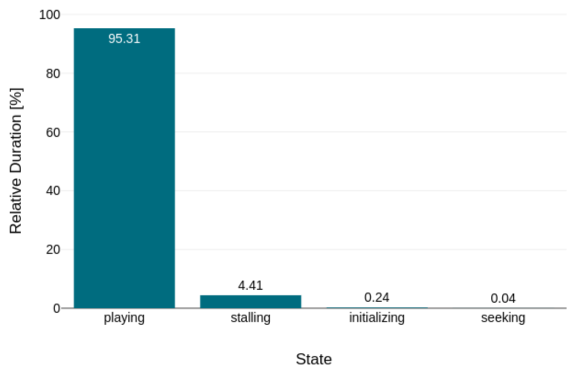

Figure 7-40 Relative duration of player states

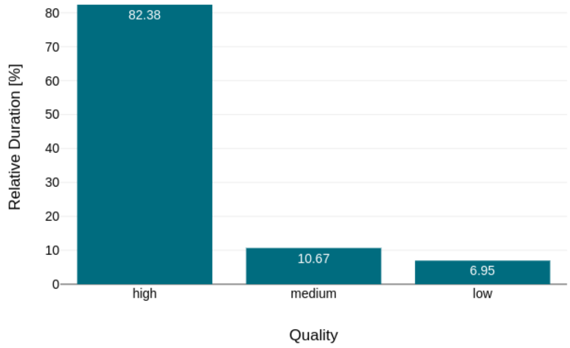

Figure 7-41 Relative duration of qualities

In Figure 7-40, we show the relative duration of the video player states over all the impressions. The video player spends a healthy 95.31 percent on playback, while stalls demand no more than 4.41 percent. Initializing and seeking constitute a negligible fraction. Other states are not shown on the plot. Pauses are discounted altogether.

We then analyzed the relative duration of the video qualities. Note that the stream is encoded at three different bitrate levels to accommodate varying network conditions. Therefore, the encoded bitrate of the stream is arguably the single most important factor inherent to the stream itself that influences the quality: it restricts how much information goes into the stream. Lower bitrate means, of course, lower quality, and vice versa. Hence, the relative duration of bitrate levels provides a direct measure of a decisive quality indicator.

In the playing state, the video player requests segments from the media server using an Adaptive Bit Rate (ABR) algorithm. ABR dynamically chooses the best quality level given the network conditions and the current buffering state. In this work, we use Apple HLS, which is one of the most common ABR streaming technologies supported by the vast majority of devices. We further use three different bitrates/qualities in our experiments: 1000 kbps (low), 4000 kbps (medium), and 6000 kbps (high). Figure 7-41, we illustrate the relative duration of qualities over all viewing sessions. By far the most time, 82.38 percent, is spent on the high level. The medium and low levels constitute 10.67 percent and 6.95 percent, respectively. This shows that the network can support the high-quality video for the majority of the viewing sessions.

Finally, we analyzed the frequencies of the quality switches and stalls. Note that a quality switch may occur in two circumstances: either if the video player finds that the current network capacity is insufficient for sustaining delivery of a higher-quality representation, or if the video player finds that network conditions have improved sufficiently to sustain a higher-quality representation than the one currently being served. In either case, the video player requests the representation that is most likely to provide uninterrupted playback of the highest possible quality. As long as bandwidth fluctuations are not too extreme, the video player should be able to continue playback without stalling—though the quality may vary.

The ABR algorithm attempts to buffer segments to aid uninterrupted playback even in the case of intermittent network outage. Of course, in a livestreaming scenario, buffering future segments is at odds with keeping the delay as short as possible. A stall will happen whenever the video player has exhausted its buffer and is no longer able to sustain playback. The video player then requests more segments of the lowest quality, while the user has to wait for new segments to arrive before he can resume the video. The stalls can be brief if caused by a transitory drop in bandwidth, or it may last indefinitely if the connection between client and server is severed. Stalls are arguably the single most detrimental factor to the viewing experience.

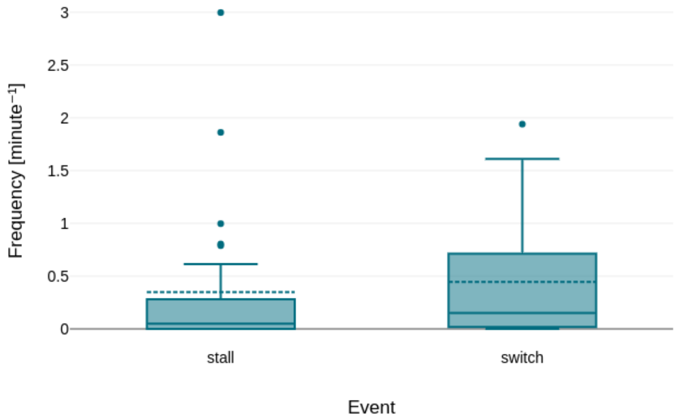

Figure 7-42 Distributions of frequencies of quality switches and stalls

We show the frequency of quality switches and stalls as distributions across viewing sessions in Figure 7-42. The median is marked by a solid line, the average by a dashed line, within each box. The maxima and minima are set at a distance of 1.5 times the interquartile range from the third and first quartiles, respectively. While quality switches are momentary, the stalls have a certain duration. Note, however, that Figure 7-42 does not illustrate the duration of the stall but only its frequency, i.e. any one stall may be brief or lengthy. Stall frequencies are distributed around a median of 0.05 stalls per minute, corresponding to one stall every 20 minutes. Quality switch frequencies, being more common, are distributed around a median of 0.15 switches per minute, corresponding to one switch every 6 minutes and 40 seconds. Some high-value outliers are seen, particularly with respect to stalls.

7.3.4.4. Discussion and Conclusion

The relative duration of the video player states show that, throughout the experiment, the video player was indeed operating normally more than 95 percent of the time. The relative duration of qualities also shows that most of the time is spent at high quality. This is expected as the network was never congested: 232 viewing sessions over several hours did not impose a heavy load; the used bandwidth was more than sufficient to supply the highest available quality. The medium and low qualities account for a comparatively shorter, but still notable, fraction of time. This is likely an effect of the adaptation algorithm, which often requests a lower quality at the outset of the viewing session—to get things going right away—and, subsequently, requests a higher quality if network conditions allow.

Stalls, even brief ones, are particularly undesirable. Although a few stalls at the beginning of the stream can be tolerated, many stalls scattered throughout the viewing session drastically reduces the viewing experience. Half of all viewing sessions experienced (the equivalent of) at most one stall in the course of 20 minutes; 75 percent of them experienced (the equivalent of) less than six stalls in the same time span.

Though not as bad as stalls, quality switches adversely impact the viewing experience. A few quality switches are expected, especially at the beginning of the stream. Sometimes the player needs to fall back to a lower level again. It may take some trial and error before it arrives at a stable choice. This behavior is apparent from the observation that many viewing sessions contain a number of quality switches higher than or equal to the number of available qualities, that is, three.

Overall, the player spends most of its time in playing state and the high-quality level dominates the sessions. While all these results are obtained in the absence of network congestion, the assessment of the quality criteria indicates that the system works satisfactorily and delivers good video quality.

7.3.5. 360° video experiments at Fraunhofer FOKUS

To evaluate the scalability of 5G in supporting 360º video, we have also run controlled experiments in the Fraunhofer FOKUS testbed. In this experiment, we used a setup at the FOKUS Laboratory and the MONROE measurement probe, called 360-dash (c.f. deliverable D3.6) that enabled us to emulate a 360º video client, and to perform controlled experiments; allowing us to evaluate the video quality over 5G under different network settings, as well as assessing the scalability of our solution.

Figure 7-43 illustrates the testbed employed in this experiment. A DASH HTTP video server, running nginx, was connected to the testbed through the Tenant network. A select of 360º video files were hosted on the DASH server. The client, 360-dash, ran within MONROE VN and was connected via a CPE to the actual 5G network via the CPE-n network. (The second data path through the emulated gNB provided by the Open5GCore via the CPE-1 network was not used in this experiment.) Tests were run with 1, 10 and 50 DASH clients, and each test was repeated 30 times. In each test run, a 360º video hosted at the server was streamed to the 360-DASH clients, and the delivery and video representation rates as well as the number of rebuffering events were measured at the respective client.

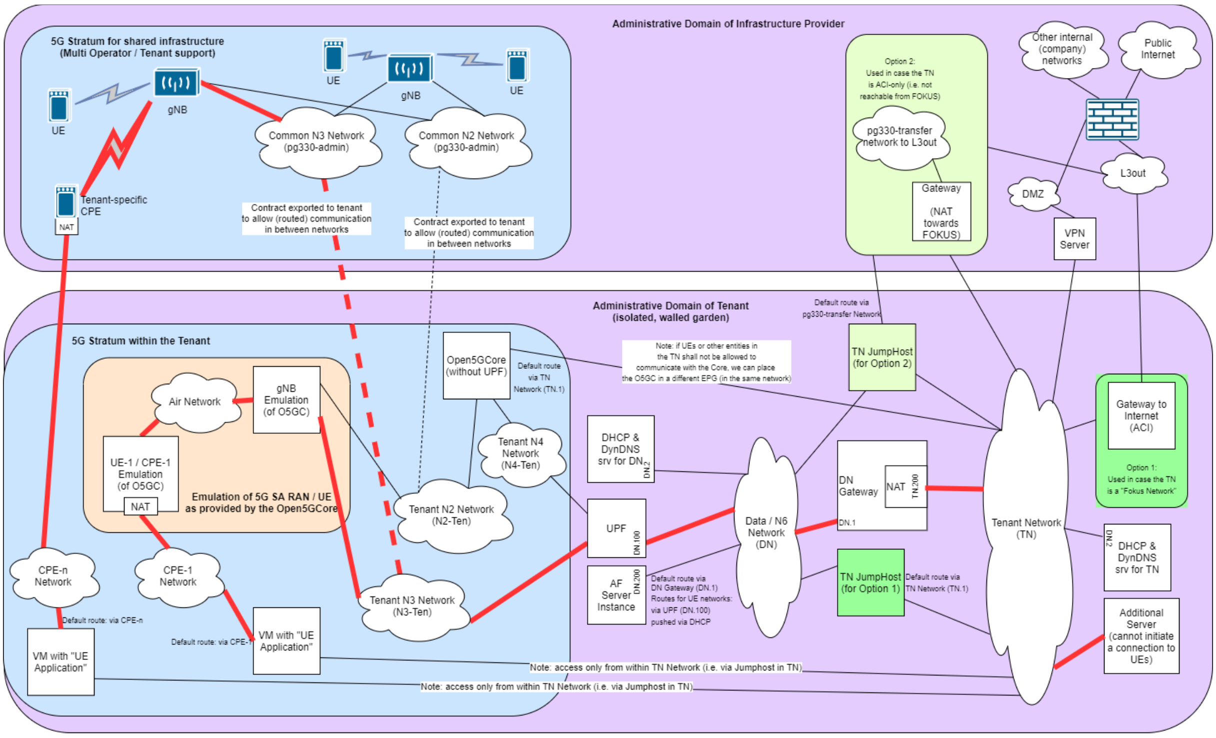

Figure 7-43 Testbed infrastructure setup for 360º video use case at Fraunhofer FOKUS.

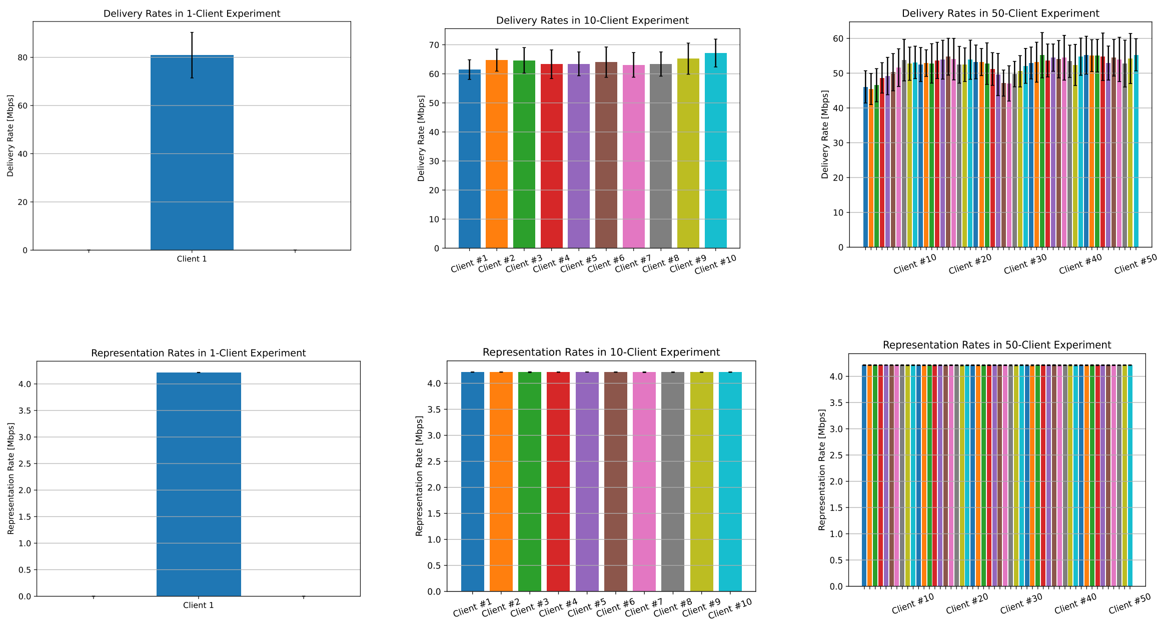

Figure 7-44 The delivery and representation rates in experiments with 1, 10 and 50 DASH clients

The results from this experiment are summarized in Figure 7-44. Looking at the top row of the figure, we can see that the delivery rate reduced from around 80 Mbps for one video client to between 60 and 70 Mbps for ten clients and to around 50 Mbps for 50 clients. However, looking at the representation rate (bottom row of the figure), we see that the delivery rate stayed constant at just above 4 Mbps, irrespective of the number of clients. Although the delivery rate decreased with an increase in the number of clients, it had no significant effect on the representation rate and thus not on the video resolution. Furthermore, we did not observe any buffering underruns, i.e., rebuffering events, when we increased the number of DASH clients from 1 to 50. Thus, our results suggest that 5G scales up well to meet the demands of an increasing number of DASH video streams.

7.4. Evaluation of 5G SA equipment at FOKUS facility

The 5G SA radio equipment installed at the FOKUS facility was evaluated using the Open5GCore. The radio deployment covers three distinct locations at FOKUS: 1) one Nokia cell deployed on the rooftop of the building covers the *Goslarer Platz*, 2) another Nokia cell along with a Huawei cell covers the underground parking deck two, and 3) a second Huawei cell is installed in a server room on the third floor of the FOKUS site.

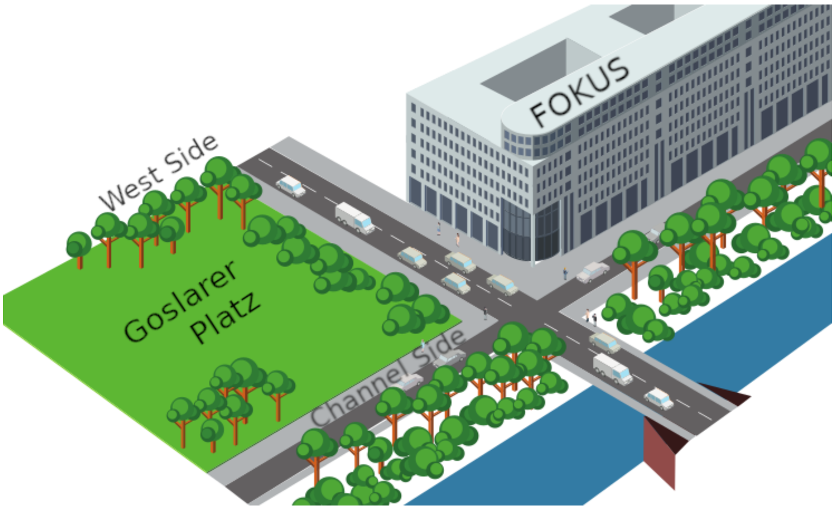

Figure 7-45 5G outdoor coverage: Goslarer Platz at FOKUS institute

As depicted in Figure 7-45, Goslarer Platz is directly in front of the FOKUS building and is divided into its sublocations West Side and Channel Side, for which experiments were executed. In addition, the underground parking deck (at sublevel 2) is divided in its sublocations front and back. Please note that, while both cells in the underground parking deck cover both areas (front and back), each of the cells are deployed in opposite areas only. Measurement probes were deployed on mobile phones as also on VMs that were connected to CPEs at all three locations.

Phones/UEs: Samsung S21+ 5G and Huawei P40
CPE: Huawei 5G CPE Pro 2

The Huawei CPE Pro 2 devices were connected to the testbed network. Using a dedicated, virtual network, VMs from the compute cluster connected to them. These VMs provided the hosts for the respective measurement probes.

Two type of KPIs were analyzed: RTT and Throughput. To evaluate the coverage of the rooftop cell (Goslarer Platz), COTS smart phones from Samsung and Huawei were prepared with measurement probes and were connected to the 5Genesis non-public network. Similarly, to evaluate the coverage of the sublevel parking deck cells, two CPEs were used, both of the exact same model as listed above. For the experiments regarding the cell in the third floor (server room), the same COTS smart phones as in the case of rooftop cell analysis were used.

7.4.1. Round-Trip Time Analysis

To be able to analyze RTT, ICMP messages were exchanged between the CPEs/cell phones and an endpoint (inside a VM) directly after the 5G core network. While originally more measurements were executed, the following figures discuss only those RTT experiments with the case of 1400 bytes payload. The main varied parameter for those measurements is the CPEs/cell phones and the cell they were connecting to. Each experiment was running for three minutes. While the experiments that were executed on the Goslarer Platz were repeated only 20 times, most of the other experiments were repeated a lot more (since they were running over night). The following three figures show: Average RTT (Figure 7-48), Minimum RTT (Figure 7-49), and Maximum RTT (Figure 7-46) of those executed experiments.

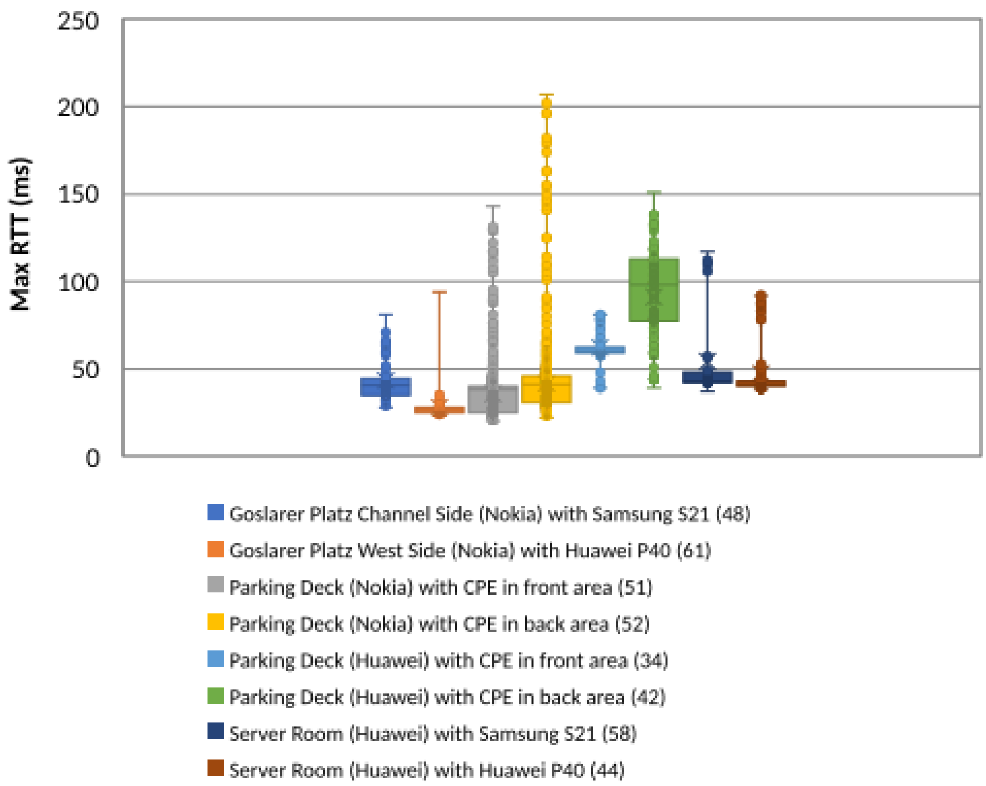

Figure 7-46 Comparing Maximum RTT for all 5G FOKUS experiments

When inspecting Maximum RTT, the values for experiment IDs 51, 52, and 42 are standing out, in regards to outliers. Initially, by analyzing the RTT experiment data further, we could not find any proper explanation for this behavior but when additionally comparing the data with other measurements, stored in our experiment database, we could find out the reason.

Those experiments that were run inside the FOKUS building, usually, were run over longer time periods. In the same time, some additional, short running throughput tests had been executed manually that interfered with the RTT experiments. For ICMP-based RTT measurement this is expected behavior – when additional IP frames try to saturate the channel/connection simultaneously. The RTT experiments were originally planned as running without any (relevant) background traffic.

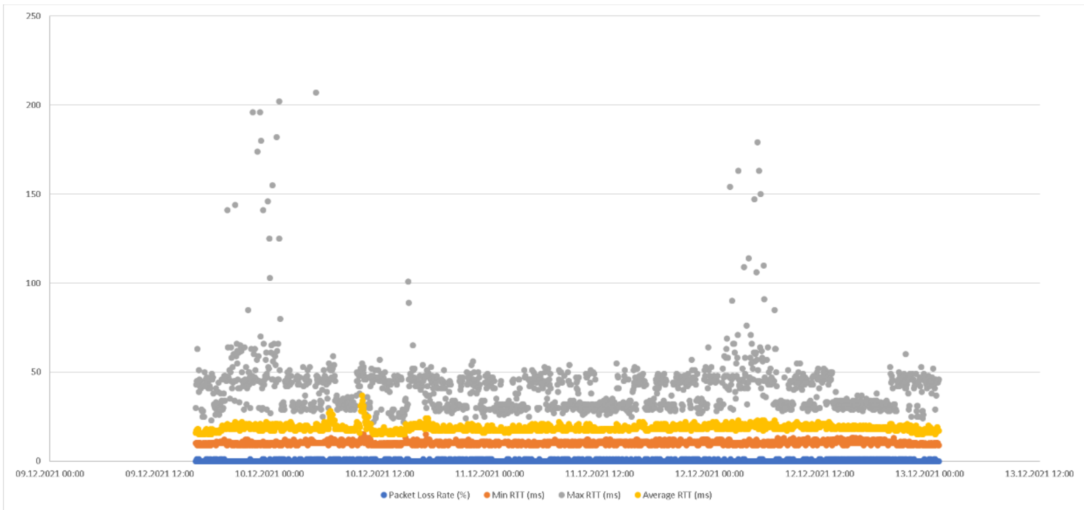

Figure 7-47 Negative influence on an example RTT experiment

An example of background traffic interference is shown in Figure 7-47, here, for the experiment in the sublevel parking deck (the CPE located in the back area attached to the Nokia cell). The experiment was running over the duration of a couple of days. As to be seen in the figure, at three phases in the RTT experiment the values of maximum RTT (light gray dots) increased significantly. As our measurement data base showed that at those times manually started throughput tests had been executed, in the background. Though, though it is not clear why these tests were executed.

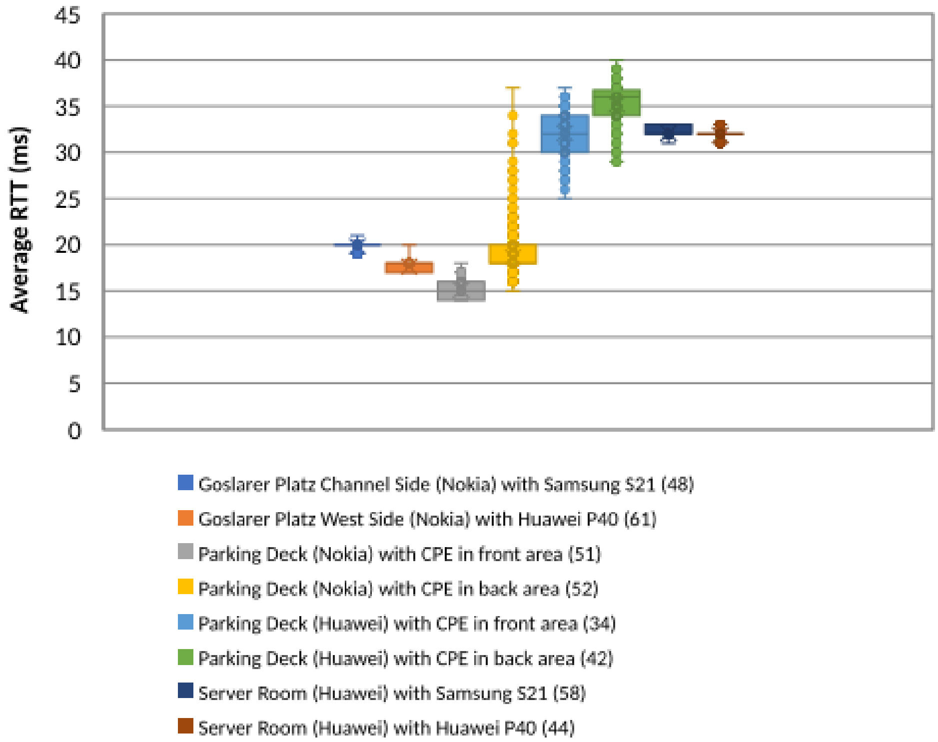

Figure 7-48 Comparing Average RTT for all 5G FOKUS experiments

When comparing Average RTT values (Figure 7-48), the cases of parking area experiments with Nokia cell (ID 51 and 52) show some of the overall lowest values. The same is true for the experiments executed in the outdoor locations at Goslarer Platz (IDs 48 and 61). Again, please note, the rather high outlier range of experiment 52 is due to the negative interference, as described above.

Surprising values can be observed in the case of indoor Server Room (IDs 58 and 44). Here, rather high average values were observed. Seemingly, for RTT the conditions in the room seem far from ideal, especially when considering the close location of cell phones and cell. This surprising result is also observable in the case of Minimum RTT (Figure 7-49).

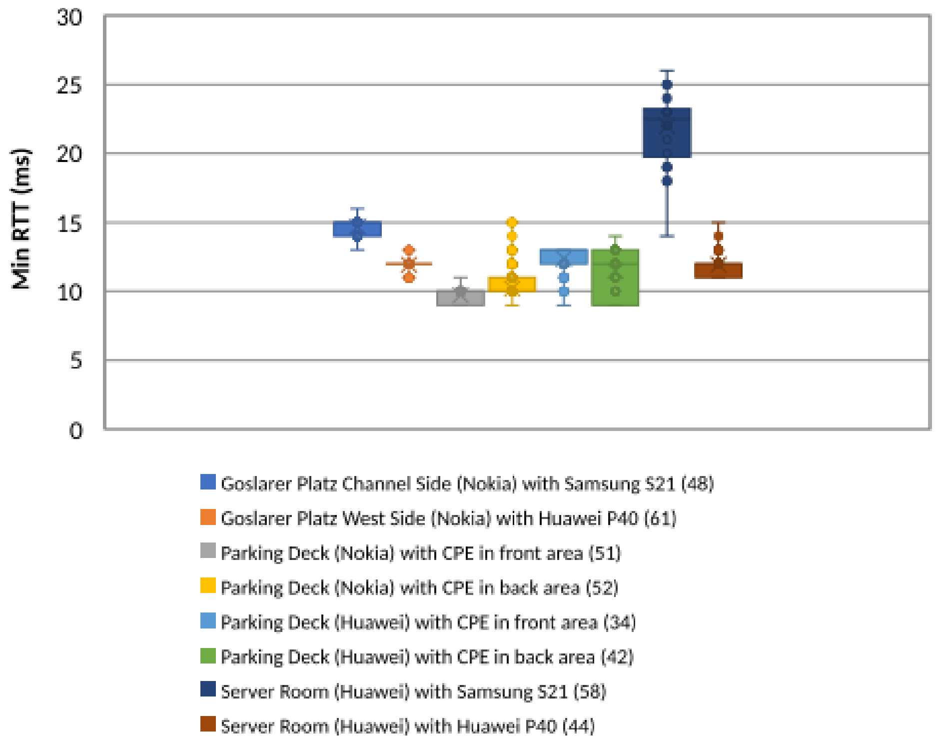

Figure 7-49 Comparing Minimum RTT for all 5G FOKUS experiments

All other cases of RTT experiments are roughly in the same area of 10 to 15ms.

7.4.2. Throughput Analysis

In addition to the RTT experiments, throughput experiments have been executed. Similar to the RTT experiments, within the throughput experiments the throughput between the CPE/cell phone and an endpoint located directly after the 5G core network was evaluated. Each experiment was running for two minutes. Again, the number of repetitions varied between the different cases. Especially, in the cases of Goslarer Platz only 20 iterations were executed. For all other cases even more iterations could be executed.

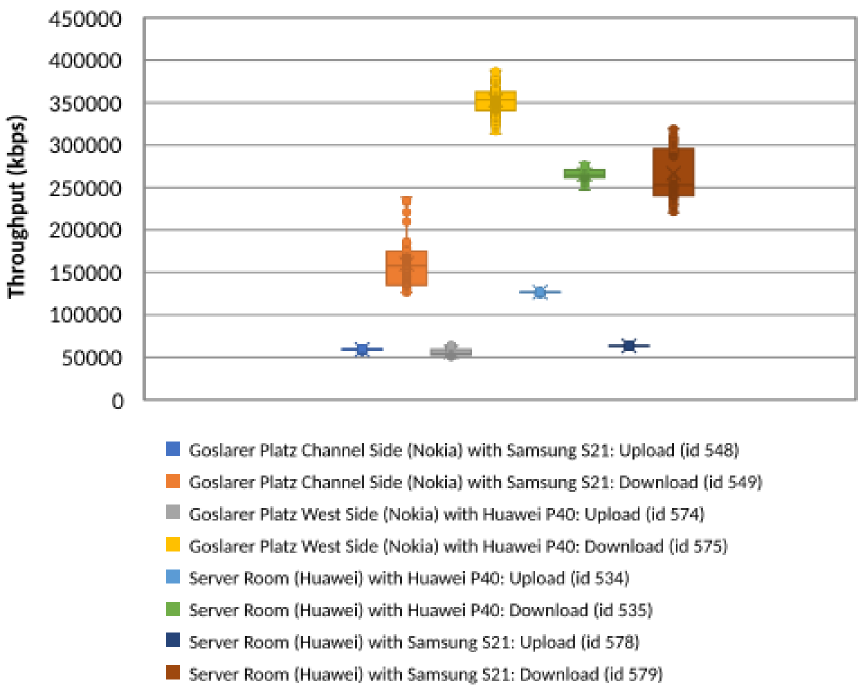

Figure 7-50 Throughput of COTS cell phones in outdoor scenario at Goslarer Platz (Nokia cell) and indoor scenario at Server Room (Huawei cell)

The results of those experiments that leverage COTS cell phones are shown Figure 7-50. With the Nokia cell mounted on the rooftop, throughput measurements were executed for the West side of Goslarer Platz as also for the Channel side. A Samsung S21 cell phone was used for the West Side, whereas a Huawei P40 was used for the Channel Side. Additionally, a Huawei cell is located in the one of the server rooms inside the FOKUS facility. The throughput experiments executed for this cell were executed with the same COTS cell phones as used for the Goslarer Platz measurements.

Another set of experiments were executed using CPEs. Here, the two cells located in the sublevel parking deck were used. The results of these experiments are shown in Figure 7-51. Please note that the Nokia cell is located in the back part of the parking deck, whereas the Huawei cell is located in the front part of the parking deck. Both cells cover whole parking deck area.

For all throughput experiments measured values are in an expectable range. For the future, firmware updates – for the COTS cell phones but also for the cells – may even allow better throughputs. A typical significant difference in measured throughput can be observed when upload and download scenarios are compared (e.g. experiment IDs 548 and 549 in Figure 7-50). While for the upload case, a value of 50 Mbps is common in our setups, in the cases where the 5G cell and the CPE/cell phone was provided by the same vendor even higher throughputs could be observed. In comparison to this, in the download case, observed throughput values may differ significantly. While stable values of 350 Mbps were observed in the outdoor cases, significantly higher throughput (of even over 700 Mbps) could be achieved – in the case of indoor, close distance measurements.

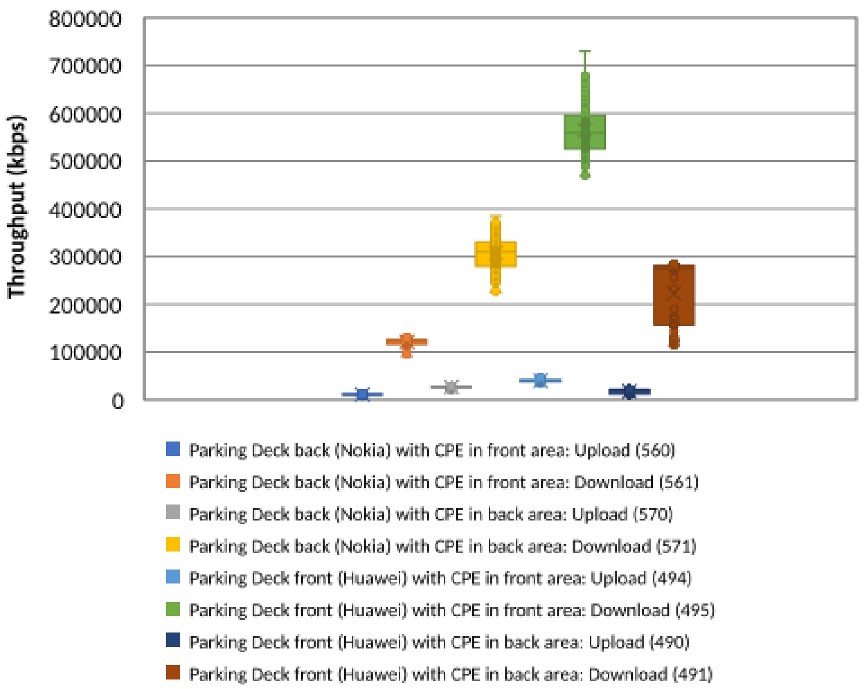

Figure 7-51 Throughput of CPEs at parking deck front and back locations

The significance of location – of CPE/cell phones – could be observed as well. Especially, in our experiments in the sublevel parking deck this is seen best (Figure 7-51). Here, the two compared cells are located in opposite parking deck areas (front and back). When executing our experiments with CPEs placed in different areas, a significant change of observed throughput is visible. For each of the cases of upload, the throughput values degrade to halve (or even less), when the location is varied. This is true for both of the in opposite deployed cells. A similar observation was possible for the case of download.

7.4.3. Conclusion

In summary, in regards to throughput, the deployed 5G system performs as expected, only little surprises were found. The case of up to 700 Mbps for the case of Huawei cell in the parking deck, came as a surprise. While we would have considered that using same vendor equipment would create better results, we would not have considered the difference that stark, as also the conditions in the parking deck good enough for this.

While analyzing the RTT experiments brought up a rather awkward problem of measuring method, we decided to keep this documented to also remind others to keep this in mind when troubleshooting. While we would have expected the RTT results to be slightly lower as observed – especially in the case of Minimum RTT – the reader is to be reminded that the full path through a 5G network is measured, leveraging ICMP.

7.5. Summary and conclusions

The 5GENESIS Berlin Platform has demonstrated the readiness of the platform to accommodate demanding ICT and vertical services on top of the several underlying infrastructures on which the 5G system was deployed. The manifold evaluation activities involving either the 5G system alone, or tested together with the different parts of the infrastructure, and as part of a video server scenario, have shown the validity of the testbed to fulfill the services' requirements and to accommodate even more demanding set of services that belong to different service classes (e.g. uRLLC).

Such 5G infrastructure makes use of the Open5GENESIS experimentation suite and features a modular design whose platform components can be replicated to establish 5G (nomadic) testbeds at either research oriented environments, industry oriented and operational facilities owned by vertical industries. The nomadic 5G testbed has therefore proven record to take part in larger field trials that will present additional challenges to the already established measurement and monitoring framework.

8. Overall Conclusions

This deliverable describes the trials and experimentation results from the third and final integration cycle of 5GENESIS. The deliverable provides analytical results from the five 5GENESIS platforms (Málaga, Athens, Limassol, Surrey and Berlin), covering 8 main KPIs through generic tests, as well as secondary KPIs associated with 14 vertical-specific use cases. The main focus has been to highlight the features and capabilities integrated during the third integration phase of the platforms, as well as to showcase the different vertical scenarios.

For all these tests, quantitative data are provided, as well as 95 % confidence intervals, based on multiple repetitions of each experiment. Specifically, for the baseline measurements the achieved throughput of 5G reached almost 1.2 Gbps subject to deployment configuration and equipment availability in each platform. Similarly, the measured RTT dropped to 10 msec for Stand-alone (SA) deployments. In any case, there were significant deviations from the theoretical values, depending on the components used (e.g. software vs. hardware radios, virtualised vs. physical core functions etc.).

During this final phase, each platform demonstrated its final configuration as well as a number of important milestones reached; the Athens platform showcased its slicing capabilities, multi-domain setups and the drone/SecaaS use cases; the Malaga platform exhibited a variety of 5G configurations (FR1/FR2), maximizing 5G performance and demonstrating demanding scenarios around MCS and large-scale events; the Limassol platform validated 5G-satellite integration with 5G SA backhauling and local break-out enhancing maritime and rural application scenarios; the Surrey platform showcased Multi-RAT scenarios with 5G/WiFi slicing and various IoT use cases engaging MQTT and CoAP; finally, the Berlin platform evaluated the interplay of 5G with a complex compute infrastructure and demonstrated the mmWave backhaul in a nomadic 5G setup.

In addition to providing measurement results, this third experimentation cycle highlighted once again the automated experimental methodology and statistical analysis as well as the incorporation in the KPI measurement process of residual measurements for each test case.

REFERENCES

[1] 5GENESIS Consoritum, "D2.1 Requirements of the Facility," 2018. [Online]. Available: https://5genesis.eu/wp-content/uploads/2018/11/5GENESIS_D2.1_v1.0.pdf.

[2] 5GENESIS Consortium, "D2.2 Initial overall facility design and specifications," 2018. [Online]. Available: https://5genesis.eu/wp-content/uploads/2018/12/5GENESIS_D2.2_v1.0.pdf.

[3] 5GENESIS Consortium, "D4.3 The Athens Platform – Release C," 2021. [Online]. Available: https://5genesis.eu/wp-content/uploads/2021/09/5GENESIS-D4.3_v1.0.pdf

[4] 5GENESIS Consortium, "D4.6 The Malaga Platform – Release C," 2021. [Online]. Available: https://5genesis.eu/wp-content/uploads/2021/10/5GENESIS_D4.6_v1.0.pdf

[5] 5GENESIS Consortium, "D4.9 The Limassol Platform – Release C," 2021. [Online]. Available: https://5genesis.eu/wp-content/uploads/2021/07/5GENESIS_D4.9_v.1.0.pdf

[6] 5GENESIS Consortium, "D4.12 The Surrey Platform – Release C," 2021. [Online]. Available: https://5genesis.eu/wp-content/uploads/2021/10/5GENESIS_D4.12_v1.0.pdf

[7] 5GENESIS Consortium, "D4.15 The Berlin Platform – Release C," 2021. [Online]. Available: https://5genesis.eu/wp-content/uploads/2021/11/5GENESIS_D4._15_v1.0.pdf

[8] 5GENESIS Consortium, "D6.1 Trials and experimentation (cycle 1)," 2019 [Online]. Available: https://5genesis.eu/wp-content/uploads/2019/12/5GENESIS_D6.1_v2.00.pdf

[9] 5GENESIS Consortium, "D6.2 Trials and experimentation (cycle 2)," 2020 [Online]. Available: https://5genesis.eu/wp-content/uploads/2020/08/5GENESIS_D6.2_v1.0_FINAL.pdf

[10] 3GPP, "3GPP TS 37.340 NR; Multi-connectivity; Overall description; Stage-2," 3GPP, Dec 2018.

[11] "TRIANGLE Project, "D2.6 First Test Scenario and Test Specifications"," September 2018. [Online]. Available: https://www.triangle-project.eu/wp-content/uploads/2018/11/TRIANGLE_D2-6.pdf.

[12] 3GPP, "3GPP TS 22.179 Mission Critical Push to Talk (MCPTT); Stage 1," 3GPP, 2019.

[13] 5GENESIS Consortium, "Deliverable D5.3 - Documentation and supporting material for 5G stakeholders," 2020. [Online]. Available: https://5genesis.eu/wp-content/uploads/2020/07/5GENESIS-D5.3_v1.0.pdf. [Accessed June 2020].

[14] 5GENESIS Consortium, "D3.3 Slice management," 2019. [Online]. Available: https://5genesis.eu/wp-content/uploads/2019/10/5GENESIS_D3.3_v1.0.pdf. [Accessed June 2020].

[15] 5GENESIS Consortium, "5GENESIS Github," [Online]. Available: https://github.com/5genesis. [Accessed June 2020].A. Author, B. Author, "Paper Title", Proc. 1st Conference Title, Location, Country, Month/Year, pp. 123-145

[16] L. Cominardi, L. M. Contreras, C. J. Bcrnardos and I. Berberana, "Understanding QoS Applicability in 5G Transport Networks," 2018 IEEE International Symposium on Broadband

Multimedia Systems and Broadcasting (BMSB), 2018, pp. 1-5, doi: 10.1109/BMSB.2018.8436847.

[17] NGMN 5G White paper, https://ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf

[18] 5GENESIS, D3.12 – 5G Radio components/ 5G Air interface (Release B), 2021

[19] "Quectel module" [Online], https://www.quectel.com/product/5g-rm500q-gl, Retrieved 12/2021

[20] "Simcom module" [Online], https://www.simcom.com/product/SIM8200EA_M2.html, Retrieved 12/2021

9. Annex 1: Detailed Experiment Results

9.1. Athens Platform Results

9.1.1. Generic tests – Results

9.1.1.1. Athens Baseline TCP Throughput Test

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_THR_TCP</td></tr>
<tr><td>General description of the test</td><td colspan="4">DL Throughput of 5G SA network</td></tr>
<tr><td>Purpose</td><td colspan="4">This test estimates the maximum user data rate in the downlink of a 5G SA network.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>NCSRD</td><td>Date:</td><td>09/12/2021</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, SRL, COS</td></tr>
<tr><td>Scenario</td><td colspan="4">One connected COTS 5G UE to the network
Radio configuration for 5G cell: 50 MHz, 2x2, Band n78, TDD, 256 QAM DL</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">N/A</td></tr>
<tr><td>Components involved</td><td colspan="4">Amarisoft RAN (eNB & gNB)
Amarisoft CN Rel. 16
OnePlus 8 PRO 5G (endpoint 1)
Dell G515 (endpoint 2)
UMA iPerf (Android Application)
OpenTAP for automated testing (iPerf TAP plugin)
NCSRD Cell Performance (Android Application)</td></tr>
<tr><td>Metric(s) under study</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Iperf2.0.10</td></tr>
<tr><td>Primary measurement results</td><td colspan="4">Mean: 293+/- 7.4 Mbps
Median: 300.14 +/- 10.2 Mbps
Standard Deviation: 46.64 +/- 3.20 Mbps
Max: 361.88 +/- 1.14 Mbps
Min: 194.88 +/- 11.2700452973 Mbps
95% Percentile: 353.3 +/- 3.87 Mbps
75% Percentile: 330.22 +/- 8.07 Mbps
25% Percentile: 260.22 +/- 10.36 Mbps
5% Percentile: 216.14 +/- 9.72 Mbps</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">Mean RSRP (dBm): -72.48 +/- 0.30
Mean RSRQ (dB): -11.00 +/- 0.21
Avg MCS: 26.9 +/- 0.2</td></tr>
</table>

9.1.1.2. Athens Baseline RTT Tests – RAN Config 1

Test Case ID	TC_RTT_e2e			
General description of the test	Measure End-to-end RTT of 5G SA network for packet sizes 32,64, 128, 512 bytes			
Purpose	Estimate the E2E RTT in a 5G SA network for different packet sizes.			
Executed by	Partner:	NCSRD	Date:	09/12/2021
Involved Partner(s)	UMA, SRL, COS			
Scenario	• One connected COTS 5G UE • Radio configuration 50 MHz, 2x2, Band n78, TDD, 256 QAM DL, 64 QAM UL			
Slicing configuration	N/A			
Components involved	• Amarisoft gNB • Amarisoft CN Rel. 16 • OnePlus 8 PRO 5G (endpoint 1) • Dell G515 (endpoint 2) • UMA Ping Agent (Android Application) • OpenTAP for automated testing (ping TAP plugin) • NCSRD Cell Performance (Android Application)			
Metric(s) under study	*Round Trip Time*			
Additional tools involved	ping			
Primary measurement results	**E2E RTT per packet size** **32 bytes:** Mean: 29.09 +/- 0.26 ms Standard deviation: 3.77 +/- 0.14 ms Median: 29.00 +/- 0.13 ms Min: 22.49 +/- 0.27 ms Max: 35.94 +/- 0.64 ms 5% Percentile: 23.45 +/- 0.26 ms 25% Percentile: 26.62 +/- 0.25 ms 75% Percentile: 31.77 +/- 0.27 ms 95% Percentile: 34.71 +/- 0.34 ms **64 bytes:** Mean: 29.03 +/- 0.08 ms Standard deviation: 3.70 +/- 0.17 ms Median: 29.02 +/- 0.28 ms Min: 22.70 +/- 0.25 ms Max: 35.98 +/- 0.83 ms 5% Percentile: 23.48 +/- 0.22 ms 25% Percentile: 26.59 +/- 0.26 ms 75% Percentile: 31.43 +/- 0.31 ms 95% Percentile: 34.53 +/- 0.39 ms			

	128 bytes: Mean: 38.85 +/- 0.12 ms Standard deviation: 3.74 +/- 0.141 ms Median: 38.49 +/- 0.26 ms Min: 32.48 +/- 0.24 ms Max: 46.12 +/- 0.71 ms 5% Percentile: 33.46 +/- 0.29 ms 25% Percentile: 36.19 +/- 0.26 ms 75% Percentile: 41.48 +/- 0.32 ms 95% Percentile: 44.79 +/- 0.30 ms **512 bytes:** Mean: 38.81 +/- 0.14 ms Standard deviation: 3.83 +/- 0.11 ms Median: 38.44 +/- 0.27 ms Min: 32.29 +/- 0.31 ms Max: 46.19 +/- 0.68 ms 5% Percentile: 33.26 +/- 0.25 ms 25% Percentile: 36.21 +/- 0.24 ms 75% Percentile: 41.63 +/- 0.31 ms 95% Percentile: 44.87 +/- 0.37 ms
Complementary measurement results	Mean RSRP (dBm): -73.2 +/- 0.15 Mean RSRQ (dB): -11.2 +/- 0.23 Failed Ratio = 0.00 +/- 0.00 Success Ratio = 1.00 +/- 0.00

9.1.1.3. Athens RTT Tests with Background Traffic – RAN Config 1

Test Case ID	TC_RTT_e2eBGTraffic			
General description of the test	Measure End-to-end RTT of 5G SA network			
Purpose	Estimate the E2E RTT in a 5G SA network when there is concurrent traffic.			
Executed by	Partner:	NCSRD	Date:	15/12/2021
Involved Partner(s)	UMA, SRL, COS			
Scenario	• One connected COTS 5G UE • Radio configuration:5G cell: 50 MHz, 2x2, Band n78, TDD, 256 QAM DL			
Slicing configuration	N/A			
Components involved	• Amarisoft gNB • Amarisoft Core Rel. 16 • Huawei P40 PRO 5G • Dell G515 (endpoint 2) • NCSRD Ping Tool (Linux Application)			

Metric(s) under study	*Round Trip Time*
Additional tools involved	Ping, iPerf3
Primary measurement results	DownLink Traffic 5Mbps – UDP End-to-end RTT [ms] - 32 bytes Mean: 24.08 ms Standard deviation: 9.35 ms Min: 6.93 ms Max: 42.64 ms End-to-end RTT [ms] - 64 bytes Mean: 24.08 ms Standard deviation: 9.34 ms Min: 7.04 ms Max: 42.20 ms End-to-end RTT [ms] - 128 bytes Mean: 34.18 ms Standard deviation: 9.24 ms Min: 16.91 ms Max: 52.97 ms End-to-end RTT [ms] - 512 bytes Mean: 34.23 ms Standard deviation: 9.23 ms Min: 17.00 ms Max: 53.53 ms DownLink Traffic 100Mbps – UDP End-to-end RTT [ms] - 32 bytes Mean: 19.21 ms Standard deviation: 6.83 ms Min: 7.64 ms Max: 37.67 ms End-to-end RTT [ms] - 64 bytes Mean: 19.17 ms Standard deviation: 6.83 ms Min: 7.01 ms Max: 38.44 ms End-to-end RTT [ms] - 128 bytes Mean: 28.58 ms Standard deviation: 6.71 ms Min: 17.30 ms Max: 51.34 ms

	End-to-end RTT [ms] - 512 bytes Mean: 29.71 ms Standard deviation: 6.66 ms Min: 16.98 ms Max: 49.82 ms **DownLink Traffic 200Mbps – UDP** **End-to-end RTT [ms] - 32 bytes** Mean: 19.59 ms Standard deviation: 6.97 ms Min: 7.67 ms Max: 41.40 ms **End-to-end RTT [ms] - 64 bytes** Mean: 19.73 ms Standard deviation: 7.10 ms Min: 7.65 ms Max: 48.99 ms **End-to-end RTT [ms] - 128 bytes** Mean: 29.93 ms Standard deviation: 7.65 ms Min: 17.65 ms Max: 53.85 ms **End-to-end RTT [ms] - 512 bytes** Mean: 30.14 ms Standard deviation: 6.88 ms Min: 18.15 ms Max: 53.14 ms
Complementary measurement results	N/A

9.1.1.4. Athens RTT Tests – RAN Config 2 (Low Latency)

Test Case ID	TC_RTT_e2e			
General description of the test	Measure End-to-end RTT of 5G SA network for packet sizes 32,64, 128, 512 bytes			
Purpose	Estimate the E2E RTT in a 5G SA network for different packet sizes.			
Executed by	Partner:	NCSRD	Date:	09/12/2021
Involved Partner(s)	UMA, SRL, COS			
Scenario	• One connected COTS 5G UE			

	• Radio configuration 50 MHz, 2x2, Band n78, TDD, 256 QAM DL, 64 QAM UL
Slicing configuration	N/A
Components involved	• Amarisoft gNB • Amarisoft CN Rel. 16 • OnePlus 8 PRO 5G (endpoint 1) • Dell G515 (endpoint 2) • UMA Ping Agent (Android Application) • OpenTAP for automated testing (ping TAP plugin) • NCSRD Cell Performance (Android Application)
Metric(s) under study	*Round Trip Time*
Additional tools involved	ping
Primary measurement results	**E2E RTT per packet size** **32 bytes:** Mean: 12.36 +/- 0.09 ms Standard deviation: 0.82 +/- 0.07 ms Median: 12.27 +/- 0.10 ms Min: 11.28 +/- 0.13 ms Max: 14.75 +/- 0.39 ms 5% Percentile: 11.41 +/- 0.11 ms 25% Percentile: 11.81 +/- 0.10 ms 75% Percentile: 12.66 +/- 0.11 ms 95% Percentile: 13.8 +/- 0.25 ms **64 bytes:** Mean: 12.43 +/- 0.07 ms Standard deviation: 0.90 +/- 0.07 ms Median: 12.3 +/- 0.07 ms Min: 11.26 +/- 0.12 ms Max: 14.98 +/- 0.35 ms 5% Percentile: 11.47 +/- 0.08 ms 25% Percentile: 11.76 +/- 0.08 ms 75% Percentile: 12.72 +/- 0.09 ms 95% Percentile: 14.09 +/- 0.23 ms **128 bytes:** Mean: 20.70 +/- 0.10 ms Standard deviation: 1.62 +/- 0.11 ms Median: 20.09 +/- 0.21 ms Min: 18.78 +/- 0.17 ms Max: 25.17 +/- 0.57 ms 5% Percentile: 19.13 +/- 0.10 ms 25% Percentile: 19.51 +/- 0.03 ms 75% Percentile: 21.65 +/- 0.15 ms

	95% Percentile: 23.64 +/- 0.34 ms **512 bytes:** Mean: 20.10 +/- 0.05 ms Standard deviation: 1.03 +/- 0.07 ms Median: 19.64 +/- 0.04 ms Min: 18.86 +/- 0.13 ms Max: 22.83 +/- 0.44 ms 5% Percentile: 19.17 +/- 0.09 ms 25% Percentile: 19.50 +/- 0.02 ms 75% Percentile: 20.50 +/- 0.18 ms 95% Percentile: 22.04 +/- 0.10 ms
Complementary measurement results	Mean RSRP (dBm): -72.2 +/- 0.22 Mean RSRQ (dB): -11.00 +/- 0.21 Failed Ratio = 0.00 +/- 0.00 Success Ratio = 1.00 +/- 0.00

9.1.1.5. Athens RTT Tests with Background Traffic – RAN Config 2

Test Case ID	TC_RTT_e2eBGTraffic			
General description of the test	Measure End-to-end RTT of 5G SA network			
Purpose	Estimate the E2E RTT in a 5G SA network when there is concurrent traffic.			
Executed by	Partner:	NCSRD	Date:	15/12/2021
Involved Partner(s)	UMA, SRL, COS			
Scenario	• One connected COTS 5G UE • Radio configuration:5G cell: 50 MHz, 2x2, Band n78, TDD, 256 QAM DL			
Slicing configuration	N/A			
Components involved	• Amarisoft gNB • Amarisoft Core Rel. 16 • Huawei P40 PRO 5G • Dell G515 (endpoint 2) • NCSRD Ping Tool (Linux Application)			
Metric(s) under study	*Round Trip Time*			
Additional tools involved	Ping, iPerf3			
Primary measurement results	**DownLink Traffic 5Mbps – UDP** **End-to-end RTT [ms] - 32 bytes** Mean: 12.07 ms Standard deviation: 1.63 ms Min: 7.80 ms Max: 18.24 ms			

	End-to-end RTT [ms] - 64 bytes Mean: 11.98 ms Standard deviation: 1.50 ms Min: 8.01 ms Max: 16.76 ms End-to-end RTT [ms] - 128 bytes Mean: 19.45 ms Standard deviation: 2.03 ms Min: 12.88 ms Max: 27.67 ms End-to-end RTT [ms] - 512 bytes Mean: 19.62 ms Standard deviation: 1.64 ms Min: 15.56 ms Max: 24.69 ms DownLink Traffic 100Mbps – UDP End-to-end RTT [ms] - 32 bytes Mean: 11.65 ms Standard deviation: 3.31 ms Min: 6.61 ms Max: 22.84 ms End-to-end RTT [ms] - 64 bytes Mean: 11. 71 ms Standard deviation: 3.11 ms Min: 6.65 ms Max: 21.64 ms End-to-end RTT [ms] - 128 bytes Mean: 12.05 ms Standard deviation: 6.4 ms Min: 6.4 ms Max: 25.68 ms End-to-end RTT [ms] - 512 bytes Mean: 18.99 ms Standard deviation: 3.24 ms Min: 13.66 ms Max: 30.57 ms DownLink Traffic 200Mbps – UDP End-to-end RTT [ms] - 32 bytes

	Mean: 12.02 ms Standard deviation: 3.34 ms Min: 6.95 ms Max: 22.51 ms End-to-end RTT [ms] - 64 bytes Mean: 12.03 ms Standard deviation: 3.51 ms Min: 6.68 ms Max: 23.00 ms End-to-end RTT [ms] - 128 bytes Mean: 12.75 ms Standard deviation: 4.05 ms Min: 6.59 ms Max: 27.86 ms End-to-end RTT [ms] - 512 bytes Mean: 20.11 ms Standard deviation: 3.57 ms Min: 13.86 ms Max: 32.25 ms
Complementary measurement results	N/A

9.2. Malaga Platform Results

9.2.1. Generic tests – Results

9.2.1.1. Malaga facility Throughput KPI

Test Case ID	TC_THR_TCP			
Purpose	Measure the downlink throughput in 5G SA deployment.			
Executed by	Partner:	UMA	Date:	10.12.2021
Partner(s)	UMA, ATHONET			
Scenario	5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier 4 layers, 4x4, 256QAM, Single beam			
Slicing configuration	Proactive scheduling activated			
Components	One plus 9. Setup 8.2 Full E2E 5G SA			
Metric(s) under study	*Throughput*			
Additional tools	TAP for automated testing, VNF, iPerf, iPerf TAP plugin,			
Primary measurement results	Mean: 1182.19 +/-6.01 Mbps Standard deviation: 30.40 +/- 6.55 Mbps			

	Median: 30.40 +/- 6.55 Mbps Min: 1040.96 +/- 49.86 Mbps Max: 1403.56 +/- 57.89 Mbps 25% Percentile: 1177.13 +/- 8.43 Mbps 75% Percentile: 1190.98 +/- 1.92 Mbps 5% Percentile: 1155.68 +/- 17.49 Mbps 95% Percentile: 1199.74 +/- 0.88 Mbps
Complementary measurement results	NR RSRP max;-81.0; NR RSRP min;-82.80; NR RSRP avg;-82.0; NR RSRQ max;-10.30; NR RSRQ min;-10.40; NR RSRQ avg;-10.3;

9.2.1.2. Malaga facility RTT KPI

Test Case ID	TC_RTT_e2e			
General description of the test	The tests assess the average, minimum, maximum, 5% percentile and 95% percentile RTT between a UE and a VNF deployed on a single compute node in the network.			
Purpose	Measure e2e RTT			
Executed by	Partner:	UMA	Date:	11.05.2020
Partner(s)	UMA, ATHONET			
Scenario	5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier 4 layers, 4x4, 256QAM, Single beam			
Slicing configuration	Proactive scheduling activated			
Components	One plus 9. Setup 8.2 Full E2E 5G SA			
Metric(s) under study	*Round Trip Time*			
Additional tools	TAP for automated testing, VNF, Ping, Ping TAP plugin,			
Primary measurement results	**Round Trip time [ms]** Mean: 19.53 +/- 0.30 ms Standard deviation: 7.55 +/- 0.17 ms Median: 18.23 +/- 0.66 ms Min: 10.02 +/- 0.01 ms Max: 35.10 +/- 0.36 ms 25% Percentile: 12.34 +/- 0.25 ms 75% Percentile: 25.72 +/- 0.36 ms 5% Percentile: 10.51 +/- 0.06 ms 95% Percentile: 32.60 +/- 0.35 ms			
Complementary measurement results	NR RSRP max;-80.6, NR RSRP min;-84.4, NR RSRP avg;-82.44, NR RSRQ max;-10.3, NR RSRQ min;-10.4, NR RSRQ avg;-10.3			

9.2.1.3. Malaga facility Reliability KPI

Test Case ID	*TC_Rel_e2e*			
General description of the test	This KPI refers to reliability measured between two endpoints of the network. It measures the amount of packets successfully delivered between these two endpoints within the time constraints specified, divided by the total number of sent packets.			
Purpose	Measure e2e reliability			
Executed by	Partner:	UMA	Date:	11.05.2020
Partner(s)	UMA, ATHONET			
Scenario	5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier 4 layers, 4x4, 256QAM, Single beam			
Slicing configuration	Proactive scheduling activated			
Components	One plus 9. Setup 8.2 Full E2E 5G SA			
Metric(s) under study	*Reliability*			
Additional tools	TAP for automated testing, VNF, Ping, Ping TAP plugin,			
Primary measurement results	40.3 % of successful packets received for a RTT < 15 55.2 % of successful packets received for a RTT < 20 ms 70.4 % of successful packets received for a RTT < 25 87.1 % of successful packets received for a RTT < 30 ms 99.6 % of successful packets received for a RTT < 35 99.97 % of successful packets received for a RTT < 40 ms Duration: 90 minutes Packet size: 32 bytes			
Complementary measurement results	Mean SINR (dB) 21.4; Max SINR 21.8; Min SINR 20.6 Mean RSRQ -10.8 dB Mean RSRP -51 dBm			

9.2.1.4. Málaga facility Location accuracy KPI

Test Case ID	TC_Loc_Acc			
General description of the test	The test case evaluates the mean, median, minimum, máximum and standard deviation of the location accuracy of a terminal on the network. The complementary measurements obtained in the tests are included for each eNB. The delay is the time elapsed between the first packet sending the terminal location information and the last packet corresponding to the location estimation.			
Purpose	Measure location accuracy with CITA method and 2 cells coverage			
Executed by	Partner:	UMA	Date:	28.06.2021

Involved Partner(s)	UMA
Scenario	Polaris NetTest Rel. 15 NSA Core with Nokia Airscale eNB with LTE band 7. The terminal can obtain measurements from 2 eNB. The E-CID+TA (CITadv) localization method is used.
Slicing configuration	-
Components involved (e.g. HW components, SW components)	Nokia Airscale eNB, Polaris NetTest Rel. 15 EPC, Creativity Software LCS, LCS EMS (Element Management System), Samsung Galaxy Note10+ 5G
Metric(s) under study	Accuracy, RSRP, RSRQ, TA1, TA2
Additional tools involved	-
Primary measurement results (those included in the test case definition)	**Accuracy [m]** Mean: 47,18 Median: 46,60 Max: 62,89 Min: 36,78 Standard deviation: 5,80
Complementary measurement results	**RSRP [dBm]** PCI: 166 / PCI: 167 Mean: -87 to -86 / -79 to -78 **RSRQ [dB]** PCI: 166 / PCI: 167 Mean: -14 to -13.5 dB / -6 to -5.5 dB **TA1** Mean: 13 **TA2** Mean: 15 **Delay [hh:mm:ss,000]** Mean: 00:00:02,806

Test Case ID	TC_Loc_Acc			
General description of the test	The test case evaluates the mean, median, minimum, máximum and standard deviation of the location accuracy of a terminal on the network. The complementary measurements obtained in the tests are included for each eNB. The delay is the time elapsed between the first packet sending the terminal location information and the last packet corresponding to the location estimation.			
Purpose	Measure location accuracy with GEO method and 2 cells coverage			
Executed by	Partner:	UMA	Date:	28.06.2021
Involved Partner(s)	UMA			
Scenario	Polaris NetTest Rel. 15 NSA Core with Nokia Airscale eNB with LTE band 7. The terminal can obtain measurements from 2 eNB. The GEO Multilateration localization method is used.			
Slicing configuration	-			
Components involved (e.g. HW components, SW components)	Nokia Airscale eNB, Polaris NetTest Rel. 15 EPC, Creativity Software LCS, LCS EMS (Element Management System), Samsung Galaxy Note10+ 5G			
Metric(s) under study	Accuracy, RSRP, RSRQ, TA1, TA2			
Additional tools involved	-			
Primary measurement results (those included in the test case definition)	**Accuracy [m]** Mean: 72,57 Median: 69,40 Max: 90,19 Min: 65,33 Standard deviation: 6,54			
Complementary measurement results	**RSRP [dBm]** <table><tr><th></th><th>PCI: 166</th><th>PCI: 167</th></tr><tr><td>Mean</td><td>-87 a -86</td><td>-81 a -80</td></tr></table> **RSRQ [dB]** <table><tr><th></th><th>PCI: 166</th><th>PCI: 167</th></tr><tr><td>Mean</td><td>-12.5 to -12</td><td>-6 to -5.5</td></tr></table> **TA1** Mean: 14 **TA2** Mean: 16 **Delay [hh:mm:ss,000]**			

	Mean: 00:00:02,807

Test Case ID	TC_Loc_Acc			
General description of the test	The test case evaluates the mean, median, minimum, máximum and standard deviation of the location accuracy of a terminal on the network. The complementary measurements obtained in the tests are included for each eNB. The delay is the time elapsed between the first packet sending the terminal location information and the last packet corresponding to the location estimation.			
Purpose	Measure location accuracy with CITA method and 3 cells			
Executed by	Partner:	UMA	Date:	28.06.2021
Involved Partner(s)	UMA			
Scenario	Polaris NetTest Rel. 15 NSA Core with Nokia Airscale eNB with LTE band 7. The terminal can obtain measurements from 3 eNB. The E-CID+TA (CITadv) localization method is used.			
Slicing configuration	-			
Components involved (e.g. HW components, SW components)	Nokia Airscale eNB, Polaris NetTest Rel. 15 EPC, Creativity Software LCS, LCS EMS (Element Management System), Samsung Galaxy Note10+ 5G			
Metric(s) under study	Accuracy, RSRP, RSRQ, TA1, TA2			
Additional tools involved	-			
Primary measurement results (those included in the test case definition)	**Accuracy [m]** Mean: 260,69 Median: 255,18 Max: 298,15 Min: 236,62 Standard deviation: 16,77			
Complementary measurement results	**RSRP [dBm]** | | PCI: 166 | PCI: 167 | PCI: 138 | | Mean | -93 to -92 | -87 to -86 | -97 to -96 | **RSRQ [dB]** | | PCI: 166 | PCI: 167 | PCI: 138 | | Mean | -11 to -10.5 | -8 to -7.5 | -16.5 to -16 | **TA1** Mean: 25			

	TA2 Mean: 26 **Delay [hh:mm:ss,000]** Mean: 00:00:02,803

Test Case ID	TC_Loc_Acc			
General description of the test	The test evaluates the mean, median, minimum, máximum and standard deviation of the location accuracy of a terminal on the network. The complementary measurements obtained in the tests are included for each eNB. The delay is the time elapsed between the first packet sending the terminal location information and the last packet corresponding to the location estimation.			
Purpose	Measure location accuracy with GEO method and 3 cells coverage			
Executed by	Partner:	UMA	Date:	28.06.2021
Involved Partner(s)	UMA			
Scenario	Polaris NetTest Rel. 15 NSA Core with Nokia Airscale eNB with LTE band 7. The terminal can obtain measurements from 3 eNB. The GEO Multilateration localization method is used.			
Slicing configuration	-			
Components involved (e.g. HW components, SW components)	Nokia Airscale eNB, Polaris NetTest Rel. 15 EPC, Creativity Software LCS, LCS EMS (Element Management System), Samsung Galaxy Note10+ 5G			
Metric(s) under study	Accuracy, RSRP, RSRQ, TA1, TA2 and Delay			
Additional tools involved	-			
Primary measurement results (those included in the test case definition)	**Accuracy [m]** Mean: 51,81 Median: 49,54 Max: 100,54 Min: 31,04 Standard deviation: 20,39			
Complementary measurement results	**RSRP [dBm]** | | PCI: 166 | PCI: 167 | PCI: 138 | | Mean | -94 to -93 | -90 to -89 | -99 to -98 | **RSRQ [dB]** | | PCI: 166 | PCI: 167 | PCI: 138 | | Mean | -11.5 to -11 | -7.5 to -7 | -16 to -15.5 |			

	TA1 Mean: 24 **TA2** Mean: 27 **Delay [hh:mm:ss,000]** Mean: 00:00:02,818

9.2.2. UC#2: Multimedia Mission Critical Services –Results

Test Case ID	TC_MCPTTAccessTime_MAL			
General description of the test	MCPTT Access time test, this test assesses the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking. It does not include the MCPTT call establishment time, since it measures the time previously defined when the request to speak is done during an ongoing call.			
Purpose	Measure time from request to speak to permission granted in a MCPTT call. The MCPTT access time calibration tests aims at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests.			
Executed by	Partner:	UMA	Date:	26.11.2021
Involved Partner(s)	UMA, ADZ			
Scenario	5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier, 4 layers, 4x4, 256QAM, Single beam The measurements are taken at the application level, in the Airbus Agnet MCS application.			
Slicing configuration	-			
Components involved *(e.g. HW components, SW components)*	ADZ MCS applications and MCS server VNF, Nokia Airscale gNB, Athonet Rel. 15 5GC, OnePlus9 5G SA UEs			
Metric(s) under study (*Refer to those in Section 4*)	*MCPTT*			
Additional tools involved	-			
Primary measurement results *(those included in the test case definition)*	MCPTT Access time MCPTT access time [ms] Mean: 27,775 95% confidence interval for Mean: Lower bound 24,855; Upper bound 30,695 MCPTT access time [ms]			

	95% Percentile	95% confidence interval for Min	
		Lower bound	Upper bound
	36,523	32,638	40,408
Complementary measurement results	n/a		

9.2.3. UC#3: Edge-based Mission Critical Services –Results

Test Case ID	*TC_MCPTTAccessTimeIncCallEstablishment_MAL*		
General description of the test	MCPTT end-to-end access time test, this test assesses the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking, including MCPTT call establishment and possibly acknowledgment from first receiving user before voice can be transmitted.		
Purpose	Measure time from request to speak to permission granted in a MCPTT call, including call establishment. The end-to-end MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further end-to-end MCPTT access time tests.		
Executed by	Partner: UMA	Date:	25.11.2021
Involved Partner(s)	UMA, NEM		
Scenario	5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier, 4 layers, 4x4, 256QAM, Single beam. The measurements are taken at the application level, in the Nemergent MCS application.		
Slicing configuration	-		
Components involved *(e.g. HW components, SW components)*	NEM MCS applications and MCS server VNF, Nokia Airscale gNB, Athonet Rel. 155GC, OnePlus9 5G SA UEs		
Metric(s) under study (*Refer to those in Section 4*)	*MCPTT*		
Additional tools involved	Logcat Android log command-line tool		
Primary measurement results *(those included in the test case definition)*	MCPTT end-to-end access time		

End-to-end MCPTT access time [ms]		
Mean	95% confidence interval for Mean	
	Lower bound	Upper bound
145,288	130,256	160,320

	End-to-end MCPTT access time [ms]		
	95% Percentile	95% confidence interval for Min	
		Lower bound	Upper bound
	221,562	197,248	245,876
Complementary measurement results	n/a		

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_MCPTTAccessTime_MAL</td></tr>
<tr><td>General description of the test</td><td colspan="4">MCPTT Access time test, this test assesses the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking. It does not include the MCPTT call establishment time, since it measures the time previously defined when the request to speak is done during an ongoing call.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure time from request to speak to permission granted in a MCPTT call. The MCPTT access time calibration tests aims at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>25.11.2021</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEM</td></tr>
<tr><td>Scenario</td><td colspan="4">5G SA- LOS (Line of Sight). Radio configuration TDD, 100 MHz, 1 Carrier
4 layers, 4x4, 256QAM, Single beam The measurements are taken at the application level, in the Nemergent MCS application.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">-</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">NEM MCS applications and MCS server VNF, Nokia Airscale gNB, Athonet Rel. 15 5GC, OnePlus9 5G SA UEs</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">MCPTT</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Logcat Android log command-line tool</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">MCPTT Access time

MCPTT access time [ms]
<table><tr><td rowspan="2">Mean</td><td colspan="2">95% confidence interval for Mean</td></tr><tr><td>Lower bound</td><td>Upper bound</td></tr><tr><td>23,509</td><td>21,449</td><td>25,570</td></tr></table>
MCPTT access time [ms]
<table><tr><td rowspan="2">95% Percentile</td><td colspan="2">95% confidence interval for Min</td></tr><tr><td>Lower bound</td><td>Upper bound</td></tr><tr><td>35,723</td><td>32,625</td><td>38,822</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Test Case ID	*TC_MCPTTAccessTimeIncCallEstablishment_MAL*			
General description of the test	MCPTT end-to-end access time test, this test assesses the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking, including MCPTT call establishment and possibly acknowledgment from first receiving user before voice can be transmitted.			
Purpose	Measure time from request to speak to permission granted in a MCPTT call, including call establishment. The end-to-end MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further end-to-end MCPTT access time tests.			
Executed by	Partner:	UMA	Date:	20.12.2021
Involved Partner(s)	UMA, NEM			
Scenario	Polaris NetTest Rel. 15 Core using a Local breackout setup, in which the SGW and PGW (dataplane components) have been deployed in the Edge Data Center, while the rest of the EPC components are deployed at the Main Data center. Nokia Airscale eNB and gNB with LTE band 7 and 5G NR band 78 has been used as RAN. The measurements are taken at the application level, in the Nemergent MCS application.			
Slicing configuration	-			
Components involved *(e.g. HW components, SW components)*	NEM MCS applications and MCS server VNF, Nokia Airscale gNB, Polaris NetTest Rel. 15 EPC, Málaga Platform’s Edge Data Center deployment, OnePlus9 5G SA UEs			
Metric(s) under study (*Refer to those in Section 4*)	*MCPTT*			
Additional tools involved	Logcat Android log command-line tool			
Primary measurement results *(those included in the test case definition)*	MCPTT end-to-end access time (see tables below)			
Complementary measurement results	n/a			

MCPTT end-to-end access time

End-to-end MCPTT access time [ms]		
Mean	95% confidence interval for Mean	
	Lower bound	Upper bound
151,680	148,492	154,868

End-to-end MCPTT access time [ms]		
95% Percentile	95% confidence interval for Min	
	Lower bound	Upper bound
181,440	173,264	189,616

Test Case ID	TC_MCPTTAccessTime_MAL			
General description of the test	MCPTT Access time test, this test assesses the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking. It does not include the MCPTT call establishment time, since it measures the time previously defined when the request to speak is done during an ongoing call.			
Purpose	Measure time from request to speak to permission granted in a MCPTT call. The MCPTT access time calibration tests aims at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests.			
Executed by	Partner:	UMA	Date:	20.12.2021
Involved Partner(s)	UMA, NEM			
Scenario	Polaris NetTest Rel. 15 Core using a Local breackout setup, in which the SGW and PGW (dataplane components) have been deployed in the Edge Data Center, while the rest of the EPC components are deployed at the Main Data center. Nokia Airscale eNB and gNB with LTE band 7 and 5G NR band 78 has been used as RAN. The measurements are taken at the application level, in the Nemergent MCS application.			
Slicing configuration	-			
Components involved *(e.g. HW components, SW components)*	NEM MCS applications and MCS server VNF, Nokia Airscale gNB, Polaris NetTest Rel. 15 EPC, Málaga Platform's Edge Data Center deployment, OnePlus9 5G SA UEs			
Metric(s) under study (*Refer to those in Section 4*)	*MCPTT*			
Additional tools involved	Logcat Android log command-line tool			
Primary measurement results *(those included in the test case definition)*	MCPTT Access time (see tables below)			
Complementary measurement results	n/a			

MCPTT Access time

MCPTT access time [ms]		
Mean	95% confidence interval for Mean	
	Lower bound	Upper bound
26,304	25,865	26,743

MCPTT access time [ms]		
95% Percentile	95% confidence interval for Min	
	Lower bound	Upper bound
31,960	31,763	32,157

9.3. Limassol Platform Results

9.3.1. Generic tests – Results

9.3.1.1. 5G SA setup with Open5GS core co-located with the RAN

Test Case ID	TC_THR_Tcp			
General description of the test	DL Throughput for Edge node – UE traffic			
Purpose	Measure the DL throughput in 5G SA mode			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode DL throughput			
Slicing configuration	N/A			
Components involved	• UMA iPerf Agents • OpenTAP for automated testing (iPerf TAP plugin)			
Metric(s) under study	*Throughput*			
Additional tools involved	Iperf3			
Test Case Statistics	Mean: 111.74 +/- 5.24 Mbps Standard deviation: 32.8 +/- 2.49 Mbps Median: 120.63 +/- 7.12 Mbps Min: 18.32 +/- 1.39 Mbps Max: 147.88 +/- 4.95 Mbps 25% Percentile: 105.85 +/- 7.44 Mbps 75% Percentile: 130.72 +/- 6.94 Mbps 5% Percentile: 41.0 +/- 3.87 Mbps 95% Percentile: 142.38 +/- 5.36 Mbps			

Test Case ID	TC_THR_Tcp			
General description of the test	UL Throughput for Edge node – UE traffic			
Purpose	Measure the UL throughput in 5G SA mode			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode UL throughput			
Slicing configuration	N/A			
Components involved	• UMA iPerf Agents • OpenTAP for automated testing (iPerf TAP plugin)			
Metric(s) under study	*Throughput*			
Additional tools involved	Iperf3			

Test Case Statistics	Mean: 35.94 +/- 1.39 Mbps Standard deviation: 2.07 +/- 0.76 Mbps Median: 36.12 +/- 1.53 Mbps Min: 32.16 +/- 2.36 Mbps Max: 38.90 +/- 0.53 Mbps 25% Percentile: 34.51 +/- 1.97 Mbps 75% Percentile: 37.53 +/- 1.13 Mbps 5% Percentile: 32.96 +/- 2.19 Mbps 95% Percentile: 38.62 +/- 0.59 Mbps

Test Case ID	TC_RTT_e2e			
General description of the test	RTT for Edge node – UE			
Purpose	Measure RTT in 5G SA mode			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode RTT			
Slicing configuration	N/A			
Components involved	• UMA ping Agents • OpenTAP for automated testing (ping TAP plugin)			
Metric(s) under study	RTT			
Additional tools involved	ping			
Test Case Statistics	Mean: 13.70 +/- 0.62 ms Standard deviation: 4.81 +/- 5.34 ms Median: 13.70 +/- 0.62 ms Min: 9.65 +/- 0.34 ms Max: 49.71 +/- 54.33 ms 25% Percentile: 11.94 +/- 0.09 ms 75% Percentile: 14.60 +/- 0.11 ms 5% Percentile: 10.69 +/- 0.16 ms 95% Percentile: 16.41 +/- 0.55 ms			

9.3.1.2. 5G SA with Open5GS core fully backhauled over satellite

Test Case ID	TC_THR_Tcp			
General description of the test	DL Throughput for Edge node – UE traffic using Open5Gs core functions			
Purpose	Measure the DL throughput in 5G SA mode using Open5Gs			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode DL throughput using Open5Gs core functions			
Slicing configuration	N/A			
Components involved	• UMA iPerf Agents			

	• OpenTAP for automated testing (iPerf TAP plugin)
Metric(s) under study	*Throughput*
Additional tools involved	Iperf3, Open5Gs
Test Case Statistics	Mean: 2.22 +/- 0.09 Mbps Standard deviation: 1.11 +/- 0.07 Mbps Median: 2.22 +/- 0.13 Mbps Min: 0.35 +/- 0.10 Mbps Max: 4.56 +/- 0.29 Mbps 25% Percentile: 1.44 +/- 0.11 Mbps 75% Percentile: 2.92 +/- 0.15 Mbps 5% Percentile: 0.61 +/- 0.09 Mbps 95% Percentile: 3.93 +/- 0.16 Mbps

Test Case ID	TC_RTT_e2e			
General description of the test	RTT for Edge node – UE using Open5Gs core functions			
Purpose	Measure RTT in 5G SA mode using Open5Gs			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode RTT using Open5Gs core functions			
Slicing configuration	N/A			
Components involved	• UMA ping Agents • OpenTAP for automated testing (ping TAP plugin)			
Metric(s) under study	RTT			
Additional tools involved	Ping, Open5Gs			
Test Case Statistics	Mean: 838.34 +/- 2.27 ms Standard deviation: 38.48 +/- 2.00 ms Median: 822.80 +/- 2.89 ms Min: 794.80 +/- 2.70 ms Max: 972.60 +/- 26.98 ms 25% Percentile: 810.45 +/- 2.02 ms 75% Percentile: 860.11 +/- 4.96 ms 5% Percentile: 800.26 +/- 2.43 ms 95% Percentile: 912.30 +/- 3.66 ms			

9.3.1.3. 5G SA with Open5GS core, satellite backhaul and local break-out (LBO)

Test Case ID	TC_THR_Tcp			
General description of the test	DL Throughput for Edge node – UE traffic using local break out			
Purpose	Measure the DL throughput in 5G SA mode using local break out			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			

Scenario	5G SA mode DL throughput using local break out
Slicing configuration	N/A
Components involved	• UMA iPerf Agents • OpenTAP for automated testing (iPerf TAP plugin)
Metric(s) under study	*Throughput*
Additional tools involved	Iperf3
Test Case Statistics	Mean: 30.76 +/- 1.23 Mbps Standard deviation: 3.44 +/- 0.45 Mbps Median: 30.65 +/- 1.32 Mbps Min: 25.46 +/- 1.36 Mbps Max: 36.70 +/- 1.59 Mbps 25% Percentile: 28.80 +/- 1.37 Mbps 75% Percentile: 32.64 +/- 1.30 Mbps 5% Percentile: 26.29 +/- 1.31 Mbps 95% Percentile: 35.54 +/- 1.35 Mbps

Test Case ID	TC_RTT_e2e			
General description of the test	RTT for Edge node – UE traffic using local break out			
Purpose	Measure RTT in 5G SA mode traffic using local break out			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode RTT traffic using local break out			
Slicing configuration	N/A			
Components involved	• UMA ping Agents • OpenTAP for automated testing (ping TAP plugin)			
Metric(s) under study	RTT			
Additional tools involved	ping			
Test Case Statistics	Mean: 25.70 +/- 0.83 ms Standard deviation: 6.71 +/- 1.05 ms Median: 25.33 +/- 1.03 ms Min: 15.89 +/- 0.97 ms Max: 39.14 +/- 4.54 ms 25% Percentile: 21.09 +/- 1.02 ms 75% Percentile: 29.61 +/- 0.88 ms 5% Percentile: 16.99 +/- 0.89 ms 95% Percentile: 35.24 +/- 2.36 ms			

9.3.2. UC#1: 5G Maritime Communications – Results

9.3.2.1. End-to-end measurements

Test Case ID	TC_THR_Tcp			
General description of the test	DL Throughput for CoreDC – UE traffic measured on the vessel			
Purpose	Measure the DL throughput between CoreDC and UE in 5G SA mode measured on the vessel			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode DL throughput between CoreDC and UE measured on the vessel			
Slicing configuration	N/A			
Components involved	• UMA iPerf Agents • OpenTAP for automated testing (iPerf TAP plugin)			
Metric(s) under study	*Throughput*			
Additional tools involved	Iperf3			
Test Case Statistics	Mean: 0.67 +/- 0.04 Mbps Standard deviation: 0.40 +/- 0.03 Mbps Median: 0.64 +/- 0.06 Mbps Min: 0.12 +/- 0.01 Mbps Max: 1.41 +/- 0.12 Mbps 25% Percentile: 0.34 +/- 0.04 Mbps 75% Percentile: 0.94 +/- 0.05 Mbps 5% Percentile: 0.16 +/- 0.02 Mbps 95% Percentile: 1.25 +/- 0.09 Mbps			

Test Case ID	TC_RTT_e2e			
General description of the test	RTT for CoreDC – UE traffic measured on the vessel			
Purpose	Measure RTT in 5G SA mode between CoreDC and UE in 5G SA mode measured on the vessel			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode RTT between CoreDC and UE measured on the vessel			
Slicing configuration	N/A			
Components involved	• UMA ping Agents • OpenTAP for automated testing (ping TAP plugin)			
Metric(s) under study	RTT			
Additional tools involved	Ping			
Test Case Statistics	Mean: 687.94 +/- 4.60 ms Standard deviation: 27.74 +/- 7.11 ms Median: 681.48 +/- 4.16 ms			

	Min: 662.00 +/- 4.44 ms Max: 750.33 +/- 23.61 ms 25% Percentile: 671.29 +/- 3.98 ms 75% Percentile: 695.64 +/- 6.14 ms 5% Percentile: 663.82 +/- 4.43 ms 95% Percentile: 730.63 +/- 14.82 ms

9.3.2.2. Local measurements

Test Case ID	TC_THR_Tcp			
General description of the test	DL Throughput for Edge node – UE traffic on the vessel			
Purpose	Measure the DL throughput in 5G SA mode on the vessel			
Executed by	Partner:	SHC	Date:	10/12/2021
Involved Partner(s)	PLC, AVA, EKI, SRL, UMA			
Scenario	5G SA mode DL throughput measured on vessel			
Slicing configuration	N/A			
Components involved	• UMA iPerf Agents • OpenTAP for automated testing (iPerf TAP plugin)			
Metric(s) under study	*Throughput*			
Additional tools involved	Iperf3			
Test Case Statistics	Mean: 164.30 +/- 43.35 Mbps Standard deviation: 57.32 +/- 13.53 Mbps Median: 180.68 +/- 46.13 Mbps Min: 48.52 +/- 31.78 Mbps Max: 228.69 +/- 51.19 Mbps 25% Percentile: 143.52 +/- 45.24 Mbps 75% Percentile: 200.73 +/- 46.77 Mbps 5% Percentile: 71.86 +/- 39.08 Mbps 95% Percentile: 220.85 +/- 49.89 Mbps			

9.4. Surrey Platform Results

9.4.1. Generic tests – Results

No generic tests were executed during Phase 3, as these were performed in Phase 2 and reported in D6.2.

9.4.2. UC#1: Multi-RAT Support for Sensor Measurements –Results

Test Case ID	TC_IoT_PacketDelay_WIFI_SUR			
General description of the test	The main KPI is ‘packet delays’. Under heavy load, the servers are able to receive, treat (decoding JSON and storing data in MySQL data base) and answer all 3 protocols requests without accumulating delays. The three protocols (HTTP POST, MQTT and CoAP) are used relying on traditional WiFi Access Point for the radio. The number of JSON data packets per board is known in advance (i.e., 1500). Only one slice is used and each protocol is covered by 10 boards.			
Purpose	Check performance (packet delay) and reliability (packet loss) of the servers when having to handle large number of packets using 3 protocols simultaneously.			
Executed by	Partner:	UNIS/INF	Date:	01/11/2021
Involved Partner(s)	Same as above			
Scenario	See the *TC_IoT_PacketDelay_WIFI_SUR* test case description			
Slicing configuration	n/a			
Components involved *(e.g. HW components, SW components)*	Surrey Platform: openStack, SDN switch, Apache server, mySQL + pycom/pysense boards			
Metric(s) under study (*Refer to those in Section 4*)	Packet delay & packet loss			
Additional tools involved	n/a			
Primary measurement results for HTTP POST *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	15000	0	140.6
Primary measurement results for CoAP *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	15000	0	116.5
Primary measurement results for MQTT *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	14407	593	142.3
Complementary measurement results	n/a			

Test Case ID	TC_IoT_PacketDelay_5G_SUR
General description of the test	The main KPI is ‘packet delays’. Under heavy load, the servers are able to receive, treat (decoding JSON and storing data in MySQL data base) and answer all 3 protocols requests without accumulating delays. The three protocols (HTTP POST, MQTT and CoAP) are used relying on 5G for the radio.

	The number of JSON data packets per board is known in advance (i.e., 1500). Only one slice is used and each protocol is covered by 10 boards.			
Purpose	Check performance (packet delay) and reliability (packet loss) of the servers when having to handle large number of packets using 3 protocols simultaneously.			
Executed by	Partner:	UNIS/INF	Date:	03/11/2021
Involved Partner(s)	Same as above			
Scenario	See the *TC_IoT_PacketDelay_5G_SUR* test case description			
Slicing configuration	n/a			
Components involved *(e.g. HW components, SW components)*	Surrey Platform: openStack, SDN switch, Apache server, mySQL + pycom/pysense boards			
Metric(s) under study (*Refer to those in Section 4*)	Packet delay & packet loss			
Additional tools involved	n/a			
Primary measurement results for HTTP POST *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	15000	0	163.8
Primary measurement results for MQTT *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	14754	246	155.8
Primary measurement results for CoAP *(those included in the test case definition)*	Total Number of Sent Packets	Total Number of Rec. Packets	Lost	Average Delay (ms)
	15000	14863	137	130.1
Complementary measurement results	n/a			

Test Case ID	TC_IoT_PacketDelay_WIFI/5G_2SLICES_SUR			
General description of the test	The main KPI is 'packet delays'. Under heavy load, the servers are able to receive, treat (decoding JSON and storing data in MySQL data base) and answer both protocols requests (CoAP and MQTT) without accumulating delays. WIFI and 5G CPE access point are used respectively using 2 slices.			
Purpose	Check performance (packet delay) and reliability (packet loss) of the servers when having to handle large number of packets using 2 protocols over 2 RATs simultaneously.			
Executed by	Partner:	UNIS/INF	Date:	03/12/2021
Involved Partner(s)	Same as above			
Scenario	See the *TC_IoT_PacketDelay_WIFI/5G_2SLICES_SUR* test case description			
Slicing configuration	2 slices			

Components involved *(e.g. HW components, SW components)*	Surrey platform: openStack, SDN switch, Apache server, mySQL + pycom/pysense boards
Metric(s) under study (*Refer to those in Section 4*)	Packet delay & packet loss
Additional tools involved	n/a
Primary measurement results for Slice 1 (CoAP/WiFi) *(those included in the test case definition)*	Total Number of Sent Packets: 22500 Total Number of Rec. Packets: 22413 Lost: 87 Average Delay (ms): 236.07
Primary measurement results for Slice 2 (MQTT/5G) *(those included in the test case definition)*	Total Number of Sent Packets: 22500 Total Number of Rec. Packets: 21834 Lost: 666 Average Delay (ms): 128.07
Complementary measurement results	n/a

9.4.3. UC#2: Coverage Evaluation–Results

Test Case ID	TC_COVERAGE_DL_SURREY			
General description of the test	This test calculates the Maximum Coupling Loss in different parts of the University of Surrey campus based on the gNB transmission power and the received SINR.			
Purpose	To evaluate the downlink 5G coverage in the campus of the University of Surrey (home site of the Surrey Platform)			
Executed by	Partner:	UNIS	Date:	July 2021
Involved Partner(s)	UNIS			
Scenario	The scenario included One 5G CPE, with multiple measurements (50 iterations) over 5 sites.			
Slicing configuration	N/A			
Components involved *(e.g. HW components, SW components)*	– gNBs – 5G CPE – Laptop (results storage/post-processing)			
Metric(s) under study (*Refer to those in Section 4*)	*Maximum Coupling Loss (Primary)* *PDCP-Level Throughput/SINR/RSRP (Complementary)*			
Additional tools involved	N/A			

	Maximum Coupling Loss (calc. per Site)						
	#	MCL (dB) - [DL Calc.]	MCL (dB) - [DL Calc.]	MCL (dB) - [DL Calc.]	MCL (dB) - [DL Calc.]	MCL (dB) - [DL Calc.]	MCL (dB) - [DL Calc.]
		site # 25	site # 26	site # 27	site # 28	site # 29	site # 30
Primary measurement results *(those included in the test case definition)*	1	135.0	123.0	130.0	120.0	142.0	126.0
	2	128.0	142.0	126.0	128.0	147.0	135.0
	3	144.0	142.0	146.0	139.0	137.0	132.0
	4	134.0	120.0	121.0	128.0	147.0	130.0
	5	119.0	132.0	148.0	138.0	137.0	142.0
	6	142.0	138.0	142.0	126.0	144.0	142.0
	7	119.0	144.0	133.0	146.0	148.0	123.0
	8	140.0	123.0	145.0	146.0	146.0	143.0
	9	122.0	143.0	137.0	138.0	144.0	131.0
	10	130.0	141.0	124.0	132.0	144.0	148.0
	11	145.0	126.0	138.0	119.0	121.0	142.0
	12	145.0	123.0	119.0	127.0	133.0	147.0
	13	142.0	138.0	120.0	133.0	124.0	143.0
	14	133.0	139.0	119.0	124.0	138.0	123.0
	15	124.0	140.0	133.0	138.0	142.0	139.0
	16	119.0	144.0	127.0	143.0	120.0	134.0
	17	132.0	136.0	138.0	144.0	134.0	145.0
	18	126.0	122.0	147.0	146.0	148.0	149.0
	19	143.0	121.0	135.0	136.0	128.0	135.0
	20	143.0	119.0	128.0	140.0	147.0	149.0
	21	134.0	131.0	124.0	140.0	146.0	122.0
	22	132.0	135.0	119.0	125.0	145.0	146.0
	23	119.0	149.0	123.0	131.0	130.0	142.0
	24	137.0	123.0	134.0	125.0	132.0	128.0
	25	133.0	123.0	122.0	141.0	135.0	119.0
	26	143.0	120.0	147.0	129.0	130.0	140.0
	27	120.0	146.0	142.0	138.0	143.0	126.0
	28	140.0	137.0	144.0	128.0	124.0	134.0
	29	138.0	146.0	125.0	123.0	126.0	149.0
	30	119.0	132.0	128.0	145.0	140.0	121.0
	31	132.0	120.0	135.0	148.0	128.0	124.0
	32	136.0	138.0	127.0	126.0	124.0	140.0
	33	139.0	143.0	120.0	146.0	128.0	145.0
	34	146.0	148.0	123.0	147.0	119.0	129.0
	35	141.0	123.0	124.0	135.0	149.0	141.0
	36	130.0	135.0	135.0	143.0	127.0	147.0
	37	138.0	147.0	149.0	138.0	130.0	120.0
	38	133.0	134.0	127.0	124.0	147.0	140.0
	39	133.0	131.0	126.0	145.0	132.0	134.0
	40	142.0	126.0	125.0	127.0	126.0	144.0
	41	147.0	125.0	125.0	127.0	140.0	144.0
	42	145.0	136.0	123.0	144.0	145.0	131.0
	43	132.0	128.0	124.0	132.0	145.0	119.0
	44	127.0	127.0	141.0	123.0	129.0	148.0
	45	149.0	138.0	136.0	131.0	122.0	133.0
	46	135.0	121.0	123.0	131.0	137.0	138.0
	47	148.0	138.0	149.0	132.0	149.0	136.0

	48	149.0	147.0	146.0	139.0	144.0	140.0
	49	127.0	136.0	137.0	122.0	144.0	121.0
	50	126.0	142.0	133.0	128.0	144.0	143.0

Complementary measurement results

PDCP-Level Throughput (Mb/s)

#	PDCP-level Throughput (Mbps) - [DL measured]	PDCP-level Throughput (Mbps) - [DL measured]	PDCP-level Throughput (Mbps) - [DL measured]	PDCP-level Throughput (Mbps) - [DL measured]	PDCP-level Throughput (Mbps) - [DL measured]	PDCP-level Throughput (Mbps) - [DL measured]
	site # 25	site # 26	site # 27	site # 28	site # 29	site # 30
1	664	833	912	654	791	928
2	783	828	733	640	700	894
3	673	644	727	849	699	878
4	639	809	709	878	999	944
5	912	627	686	972	844	714
6	758	792	670	734	835	937
7	601	965	815	668	704	726
8	843	648	729	816	739	914
9	889	837	822	876	720	677
10	976	728	623	762	696	980
11	686	613	899	919	1000	900
12	802	711	684	991	978	933
13	711	943	627	906	891	912
14	716	664	733	943	714	844
15	824	989	765	796	676	606
16	811	862	709	709	651	680
17	809	706	968	924	777	655
18	970	808	684	706	677	734
19	722	686	782	1000	616	805
20	827	945	726	604	632	682
21	654	853	776	728	818	717
22	711	722	672	712	970	807
23	849	715	768	780	885	619
24	660	784	946	703	781	986
25	924	814	862	760	980	626
26	873	624	858	759	728	691
27	936	958	823	891	667	763
28	880	809	936	720	687	934
29	974	656	971	736	789	711
30	978	800	957	896	938	611
31	906	934	982	838	753	886
32	966	884	916	794	700	668
33	730	887	845	639	631	721
34	960	823	697	795	689	994
35	864	841	804	683	880	862
36	713	609	671	745	978	739
37	940	823	939	983	690	601
38	693	660	902	984	713	895
39	816	811	886	694	966	651
40	798	765	847	782	980	782

	41	609	697	899	637	774	997
	42	989	875	612	955	676	987
	43	901	761	793	729	764	642
	44	701	930	977	876	889	641
	45	654	849	939	895	899	753
	46	749	633	617	905	946	688
	47	882	896	871	875	795	743
	48	916	790	918	895	754	931
	49	633	944	602	731	853	915
	50	627	761	980	894	992	764

9.4.4. UC#3: WSMP – Results

Test Case ID	TC_WiFi_SCT_e2e			
General description of the test	The test calculates the elapsed time between a new WiFi Slice request is received in the WiFi Slice Controller and the moment the service is up and running. Thus, a new WiFi Slice request is received, the WSC notifies the AP Manager with the required information to set up the WiFi Slice. Meanwhile, the monitoring service (WSAM) is actively monitoring the AP Manager. This process is repeated until at least one AP has no more slice slots available.			
Purpose	Measure WiFi RAT contribution to SCT			
Executed by	Partner:	FON	Date:	12.08.2021
Partner(s)	UNIS, FON			
Scenario	5G NSA			
Slicing configuration	Proactive scheduling activated			
Components	APs (2x Ruckus 550, 2x Ruckus 650), UEs (OnePlus Nord, LG K92 5G)			
Metric(s) under study	*Service Creation Time*			
Additional tools	Ruckus vSZ 5.2.0			
Primary measurement results	**Service Creation Time [s]** Cold start phase (0-3 pre-existing WiFi Slices) Mean: 13.006 s Median: 12.764 s Standard deviation: 4.08 s Medium load phase (4-7 pre-existing WiFi Slices) Mean: 10.506 s Median: 10.790 s Standard deviation: 4.058 s Heavy load phase (8+ pre-existing WiFi Slices) Mean: 23.360 s Median: 24.502 s Standard deviation: 5.378 s			

Complementary measurement results	-

9.4.5. UC#4: CoAP over LTE and 5G Radios – Results

Test Case ID	TC_RTT_COAP_SUR			
General description of the test	The main KPI is 'round trip time'. The CoAP client sends and receive messages using different access technologies and various loads. The round trip time is calculated from the different between send time and receive time.			
Purpose	Check the performance of LTE/5G radio interface for CoAP based transmission. Check *TC_RTT_COAP_SUR* test case description out for more detail			
Executed by	Partner:	KAU	Date:	21/11/2021
Involved Partner(s)	KAU and UNIS			
Scenario	See the test case description			
Slicing configuration	n/a			
Components involved *(e.g. HW components, SW components)*	Surrey platform: MONROE node (LTE modem), 5G CPE (Ethernet connected with MONROE node), 4G eNB, MONROE subsystem, in-house python based CoAP client. Karlstad University: In house python based CoAP echo server.			
Metric(s) under study (*Refer to those in Section 4*)	Round trip time			
Additional tools involved	n/a			
Primary measurement results *(those included in the test case definition)*	See table below			

Average and Median CoAP RTT

Interface	MsgLength[Bytes]	MsgInterval[s]	Mean RTT ± CI [ms]	Median RTT [ms]
5G	100	0.5	59.5 ± 1.7	56.9
		1	70.8 ± 1.9	71.5
		2	72.1 ± 1.6	72.8
	200	0.5	62.8 ± 3.0	57.0
		1	69.3 ± 1.5	71.5
		2	71.1 ± 1.5	72.3
	400	0.5	71.3 ± 1.6	71.3
		1	74.1 ± 2.1	73.7
		2	74.7 ± 1.8	73.7
LTE	100	0.5	62.5 ± 1.3	59.9
		1	62.5 ± 1.3	59.7
		2	62.3 ± 1.3	60.6

	200	0.5	67.3 ± 1.6	64.5
		1	67.0 ± 1.4	64.2
		2	66.8 ± 1.5	64.8
	400	0.5	68.8 ± 1.7	66.1
		1	69.1 ± 2.3	65.2
		2	68.0 ± 1.3	65.4

Complementary measurement results	**Average and Median Ping RTT**		
	Interface	Mean Ping RTT [ms] ± CI	Median Ping RTT [ms]
	5G	45.4 ± 0.34	44.2
	LTE	43.2 ± 0.16	43.2

9.5. Berlin Platform Results

9.5.1. Generic tests – Results

9.5.1.1. Intra-Compute Throughput

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between two VMs running on the same compute node of the Berlin Platform test bed infrastructure			
Purpose	The test assesses the performance of the end to end connection and the of the components deployed in the field trial.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			
Primary measurement results	**Throughput**	**Compute 101 (a to d)**		
	Mean:	31579.856304	+/- 772.066962	Mbps
	Standard deviation:	1510.712492	+/- 184.357267	
	Median:	31638.76	+/- 814.735065	
	Min:	25106.08	+/- 1706.864141	

	Max:	36184.76	+/- 995.974280	
	25% Percentile:	30807.42	+/- 895.766554	
	75% Percentile:	32546.9	+/- 678.397590	
	5% Percentile:	29148.24	+/- 845.687591	
	95% Percentile:	33613.246	+/- 651.483253	
	Throughput	Compute 101 (d to a)		
	Mean:	20329.825869	+/- 148.608022	
	Standard deviation:	361.037554	+/- 155.893152	
	Median:	20394.46	+/- 149.159215	
	Min:	18351.04	+/- 742.156643	
	Max:	21064.24	+/- 151.256603	Mbps
	25% Percentile:	20238.65	+/- 150.773540	
	75% Percentile:	20511.76	+/- 147.729220	
	5% Percentile:	19585.932	+/- 663.251379	
	95% Percentile:	20679.3	+/- 137.537051	
	Throughput	Compute 104 (a to b)		
	Mean:	29037.507556	+/- 827.108858	
	Standard deviation:	2334.725703	+/- 400.021714	
	Median:	29452.36	+/- 752.880604	
	Min:	22673.4	+/- 1371.804472	
	Max:	33662.68	+/- 568.787886	Mbps
	25% Percentile:	27106.95	+/- 1249.917938	
	75% Percentile:	30825.52	+/- 603.0137491	
	5% Percentile:	25481.408	+/- 1332.695347	
	95% Percentile:	32222.378	+/- 522.488472	
	Throughput	Compute 104 (b to a)		
	Mean:	20075,45444	+/- 131,1217364	
	Standard deviation:	380,1807832	+/- 96,80719556	
	Median:	20101,74	+/- 132,5896761	
	Min:	17997,4	+/- 671,9571214	
	Max:	21143,6	+/- 201,7016784	Mbps
	25% Percentile:	19920,77	+/- 138,6586013	
	75% Percentile:	20281,07	+/- 123,7096502	
	5% Percentile:	19572,99	+/- 161,9695516	
	95% Percentile:	20531,676	+/- 110,2262523	
	Throughput	Compute 104 (a to c)		
	Mean:	20212.795	+/- 161.246826	
	Standard deviation:	521.580169	+/- 155.358581	
	Median:	20304.98	+/- 133.611962	Mbps
	Min:	17633.44	+/- 917.143049	
	Max:	21320.32	+/- 202.342798	
	25% Percentile:	20117.87	+/- 139.507648	

	75% Percentile:	20465.93	+/- 131.919629	
	5% Percentile:	19410.374	+/- 520.390730	
	95% Percentile:	20699.766	+/- 131.468903	
	Throughput	Compute 104 (c to a)		
	Mean:	20012,01174	+/- 131,3201491	
	Standard deviation:	364,4477092	+/- 28,9567432	
	Median:	20046,52	+/- 128,8001205	
	Min:	17759	+/- 538,7144916	
	Max:	20989,16	+/- 217,8309332	Mbps
	25% Percentile:	19876,8	+/- 124,8343924	
	75% Percentile:	20210,26	+/- 135,6069294	
	5% Percentile:	19477,986	+/- 146,7587678	
	95% Percentile:	20448,138	+/- 145,3040656	
	Throughput	Compute 105 (e to f 0)		
	Mean:	32728,78064	+/- 714,8266151	
	Standard deviation:	2979,024803	+/- 439,3065489	
	Median:	32731,62	+/- 589,8549371	
	Min:	25312,52	+/- 1064,371025	
	Max:	42061,08	+/- 865,7638554	Mbps
	25% Percentile:	30835,7	+/- 1005,877652	
	75% Percentile:	34326,14	+/- 662,2234393	
	5% Percentile:	28276,46	+/- 1180,816339	
	95% Percentile:	37803,99	+/- 1163,60242	
	Throughput	Compute 105 (e to f 1)		
	Mean:	33138,71085	+/- 801,5801758	
	Standard deviation:	3007,319869	+/- 516,7194304	
	Median:	33025,54	+/- 635,3847758	
	Min:	25790,6	+/- 1101,970514	
	Max:	42634,16	+/- 1074,688996	
	25% Percentile:	31250,15	+/- 1112,212114	
	75% Percentile:	34685,39	+/- 701,7371994	
	5% Percentile:	28747,694	+/- 1335,479918	
	95% Percentile:	38478,724	+/- 1390,595401	
	Throughput	Compute 105 (e to f 2)		
	Mean:	33129,80255	+/- 635,7635944	
	Standard deviation:	3042,617432	+/- 499,6187805	
	Median:	32905,9	+/- 568,3270799	
	Min:	25278,64	+/- 2206,477295	
	Max:	42016,08	+/- 1252,481389	
	25% Percentile:	31326,97	+/- 866,3351012	
	75% Percentile:	34668,16	+/- 646,8749413	
	5% Percentile:	28649,508	+/- 1097,22094	

	95% Percentile:	38258,106	+/- 1364,966365
	Throughput	Compute 105 (e to f 3)	
	Mean:	15876,35872	+/- 4032,635559
	Standard deviation:	1442,868506	+/- 1114,928058
	Median:	16327,5	+/- 4296,531592
	Min:	12462,96	+/- 2378,720404
	Max:	18685,72	+/- 5801,281775
	25% Percentile:	14515,2	+/- 3517,124857
	75% Percentile:	16974,25	+/- 4691,307559
	5% Percentile:	13843,78	+/- 3051,975619
	95% Percentile:	17620,678	+/- 5098,520209
	Throughput	Compute 105 (f to e 0)	
	Mean:	15876,35872	+/- 4032,635559
	Standard deviation:	1442,868506	+/- 1114,928058
	Median:	16327,5	+/- 4296,531592
	Min:	12462,96	+/- 2378,720404
	Max:	18685,72	+/- 5801,281775
	25% Percentile:	14515,2	+/- 3517,124857
	75% Percentile:	16974,25	+/- 4691,307559
	5% Percentile:	13843,78	+/- 3051,975619
	95% Percentile:	17620,678	+/- 5098,520209
	Throughput	Compute 105 (f to e 1)	
	Mean:	9348,953191	+/- 3,059129973
	Standard deviation:	11,22605379	+/- 4,749194364
	Median:	9351,04	+/- 0,403737105
	Min:	9270,88	+/- 29,1500423
	Max:	9363,24	+/- 0,382053584
	25% Percentile:	9344,94	+/- 6,018446994
	75% Percentile:	9354,56	+/- 0,284613532
	5% Percentile:	9335,13	+/- 12,70757725
	95% Percentile:	9359,824	+/- 0,363539196
	Throughput	Compute 105 (f to e 2)	
	Mean:	9350,834468	+/- 0,388961624
	Standard deviation:	6,718049101	+/- 1,670487295
	Median:	9351,34	+/- 0,175936508
	Min:	9304,84	+/- 13,72954386
	Max:	9363	+/- 0,734546882
	25% Percentile:	9348,04	+/- 0,187652157
	75% Percentile:	9354,47	+/- 0,200545389
	5% Percentile:	9342,898	+/- 0,576830418
	95% Percentile:	9359,626	+/- 0,232675006
	Throughput	Compute 105 (f to e 3)	

	Mean:	22995,87809	+/- 4704,515259	
	Standard deviation:	2129,564163	+/- 830,6999485	
	Median:	23012,12	+/- 4713,912624	
	Min:	17314,6	+/- 3417,501984	
	Max:	30348,68	+/- 6703,90762	
	25% Percentile:	21942,49	+/- 4375,499808	
	75% Percentile:	23862,06	+/- 5000,912966	
	5% Percentile:	19532,74	+/- 3917,885476	
	95% Percentile:	26639,078	+/- 6018,286193	
	Throughput	**Compute 106 (a to b)**		
	Mean:	30619,82267	+/- 525,641912	
	Standard deviation:	1813,247968	+/- 190,2169399	
	Median:	30828,06	+/- 662,8694759	
	Min:	21727,44	+/- 1305,944217	
	Max:	34709,96	+/- 537,8984983	Mbps
	25% Percentile:	29851,49	+/- 585,5184352	
	75% Percentile:	31766,25	+/- 568,4309386	
	5% Percentile:	27463,652	+/- 571,8220941	
	95% Percentile:	32757,352	+/- 501,0404832	
	Throughput	**Compute 106 (b to a)**		
	Mean:	19977,94378	+/- 145,095415	
	Standard deviation:	372,152591	+/- 38,51636545	
	Median:	20005,74	+/- 143,656509	
	Min:	17864,32	+/- 554,9442231	
	Max:	21025,36	+/- 197,9428245	Mbps
	25% Percentile:	19802,31	+/- 144,9814166	
	75% Percentile:	20193,4	+/- 144,0833107	
	5% Percentile:	19441,616	+/- 156,2744412	
	95% Percentile:	20471,24	+/- 146,0561053	
Complementary measurement results	n/a			

9.5.1.2. Intra-Compute Two Parallel Streams

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between two pairs of VMs running on the same compute node of the Berlin Platform test bed infrastructure. All measurements are run in parallel.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			

Scenario	
Slicing configuration	An exclusive slice / software-defined network is used for the test.
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement
Metric(s) under study	*Throughput*
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints
Primary measurement results	**Throughput** **Compute 105 (a to b)** Mean: 17184,41888 +/- 8099,280588 Standard deviation: 3391,077525 +/- 2526,436468 Median: 16294,16667 +/- 8574,089483 Min: 11232,66667 +/- 6811,867976 Max: 25938,66667 +/- 9337,085483 Mbps 25% Percentile: 15018,3125 +/- 8327,306523 75% Percentile: 20488,91667 +/- 8597,386764 5% Percentile: 13107,30833 +/- 7663,493062 95% Percentile: 22623,54583 +/- 9279,482482 **Throughput** **Compute 105 (c to d)** Mean: 16981,28191 +/- 8307,101753 Standard deviation: 2498,713905 +/- 2243,690015 Median: 17672,70833 +/- 8693,934733 Min: 11517,58333 +/- 7039,946898 Max: 22192,66667 +/- 8532,583117 Mbps 25% Percentile: 15002,5625 +/- 8576,952894 75% Percentile: 19032,3125 +/- 8749,328216 5% Percentile: 13645,25 +/- 8092,357678 95% Percentile: 20221,17083 +/- 8908,051173
Complementary measurement results	n/a

9.5.1.3. Intra Compute TCP Throughput 3 Parallel Streams

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between three pairs of VMs running on the same compute node of the Berlin Platform test bed infrastructure. All measurements are run in parallel.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			

Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			
Primary measurement results	**Throughput**	**Compute 106 (a to b)**		
	Mean:	19863,31723	+/- 4326,343483	
	Standard deviation:	2043,311723	+/- 1003,263247	
	Median:	19644,34	+/- 4559,13611	
	Min:	13858	+/- 2785,167065	
	Max:	24930,24	+/- 5834,246014	Mbps
	25% Percentile:	18818,75	+/- 4202,814593	
	75% Percentile:	20997,44	+/- 4762,415152	
	5% Percentile:	16912,252	+/- 3453,149618	
	95% Percentile:	22850,58	+/- 5063,56411	
	Throughput	**Compute 106 (c to d)**		
	Mean:	22957,7966	+/- 4013,408271	
	Standard deviation:	2790,685539	+/- 1224,182478	
	Median:	23894,1	+/- 4097,606386	
	Min:	15111,52	+/- 3527,304018	
	Max:	27824,2	+/- 3960,894878	Mbps
	25% Percentile:	21277,18	+/- 4275,246708	
	75% Percentile:	24730,22	+/- 4263,314303	
	5% Percentile:	17919,548	+/- 4218,620892	
	95% Percentile:	25798,196	+/- 4457,592706	
	Throughput	**Compute 106 (e to f)**		
	Mean:	24160,04106	+/- 4257,578883	
	Standard deviation:	2096,176628	+/- 1016,515645	
	Median:	24675,34	+/- 4275,510573	
	Min:	18493,88	+/- 3887,419603	
	Max:	28664,2	+/- 4488,265859	Mbps
	25% Percentile:	22531,26	+/- 4397,121956	
	75% Percentile:	25598,89	+/- 4458,874799	
	5% Percentile:	20585,044	+/- 4296,560982	
	95% Percentile:	26570,94	+/- 4654,125613	
Complementary measurement results	n/a			

9.5.1.4. Inter-Compute TCP Throughput

Test Case ID	TC_THR_Tcp

General description of the test	Throughput between a pair of, where each one runs on a different compute node of the Berlin Platform test bed infrastructure.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			

Primary measurement results

Throughput	**Compute 101 to 102**		
Mean:	9344,748889	+/- 2,152385775	
Standard deviation:	20,77159763	+/- 7,677972526	
Median:	9350,68	+/- 0,613062415	
Min:	9220,56	+/- 69,34740936	
Max:	9364,44	+/- 0,975942538	Mbps
25% Percentile:	9341,88	+/- 2,558651598	
75% Percentile:	9354,72	+/- 0,283112923	
5% Percentile:	9315,01	+/- 6,905095407	
95% Percentile:	9360,058	+/- 0,322373644	
Throughput	**Compute 101 to 103**		
Mean:	9342,608889	+/- 5,544495233	
Standard deviation:	17,99759743	+/- 5,53546239	
Median:	9348,62	+/- 4,008818584	
Min:	9267,2	+/- 22,87391876	
Max:	9364,2	+/- 0,684517713	Mbps
25% Percentile:	9335,68	+/- 9,242598491	
75% Percentile:	9354,71	+/- 0,454409348	
5% Percentile:	9307,104	+/- 16,40797833	
95% Percentile:	9360,488	+/- 0,357044509	
Throughput	**Compute 101 to 104**		
Mean:	9351,007333	+/- 2,139974083	
Standard deviation:	9,420959205	+/- 3,902276143	
Median:	9352,76	+/- 0,430130193	
Min:	9294,24	+/- 14,35573621	
Max:	9364	+/- 0,445853047	Mbps
25% Percentile:	9348,98	+/- 1,262089862	
75% Percentile:	9355,88	+/- 0,389688972	
5% Percentile:	9335,328	+/- 14,08353713	

	95% Percentile:	9360,964	+/- 0,220888577
	Throughput	**Compute 101 to 105**	
	Mean:	9344,585556	+/- 0,623783058
	Standard deviation:	17,71197333	+/- 2,588296028
	Median:	9349,82	+/- 0,314589418
	Min:	9245,72	+/- 41,68164291
	Max:	9363,72	+/- 0,512800231
	25% Percentile:	9342,48	+/- 0,63188112
	75% Percentile:	9354,02	+/- 0,293274587
	5% Percentile:	9307,952	+/- 3,843561928
	95% Percentile:	9359,888	+/- 0,245601038
	Throughput	**Compute 101 to 106**	
	Mean:	9344,103111	+/- 1,108047266
	Standard deviation:	20,75418807	+/- 2,601641991
	Median:	9351,52	+/- 0,312324526
	Min:	9249,24	+/- 23,30173467
	Max:	9364,28	+/- 0,577891597
	25% Percentile:	9345,23	+/- 1,50399306
	75% Percentile:	9355,21	+/- 0,259154224
	5% Percentile:	9301,184	+/- 4,597144755
	95% Percentile:	9360,456	+/- 0,264060745
Complementary measurement results	n/a		

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between a pair of VMs, where each runs on a different compute node of the Berlin Platform test bed infrastructure.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			

Primary measurement results	**Throughput**	**Compute 104 to 101**		
	Mean:	9344,061889	+/- 1,259658957	
	Standard deviation:	22,40386375	+/- 4,661637209	Mbps
	Median:	9351,3	+/- 0,326331016	

	Min:	9181,88	+/- 80,87869774	
	Max:	9364,32	+/- 0,831728074	
	25% Percentile:	9342,85	+/- 2,299804183	
	75% Percentile:	9355,02	+/- 0,224750246	
	5% Percentile:	9308,638	+/- 5,581463402	
	95% Percentile:	9360,446	+/- 0,24031497	
	Throughput	**Compute 104 to 102**		
	Mean:	9343,955556	+/- 1,216646116	
	Standard deviation:	19,4259962	+/- 2,427388301	
	Median:	9351,06	+/- 0,378318855	
	Min:	9257,16	+/- 26,63913458	
	Max:	9364,28	+/- 1,007723703	Mbps
	25% Percentile:	9341,18	+/- 3,512990392	
	75% Percentile:	9354,81	+/- 0,200456867	
	5% Percentile:	9306,754	+/- 3,008563842	
	95% Percentile:	9360,29	+/- 0,302744209	
	Throughput	**Compute 104 to 103**		
	Mean:	9351,696889	+/- 0,868873443	
	Standard deviation:	11,86038744	+/- 4,079008127	
	Median:	9353,02	+/- 0,241866839	
	Min:	9264,28	+/- 30,49022057	
	Max:	9363,96	+/- 0,325895617	Mbps
	25% Percentile:	9349,79	+/- 0,357080299	
	75% Percentile:	9356,2	+/- 0,291879334	
	5% Percentile:	9344,246	+/- 1,267393373	
	95% Percentile:	9361,17	+/- 0,256815115	
	Throughput	**Compute 104 to 105**		
	Mean:	9346,305333	+/- 1,614772498	
	Standard deviation:	13,02715989	+/- 2,505612859	
	Median:	9350,42	+/- 0,4865142	
	Min:	9281	+/- 17,65809662	
	Max:	9362,4	+/- 0,491306129	
	25% Percentile:	9343,22	+/- 2,385477493	
	75% Percentile:	9353,98	+/- 0,343449506	
	5% Percentile:	9322,848	+/- 6,670101334	
	95% Percentile:	9359,054	+/- 0,410707171	
	Throughput	**Compute 104 to 106**		
	Mean:	9351,412667	+/- 1,381548146	
	Standard deviation:	8,030577259	+/- 2,158415486	
	Median:	9352,8	+/- 0,342258857	
	Min:	9309,44	+/- 18,40116524	
	Max:	9365,2	+/- 1,473918386	

	25% Percentile: 9348,37 +/- 1,587302149 75% Percentile: 9355,94 +/- 0,260179498 5% Percentile: 9338,81 +/- 5,993105604 95% Percentile: 9360,808 +/- 0,312874134
Complementary measurement results	n/a

9.5.1.5. Inter-compute Parallel Streams

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between two pairs of VMs running on different compute nodes of the Berlin Platform test bed infrastructure. All measurements are run in parallel.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			
Primary measurement results	**Throughput** **Compute 101 to 102** Mean: 9343,631702 +/- 4,310418294 Standard deviation: 12,36225331 +/- 5,906685317 Median: 9347,04 +/- 1,142811031 Min: 9269,8 +/- 32,43671202 Max: 9359,16 +/- 0,691534358 Mbps 25% Percentile: 9338,42 +/- 8,388554416 75% Percentile: 9350,53 +/- 0,400027344 5% Percentile: 9326,666 +/- 16,14620079 95% Percentile: 9354,598 +/- 0,299412338 **Throughput** **Compute 103 to 104** Mean: 9347,535957 +/- 1,250071848 Standard deviation: 15,52114231 +/- 5,722436432 Median: 9350,08 +/- 0,384646319 Min: 9216,4 +/- 64,18669224 Mbps Max: 9362,4 +/- 0,445853047 25% Percentile: 9345,69 +/- 1,03807122 75% Percentile: 9353,56 +/- 0,271855478			

	5% Percentile:	9333,21	+/- 3,884531795	
	95% Percentile:	9358,926	+/- 0,396914757	
Complementary measurement results	n/a			

Test Case ID	TC_THR_Tcp			
General description of the test	Throughput between three pairs of VMs running on different compute nodes of the Berlin Platform test bed infrastructure. All measurements are run in parallel.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	Agent Probes deployed on different end points in the E2E path, VMWare Virtualization Platform, iPerf3 for throughput traffic generation and measurement			
Metric(s) under study	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			

Primary measurement results	**Throughput**	**Compute 101 to 102**		
	Mean:	9347,59	+/- 0,082594377	
	Standard deviation:	5,13571942	+/- 0,279378603	
	Median:	9348	+/- 0	
	Min:	9318,04	+/- 5,789324373	
	Max:	9359,44	+/- 0,642576623	Mbps
	25% Percentile:	9344,78	+/- 0,166716508	
	75% Percentile:	9350,9	+/- 0,115375437	
	5% Percentile:	9340,06	+/- 0,279897418	
	95% Percentile:	9354,902	+/- 0,235582969	
	Throughput	**Compute 103 to 104**		
	Mean:	9347,783404	+/- 0,830581781	
	Standard deviation:	14,30705208	+/- 3,332359344	
	Median:	9350,3	+/- 0,230750874	
	Min:	9232,56	+/- 37,12495541	
	Max:	9362,24	+/- 0,574935611	Mbps
	25% Percentile:	9346,12	+/- 0,351368249	
	75% Percentile:	9353,69	+/- 0,191398079	
	5% Percentile:	9334,07	+/- 3,37481601	
	95% Percentile:	9358,396	+/- 0,489892312	
	Throughput	**Compute 105 to 106**		
	Mean:	9351,015532	+/- 0,497715842	Mbps

	Standard deviation:	7,812115308	+/- 1,693604553
	Median:	9351,88	+/- 0,181497946
	Min:	9300,28	+/- 13,7025457
	Max:	9363,92	+/- 0,945196442
	25% Percentile:	9348,54	+/- 0,203532212
	75% Percentile:	9355,06	+/- 0,215394951
	5% Percentile:	9341,302	+/- 2,707442922
	95% Percentile:	9360,196	+/- 0,217683417
Complementary measurement results	n/a		

9.5.1.6. Intra-Compute RTT

Test Case ID	TC_RTT_e2e			
General description of the test	The test assesses the average, minimum, and maximum RTT between two communication endpoints (VNF) deployed on compute nodes in one data center.			
Purpose	The test acts as a calibration test to primarily assess the performance of the data center interconnection and the performance of the virtualization layer (SDN).			
Executed by	Partner:	FOKUS	Date:	2021 Q4
Involved Partner(s)	FOKUS			
Scenario	The communication endpoints of the service (VNFs) are within the same compute and storage unit on a single compute node.			
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved *(e.g. HW components, SW components)*	• VMs placed on different compute units within the same compute & storage system • Compute and storage system (NetApp HCI)			
Metric(s) under study (*Refer to those in Section 4*)	*RTT*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints			
Primary measurement results	**Round Trip time (ms)**	**102 to 101 32 Byte**		
	Mean:	0,269720513	+/- 0,003725919	
	Standard deviation:	0,059114899	+/- 0,016413734	
	Median:	0,26856	+/- 0,003409246	
	Min:	0,14188	+/- 0,00717861	
	Max:	0,70816	+/- 0,265214892	
	25% Percentile:	0,24049	+/- 0,004110722	
	75% Percentile:	0,29633	+/- 0,003808406	
	5% Percentile:	0,194676	+/- 0,005053006	
	95% Percentile:	0,337204	+/- 0,004527259	
	Round Trip time (ms)	**102 to 101 56 Byte**		

	Mean:	0,205993442	+/- 0,017567929
	Standard deviation:	0,056037213	+/- 0,034569296
	Median:	0,202041667	+/- 0,017515662
	Min:	0,117083333	+/- 0,010801521
	Max:	0,747666667	+/- 0,705884211
	25% Percentile:	0,17878125	+/- 0,016643074
	75% Percentile:	0,22725	+/- 0,018368729
	5% Percentile:	0,150410417	+/- 0,01434477
	95% Percentile:	0,267547917	+/- 0,019279193
	Round Trip time (ms)	102 to 101 1400 Byte	
	Mean:	0,267187427	+/- 0,005921057
	Min:	0,054086181	+/- 0,011573013
	Max:	0,26582	+/- 0,006003036
	25% Percentile:	0,15104	+/- 0,006891155
	75% Percentile:	0,62464	+/- 0,16769587
	5% Percentile:	0,23919	+/- 0,005249881
	95% Percentile:	0,29217	+/- 0,007087766
	Standard deviation:	0,196524	+/- 0,004574144
	Median:	0,330678	+/- 0,008606754
Complementary measurement results	n/a		

9.5.1.7. Inter-Compute Reliability RTT

Test Case ID	TC_REL_rtt_e2e			
General description of the test	The reliability (RTT) of the connection between different compute nodes of the Berlin Platform test bed infrastructure.			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	• VMs placed on different compute units within the same compute & storage system • Compute and storage system (NetApp HCI)			

Metric(s) under study	*Reliability (RTT)*							
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints							
Primary measurement results	RTT							
	Link	(-0.001, 1.0]ms	(1.0, 2.0]ms	(2.0, 3.0]ms	(3.0, 10.0]ms	(10.0, 20.0]ms	(20.0, 50.0]ms	(50.0, 1000.0] ms
	101 to 102	99,945488%	0,034365%	0,003555%	0,008295%	0,003555%	0,004740%	0%
	102 to 104	99,961486%	0,021923%	0,006517%	0,004147%	0,005332%	0,000592%	0%
	103 to 105	99,962083%	0,028437%	0,002369%	0,004739%	0,002369%	0%	0%
	106 to 102	99,966819%	0,022515%	0,003555%	0,00237%	0,00237%	0,00237%	0%
Complementary measurement results	n/a							

9.5.2. Evaluation of 5G Equipment – Results

9.5.2.1. 5G RTT Reliability

Test Case ID	TC_REL_rtt_e2e			
General description of the test	The RTT reliability of the connection between a 5G CPE in the server room 3009 at FOKUS and compute nodes of the Berlin Platform test bed infrastructure			
Purpose	The test assesses the performance of the end to end connection and the performance of the virtualization layer.			
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario				
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved	• VMs placed on different compute units within the same compute & storage system • Compute and storage system (NetApp HCI)			
Metric(s) under study	*Reliability (RTT)*			

Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication endpoints							
Primary measurement results	RTT							
	Link	(-0.001, 1.0]ms	(1.0, 2.0]ms	(2.0, 3.0]ms	(3.0, 10.0]ms	(10.0, 20.0]ms	(20.0, 50.0]ms	(50.0, 1000.0]ms
	R3009 to 101	0%	0%	0%	7,964601 %	49,754178 %	42,281219 %	0%
Complementary measurement results	n/a							

9.5.3. UC#1: 360 degree camera –Results

9.5.3.1. 60GHz backbone link

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses the throughput (TCP) capacity between two communication endpoints connected via a 60GHz backbone link.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario	60GHz backbone baseline			
Slicing configuration	-			
Components involved	• Host A and B • 60GHz backbone link			
Metric(s) under study	*Throughput*			
Additional tools involved	-			
Primary measurement results	**Throughput**	**B>A [380]**		
	Mean: Standard deviation: Median: Min: Max: 25% Percentile: 75% Percentile: 5% Percentile: 95% Percentile:	896008.3355072463 3813.0057188127926 897622.755 867715.21 899419.26 893868.97 898268.4874999999 888658.0275 898867.8975	Mbps	
	Throughput	**A>B [381]**		
	Mean: Standard deviation:	916107.523537906 4909.1300382697655	Mbps	

	Median:	917816.74	
	Min:	884896.7	
	Max:	920693.79	
	25% Percentile:	914065.57	
	75% Percentile:	918583.05	
	5% Percentile:	910941.758	
	95% Percentile:	919558.7179999999	
	Throughput	**B>A [382.1] parallel**	
	Mean:	764254.6525461251	
	Standard deviation:	36587.09014885816	
	Median:	765349.17	
	Min:	603519.45	
	Max:	827398.91	Mbps
	25% Percentile:	746511.74	
	75% Percentile:	785541.395	
	5% Percentile:	709737.075	
	95% Percentile:	817998.32	
	Throughput	**A>B [382.2] parallel**	
	Mean:	813448.1938745385	
	Standard deviation:	10601.821480049903	
	Median:	812841.56	
	Min:	780766.32	
	Max:	855148.19	Mbps
	25% Percentile:	807223.275	
	75% Percentile:	819225.995	
	5% Percentile:	797471.435	
	95% Percentile:	828542.255	
Complementary measurement results	n/a		

9.5.3.2. Facility-to-facility tunnel connection baseline

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints located in different facilities/sites, connected via an inter-facility tunnel connection.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario	Inter facility tunnel (DFN GEANT) baseline			
Slicing configuration	-			

Components involved	• Host A and F • DFN GEANT tunnel		
Metric(s) under study	*Throughput*		
Additional tools involved	-		
Primary measurement results	**Throughput**	**A>F [443]**	
	Mean:	617341.2343636364	Mbps
	Standard deviation:	87195.91477229791	
	Median:	629348.32	
	Min:	264866.12	
	Max:	783846.41	
	25% Percentile:	585050.76	
	75% Percentile:	666054.745	
	5% Percentile:	509492.90100000007	
	95% Percentile:	724061.793	
	Throughput	**F>A [444]**	
	Mean:	716526.9450000001	Mbps
	Standard deviation:	134273.9497265382	
	Median:	749287.155	
	Min:	380385.34	
	Max:	899507.51	
	25% Percentile:	620447.6275	
	75% Percentile:	819592.6825	
	5% Percentile:	442674.45200000005	
	95% Percentile:	885240.1799999999	
	Throughput	**A>F [442.1] parallel**	
	Mean:	558415.429	Mbps
	Standard deviation:	85490.86126142016	
	Median:	561406.085	
	Min:	208662.4	
	Max:	764317.38	
	25% Percentile:	508467.5425	
	75% Percentile:	597360.755	
	5% Percentile:	470820.184	
	95% Percentile:	696923.6660000001	
	Throughput	**F>A [442.2] parallel**	
	Mean:	631991.9231666666	Mbps
	Standard deviation:	208538.66459788414	
	Median:	679381.25	
	Min:	163686.51	
	Max:	892611.92	
	25% Percentile:	545725.5449999999	
	75% Percentile:	794467.515	

	5% Percentile:	223356.5255
	95% Percentile:	870553.244
Complementary measurement results	n/a	

9.5.3.3. Facility-to-facility tunnel connection with 60GHz backbone link

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints located in different facilities/sites. The two endpoints are connected via an inter facility tunnel connection in extension with the 60GHz backbone link at IHP.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario	Inter facility tunnel (DFN GEANT) + backbone link			
Slicing configuration	-			
Components involved	• Host B and F • DFN GEANT tunnel • 60GHz backbone link			
Metric(s) under study	*Throughput*			
Additional tools involved	-			

Primary measurement results	**Throughput**	**B>F [457]**	
	Mean:	500688.24631578947	Mbps
	Standard deviation:	57134.32528886534	
	Median:	510676.475	
	Min:	357231.3	
	Max:	612188.13	
	25% Percentile:	477252.26749999996	
	75% Percentile:	539308.5075000001	
	5% Percentile:	399604.82149999996	
	95% Percentile:	569636.6214999999	
	Throughput	**F>B [458]**	
	Mean:	494627.5594285715	Mbps
	Standard deviation:	149093.28468007792	
	Median:	532623.27	
	Min:	86827.44	
	Max:	695973.05	
	25% Percentile:	478076.25	
	75% Percentile:	583911.51	
	5% Percentile:	161460.47	
	95% Percentile:	661647.756	

	Throughput	**B>F [456.1] parallel**	
	Mean:	445662.987631579	Mbps
	Standard deviation:	66700.23995979429	
	Median:	449365.79500000004	
	Min:	242841.55	
	Max:	539856.15	
	25% Percentile:	422442.235	
	75% Percentile:	486954.7425	
	5% Percentile:	335901.462	
	95% Percentile:	538360.278	
	Throughput	**F>B [456.2] parallel**	
	Mean:	441672.0239473684	Mbps
	Standard deviation:	142238.58547208275	
	Median:	459687.79	
	Min:	127431.23	
	Max:	658354.99	
	25% Percentile:	347293.99250000005	
	75% Percentile:	558718.3225	
	5% Percentile:	180936.85100000002	
	95% Percentile:	597060.8994999999	
Complementary measurement results	n/a		

9.5.3.4. WiFi-to-core-infrastructure baseline

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints connected via WiFi at the IHP facility.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario	IHP WiFi baseline			
Slicing configuration	-			
Components involved	• Host A and C • WiFi AP			
Metric(s) under study	*Throughput*			
Additional tools involved	-			
Primary measurement results	**Throughput**	**C>A [22]**		
	Mean:	9083.96	Mbps	
	Standard deviation:	2607.331896924791		
	Median:	8089.08		
	Min:	6677.91		

	Max:	15330.18	
	25% Percentile:	6828.64	
	75% Percentile:	10803.97	
	5% Percentile:	6718.984	
	95% Percentile:	13304.746	
	Throughput	**A>C [23]**	
	Mean:	51239.520999999986	
	Standard deviation:	4448.570699938817	
	Median:	52099.270000000004	
	Min:	35828.0	
	Max:	60769.01	Mbps
	25% Percentile:	51516.5675	
	75% Percentile:	52585.8625	
	5% Percentile:	42584.912000000004	
	95% Percentile:	53426.8845	
Complementary measurement results	n/a		

9.5.3.5. WiFi-to-core-infrastructure with 60Ghz backbone link

Test Case ID	TC_THR_Tcp		
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints connected via WiFi at the IHP facility and is extended with the 60GHz backbone link.		
Purpose			
Executed by	Partner: FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS		
Scenario	IHP WiFi with backbone extension		
Slicing configuration	-		
Components involved	• Host B and C • WiFi AP • 60GHz backbone link		
Metric(s) under study	*Throughput*		
Additional tools involved	-		
Primary measurement results	**Throughput**	**C>B [26]**	
	Mean:	6814.644814814815	
	Standard deviation:	169.70263033114878	
	Median:	6765.29	
	Min:	6719.42	Mbps
	Max:	7597.57	
	25% Percentile:	6747.8150000000005	
	75% Percentile:	6790.415	

	5% Percentile:	6724.222	
	95% Percentile:	6985.269	
	Throughput	**B>C [27]**	
	Mean:	44082.52999999999	Mbps
	Standard deviation:	8499.665496493846	
	Median:	44755.71	
	Min:	32182.53	
	Max:	53464.57	
	25% Percentile:	35447.9	
	75% Percentile:	52586.4125	
	5% Percentile:	32936.7125	
	95% Percentile:	53372.8175	
Complementary measurement results	n/a		

9.5.3.6. Local-WiFi-to-remote-facility connectivity

Test Case ID	TC_THR_Tcp		
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints located in different facilities/sites. The two endpoints are connected via an inter facility tunnel connection, where the endpoint at IHP uses WiFi connectivity.		
Purpose			
Executed by	Partner: FOKUS	Date: Q4 2021	
Involved Partner(s)	FOKUS		
Scenario	IHP WiFi with inter-facility tunnel		
Slicing configuration	-		
Components involved	• Host F and C • WiFi AP • DFN GEANT tunnel		
Metric(s) under study	*Throughput*		
Additional tools involved	-		
Primary measurement results	**Throughput**	**C>F [465]**	
	Mean:	7297.4524	Mbps
	Standard deviation:	1202.2725594968058	
	Median:	6705.11	
	Min:	6304.37	
	Max:	11309.09	
	25% Percentile:	6481.75	
	75% Percentile:	7899.58	
	5% Percentile:	6376.898	
	95% Percentile:	9280.685999999998	

	Throughput	**F>C [466]**	
	Mean:	38764.992399999996	
	Standard deviation:	7988.202654816601	
	Median:	37536.25	
	Min:	17614.01	
	Max:	53066.28	Mbps
	25% Percentile:	35105.37	
	75% Percentile:	46106.02	
	5% Percentile:	29755.156000000003	
	95% Percentile:	50822.742	
	Throughput	**C>F [466.1] parallel**	
	Mean:	3573.2252000000003	
	Standard deviation:	2457.868193669262	
	Median:	3698.74	
	Min:	214.08	
	Max:	8711.33	Mbps
	25% Percentile:	1272.23	
	75% Percentile:	5254.08	
	5% Percentile:	321.03000000000003	
	95% Percentile:	7459.891999999999	
	Throughput	**F>C [464.2] parallel**	
	Mean:	34087.5072	
	Standard deviation:	6515.766757534834	
	Median:	34035.86	
	Min:	25456.8	
	Max:	45858.74	Mbps
	25% Percentile:	28239.48	
	75% Percentile:	38976.25	
	5% Percentile:	25773.286	
	95% Percentile:	44744.836	
Complementary measurement results	n/a		

9.5.3.7. 5G RAN and Core Network

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints, where one is connected via 5G at the IHP facility and the other located inside the network infrastructure.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			

Scenario	5G equipment baseline		
Slicing configuration	-		
Components involved	• Host E and A • 5G equipment • 5G core network deployment		
Metric(s) under study	*Throughput*		
Additional tools involved	-		
Primary measurement results	**Throughput**	**E>A [40]**	
	Mean:	63019.445999999996	Mbps
	Standard deviation:	35.36861402901304	
	Median:	63025.57	
	Min:	62859.55	
	Max:	63047.41	
	25% Percentile:	63019.01	
	75% Percentile:	63034.31	
	5% Percentile:	63002.414000000004	
	95% Percentile:	63040.86	
	Throughput	**A>E [41]**	
	Mean:	568602.7917857141	Mbps
	Standard deviation:	19970.413117229382	
	Median:	568672.01	
	Min:	518755.02	
	Max:	606383.02	
	25% Percentile:	552915.175	
	75% Percentile:	582431.455	
	5% Percentile:	545156.3545	
	95% Percentile:	600011.2914999999	
Complementary measurement results	n/a		

9.5.3.8. 5G Network with 60GHz backbone link

Test Case ID	TC_THR_Tcp			
General description of the test	The test assesses throughput (TCP) capacity between two communication endpoints, where one is connected via 5G at the IHP facility and the other is connected located behind a backbone extension.			
Purpose				
Executed by	Partner:	FOKUS	Date:	Q4 2021
Involved Partner(s)	FOKUS			
Scenario	5G with 60Ghz backbone extension			
Slicing configuration	-			
Components involved	• Host E and B			

	• 5G equipment • 5G core network deployment • 60GHz backbone link		
Metric(s) under study	*Throughput*		
Additional tools involved	-		
Primary measurement results	**Throughput**	**E>B [28]**	
	Mean:	45839.522857142874	
	Standard deviation:	21521.238442791917	
	Median:	55188.11	
	Min:	2422.57	
	Max:	63060.52	Mbps
	25% Percentile:	37666.22	
	75% Percentile:	63014.1025	
	5% Percentile:	2455.34	
	95% Percentile:	63046.429000000004	
	Throughput	**B>E [29]**	
	Mean:	463039.0910465116	
	Standard deviation:	117563.34315389463	
	Median:	502856.195	
	Min:	168647.98	
	Max:	576319.42	Mbps
	25% Percentile:	466250.2925	
	75% Percentile:	532430.47	
	5% Percentile:	186410.70249999998	
	95% Percentile:	556125.25	
Complementary measurement results	n/a		

9.5.3.9. 360° video experiments at Fraunhofer FOKUS

Test Case ID	TC_360VideoStreamingQoE_Scalability			
General description of the test	The test assesses the 360° Video Streaming QoE performance in the Fraunhofer FOKUS testbed.			
Purpose	The test involves a controlled experiment to assesses the scalability of the 5G system to support multiple 360° video clients with good QoE.			
Executed by	Partner:	KAU	Date:	26-27.11.2021
Involved Partner(s)	KAU, SRL, FhG			
Scenario	Controlled experiments in the Fraunhofer FOKUS testbed with the client connected via a CPE to the actual 5G network.			
Slicing configuration	N/A			

Components involved	• The MONROE measurement probe 360-dash that emulates a 360º video client • A DASH HTTP video server, running nginx.
Metric(s) under study	Application specific metrics (defined in TC_360VideoStreamingQoE_Scalability)
Additional tools involved	n/a
Primary measurement results	**Test with 1 client:** **Representation Rate:** Mean: 4.24 +/- 0.01 Mbps Median: 4.28 +/- 0.01 Mbps 5% Percentile: 4.28 Mbps 25% Percentile: 4.28 Mbps 50% Percentile: 4.28 Mbps 75% Percentile: 4.28 Mbps 95% Percentile: 4.28 Mbps **Frequency of rebuffering events:** Mean: 0 +/- 0 events/second Median: 0 +/- 0 events/second 5% Percentile: 0 events/second 25% Percentile: 0 events/second 50% Percentile: 0 events/second 75% Percentile: 0 events/second 95% Percentile: 0 events/second **Test with 10 clients:** **Representation Rate:** Mean: 4.24 +/- 0.00 Mbps Median: 4.28 +/- 0.00 Mbps 5% Percentile: 4.28 Mbps 25% Percentile: 4.28 Mbps 50% Percentile: 4.28 Mbps 75% Percentile: 4.28 Mbps 95% Percentile: 4.28 Mbps **Frequency of rebuffering events:** Mean: 0 +/- 0 events/second Median: 0 +/- 0 events/second 5% Percentile: 0 events/second 25% Percentile: 0 events/second 50% Percentile: 0 events/second 75% Percentile: 0 events/second 95% Percentile: 0 events/second **Test with 50 clients:** **Representation Rate:** Mean: 4.24 +/- 0.00 Mbps Median: 4.28 +/- 0.00 Mbps 5% Percentile: 4.28 Mbps 25% Percentile: 4.28 Mbps

	50% Percentile: 4.28 Mbps 75% Percentile: 4.28 Mbps 95% Percentile: 4.28 Mbps **Frequency of rebuffering events:** Mean: 0 +/- 0 events/second Median: 0 +/- 0 events/second 5% Percentile: 0 events/second 25% Percentile: 0 events/second 50% Percentile: 0 events/second 75% Percentile: 0 events/second 95% Percentile: 0 events/second
Complementary measurement results	**Test with 1 clients:** **Delivery Rate:** Mean: 81.46 +/- 0.49 Mbps **Test with 10 clients:** **Delivery Rate:** Mean: 64.50 +/- 0.09 Mbps **Test with 50 clients:** **Delivery Rate:** Mean: 52.75 +/- 0.04 Mbps

10. Annex 2: Data Anonymisation – Berlin Platform

Due to the increasing popularity of the internet over the last two decades an increasing amount of data of all types has been collected [A1]. Increasingly strict privacy laws in Europe aim at protecting people that have (knowingly or not) provided their data [A2]. It has been shown in the past that privacy protection is difficult. One famous example is the attack on the Netflix Price dataset in 2008 [A3]. Location data, as shown by the release of the NYC Taxi dataset and subsequent attacks on it [A4] and [A5], is particularly sensitive as it can contain information about home address, workplace, religion or sexual orientation. Using this information people which have provided their data can be identified.

Most datasets are not available to the public due to business or privacy reasons. But releasing data to the public as *open data* has it benefits. It gives researches better insights, can facilitate academic collaboration and improve exchange of knowledge [A6]. Public datasets like the collection of hand written numbers called MNIST [A7] are used as standard to compare approaches and ideas. For this reason finding way how to release datasets without breaching the privacy of those who donated their data has become a focus of privacy research.

An approach to release important data to the public while providing privacy to data sources is the generation of synthetic data. Hereby information and characteristics are extracted from a real world dataset to generate a synthetic dataset using anonymisation algorithms such as Differential Privacy (DP) [A8]. Differential privacy, as explained in Deliverable D3.5., eliminates the impact of single data sources and ensures in the context of synthetic dataset that no personal information is leaked by the released dataset.

In this chapter we take a look at how location data collected during the measurement campaign at the IHP in 2021 can be anonymized by generating synthetic data.

10.1. Criteria and Metrics for Synthetic Data

Synthetic data aims at providing as many properties of the original dataset as possible without leaking information about the data sources. When evaluating the utility of an algorithm for synthetic data generation, there are various metrics that can be used. In the following we will give an overview of metrics for analyzing synthetic location data which are relevant in our context.

It is important that synthetic location data follows the same distribution as the original dataset so hotpots where many people pass or come together can be correctly identified. To compare two distributions, the *Kullback-Leibler divergence* (KLD) comes to mind. It is defined as:

$$KLD(P,Q) = \sum P\,(x) log\left(\frac{P(x)}{Q(x)}\right).$$

The KLD is not a metric as it is assymetric. The *Jensen-Shannon divergence* (JSD) on the other hand is a metric based on the KLD. Unlike the KLD it is symmetric and always finite. For two distributions P and Q the JSD is defined as:

$$JSD(P,Q) = \frac{1}{2}KLD(P,M) + \frac{1}{2}KLD(Q,M) \text{with } M = \frac{1}{2}(P+Q)$$

One problem of the JSD is the fact that it performs poorly if domains do not overlap. If they do not overlap JSD will always be maximal independent of the fact if the distance between the two distributions decreases or increases [A9].

The *earth mover distance* (EMD) metric, also called *Kantorovich* or *Wasserstein metric,* measures the distance of two distributions. Picturing an area with piles of earth, the earth mover distance is the minimal cost required when piling up dirt to create a new distribution (of earth) given an existing distribution (of earth). To calculate the EMD for two distributions P and Q both have to be discretised. Using the clusters created by discretization, a flow graph between both discretised distributions is calculated. The flow is then minimized using for example the Hungary algorithm. The EMD is defined in [A10] as

$$EMD(P,Q) = min_{F=f_{ij}\in\Gamma(P,Q)} \frac{WORK(F,P,Q)}{\sum F}$$

where

- F is the minimal flow between P and Q for all flows
- d_{ij} is a distance metric, for example the Euclidean metric, between i and j
- $WORK(F,P,Q) = \sum\sum f_{ij}\, d_{ij}$

Kolouri et al. [A9] provide an example where JSD is compared to the EMD. Unlike the JSD, the EMD increases for non-overlapping distributions if their distance increases. The EMD additionally allows partial matching. This means for unequal distributions it matches all the weight of the lighter distribution to the heavier distribution but not the other way around.

Both JSD and EMD can be use to evaluate the distribution of two dimensional histograms to compare synthetic data with ground data to find hot spots.

Similar to evaluations looking at point distribution is the *query answering* metric defined by Gursoy et al. in [A11]. Here, the numbers of trajectories passing a certain area with a center and radius are counted. The relative error is used to compare query answering of original and synthetic datasets. Gursoy et al. generate a set of 500 random queries, for which the relative error is averaged. One problem of this metric is the fact that it does not account for set size. A synthetic set might be considerably larger than the original dataset. In that case, the same number of trajectories passing trough a certain region will not have the same meaning in the two datasets. We therefore divide the number of trajectories by the total number of trajectories of this dataset. We also change the bound defined by the Gursoy et al. to represent the number of trajectories instead of the number of data points in the dataset. This give us:

$$QA = \frac{\left|\frac{Q(D)}{T_D} - \frac{Q(S_D)}{T_{SD}}\right|}{max\frac{Q(D)}{T_D}, 0.01 \cdot T_D}$$

where

- T_X is the number of trajectories in dataset X
- Q(X) a counting query over an area for dataset X

For the utility of synthetic data the quality of sequential information is also an important part. A useful measure are *frequent patterns* and the preservation of these [A11, 12]. The trajectories are first discretised by applying a uniform grid. Then using the *apriori* algorithm the top k frequent patterns are calculated. The frequent patterns of the original dataset can be compared to those of the synthetic dataset. To compare frequent patterns of two datasets Gursoy et al. [A11] use the average relative error to determine the distance between two sets of frequent patterns. While using the F_1score (as done for example by He et al. in [A12]) is also an option, this will ignore patters which occur in the set of most frequent patterns but at a different rank.

Another relevant measure is *diameter error* as defined by He et al. in [A12], where the distribution of traveled distance from synthetic trajectories are compared to those of the real dataset using JSD.

Trip distribution or *trip error* can also be a meaningful measure when evaluating utility. We use the trip distribution definition of Gursoy et al. [A11] where a start and end grid cell are used to calculate the distribution of trips.

10.2. Related Work

In this section we will discuss works focusing on the generation of synthetic location data using differential privacy.

10.2.1. Machine Learning based Approaches

During the recent years there has been much work on training on spacio-temporal data [A13] and generating synthetic location data.

It has been shown during the last years that machine leanring approaches without formal privacy protection, such as TrailGAN [A14] or LSTM-TrajGAN [A15], are not sufficent to protect data. Yeom et al. [A16] have demonstrated in 2018 that for machine learning models without formal privacy protection over-fitting is one but not the only reason for privacy leaks. For this purpose they conducted membership and attribute interference attacks against various models. In 2020, Chen et al. [A17] presented a broad analysis of attacks on generative machine learning models (GANs).

There are multiple approaches how to make neural networks or their output privacy preserving and provide formal guarantees. To eliminate a single users' contribution differentially private noise can be either added to the whole network after finishing training. Another option is adding differentially private noise to the gradient during back propagation [A18]. This allows to make non private approaches like TrailGAN, LSTM-TrajGAN and SeqGAN [A19] privacy preserving.

There are also machine leaning approaches for trajectory generation designed with formal privacy guarantees in mind. For example, Acs et al. presented in 2018 an approach on how to create differentially private models to be published instead of datasets [A20]. The data is split into k clusters which each is passed to a separate GAN. Differential privacy is used during clustering and for perturbing the gradient descent. Among others the authors tested their model on a dataset of transit records.

RNN-DP proposed by Chen et al. in 2020 [A21] focuses on the release of differentially private synthetic trajectory datasets by using a recurrent neural network (RNN). This type of neural network is good at learning sequences due to it's chain structure and additional inputs from earlier time steps.

While advances towards the generation of synthetic data using machine learning based approaches have been made, issues remain. Rahman et al. [A22] have shown that to protect against membership inference attacks privacy preserving machine learning models have to trade utility against privacy. The authors suggest that to provide acceptable utility, some vulnerabilities have to be tolerated. This finding is mirrored by Chen et al. [A17] and Jayaraman and Evans [A23]. The latter also looked at different mechanisms of differential privacy and conducted attribute as well as membership inference attacks. Farokhi and Kaafar [A24] give an upper bound for the leakage of membership information when Gaussian noise is added based on the used epsilon value. Stadler et al. [A25] conducted experiments to determine the privacy guarantees for various machine learning algorithms for generating synthetic tabular data and find that it does not retain utility and does not prevent inference attacks.

Due to the privacy-utility trade off and the findings of Stadler et al. we decide against using machine learning based approaches like TrailGAN [A14], LSTM-TrajGAN [A15], Ace et al. [A20], RNN-DP [A21], Bousquet et al. [A26] and Frigerio et al. [A27] for the generation of privacy preserving differential privacy for this evaluation.

10.2.2. Hierarchical Indexes and Markov Models

Generating privacy preserving synthetic location data can be challenging due to hidden correlations between locations and many data points originating from a single user. Over the last years various algorithms have been developed specifically for location data. In the following we will discuss a selection for relevant solutions.

In 2012 Chen et al. [A28] presented the n-gram algorithm at the CCS. This solution applies the Markov assumption that one state is only defined a a fixed number of previous states on locating traces. A uniform grid is applied to the area of interest and locations are generalized to positions on the grid. Traces are split into tuples of sequences of various lengths. One specific location from a certain trace can be used in multiple tuples with the same length but also in tuples which are shorter or longer. An prefix tree is built from all tuples which represents which positions on the grid are likely followed by which other ones. Each node in the tree is associated with a number of n-grams which contain the sequence of grid positions from the root to this node. To incorporate privacy, these counts are perturbed using differential privacy. To create a synthetic dataset, data can be drawn from the prefix tree by traversing paths from root to leaf. Since the noise that has to be added to the prefix tree scales with the number of locations, only smaller areas can be covered while still maintaining good utility.

Bindschaedler and Shokri presented an approach called SGLT to generating privacy preserving data at S&P 2016 [A29]. First they compute semantic similarity between all locations of a dataset and cluster these to create a semantic location graph based on the way people interact with these locations. To generate synthetic traces first a few seed traces are sampled from the original dataset and transformed to semantic traces. Then each location is replaced with a similar location. Randomness is included in various steps of the algorithm but there is not formal privacy protection. Instead of differential privacy, the goal is statistical dissimilarity and plausible deniability. For this reason we ignore this approach in our evaluation.

Another approach is the DPT algorithm designed by He et al. [A12] in 2015. The core idea consists of using various grids of different granularity to simplify location traces and derive transition probabilities between generalized locations. For each grid size the approach of Chen et al. is used to build a prefix tree of grid position sequences. Additionally, transitions to other grids are also modeled which represent changes in velocity or transport mode. To apply Laplace noise to ensure differential privacy, a model selection mechanism is used to set the relevant parameters. For sampling trajectories for the synthetic dataset again the Markov assumption is used and only the previous k locations are taken into account when selecting the next one.

A rather different approach is taken by the DP-Star algorithm proposed by Gursoy et al. in 2018 [A11]. Here, trajectories are simplified into a sequence of locations that are representative. Then an adaptive grid which is more fine grained in areas with high density is used to discretise the area. Transitions between positions on the grid are represented as a Markov model. To derive transition probabilities trajectories are split into overlapping sets of fixed length. Particular attention is additionally paid to preserve start and end points as well as route length. To ensure privacy, the Markov model as well as the counts for start-end pairs and route length are perturbed. The authors of DP-Star compare their solution with DPT and n-grams on the GeoLife [A30], the Taxi [A31] and the an artificial Brinkhoff [A32] dataset. They find that their solution outperforms the other solutions most of the time. Only for frequent pattern metrics on the Taxi dataset the n-gram algorithm performs better than DP-Star.

Gursoy et al. improved DPStar with AdaTrace presented in 2018 at the CCS [A33]. Hereby, special attention is payed to location which are particularly sensitive such as hospitals or schools. Also attack vectors are considered where the advarsary can record sub-trajectories. The analysis compares AdaTrace with ngram, DPT and SGLT. It outperformes other approaches on realistic datasets most of the time.

The TGM algorithm proposed by Ghane et al. in 2019 [A34] pays special attention to the preservation of locations where people stay for extended periods of time, also called stay points. At first, trajectories are aggregated both in the spacial as well as in temporal dimension and mapped to a uniform grid. Then, trajectories are encoded into a k-order Markov Chain where each node represents a cell of the grid which allows transitions to neighboring cells or itself. Each node also stores the earlier locations, called prefix, of trajectories passing through and their stay time. To generate trajectories, unlike other approaches, not the model itself is perturbed using differential privacy but the process of selecting the next location when traversing the model is made private. We send a message to the authors asking for the code of TGM but have not heard back as of the 17.12.2021.

For our evaluation we selected DPT and AdaTrace. The ngram algorithm was ignored due to evaluation results of AdaTrace. Approaches such as TGM [A7], SafePath [A35] and [A36] were ignored in as no source code was available

10.3. Experiments at UC #1: 360º camera

Due to the ongoing Covid-19 pandemic, experiments at the Festival of Lights 2020 and 2021 were cancelled. Instead, experiments were conducted during end of November 2021 at IHP in Frankfurt (Oder).

10.3.1. Data Collection and Data Quality

Data was collected using a web server provided by SRL showing a video stream available both via the local Wi-Fi at IHP and the local 5G campus network. Data was collected on Wednesday 24th November 2021 at the IHP main building in Frankfurt (Oder). Local IHP personnel as well as staff from 5Genesis partners present for the experiments took part in the data collection. Due to rising Covid-19 infections in the state of Brandenburg, the IHP had asked their staff to stay at home from the 24th November 2021 onwards. Therefore the number of actual data collectors greatly diminished compared to the initial expectations.

The web tool provided by SRL collected, among other things, a session ID, longitude, latitude, altitude, location accuracy and time. If geo-location was enabled on the end device, location was determined using GPS. Otherwise an estimate was derived using the IP address. As a result, location accuracy of the data varies greatly.

Figure 10-1 All data points collected in the vicinity to the IHP main building. Outlayers have been removed

A total of 228 traces were collected containing 10,566 data points in total (see Figure 10-1). Mean number of data points per trajectory is 835.04 with a standard deviation of 995.13. The corresponding median is 387 data points, meaning there are many short trajectories. The distance traveled by trajectories of the dataset is 1,330 m with a standard deviation of 8524 m and median 0.2 m.

There are large differences in location accuracy with minimum 7.22 m, maximum 24,450 m, standard deviation 2238.5 m and median 20 m. This means while the range is large, most data points have a comparably low deviation from the measurement. Considering the fact that GPS allows for an accuracy of 4.9 m for GPS-enabled smart phones the measured accuracy comes at a surprise [A37]. This might be the result of signal reflection and indoor measurements. It is also suprising that data points collected outdoors show an error larger than the median. As these were mostly collected while walking, this might be the reason. When filtering data points with accuracy lower than 20 m, it has to be noted that not all locations that seem to be located outside the IHP building in Figure 10-1 were actually recorded there. Figure 10-2 shows for trajectories the accuracy of the corresponding data points. It can be seen that most data points that were collected inside but show up on the map as outside the building also have a large overlap with the inside.

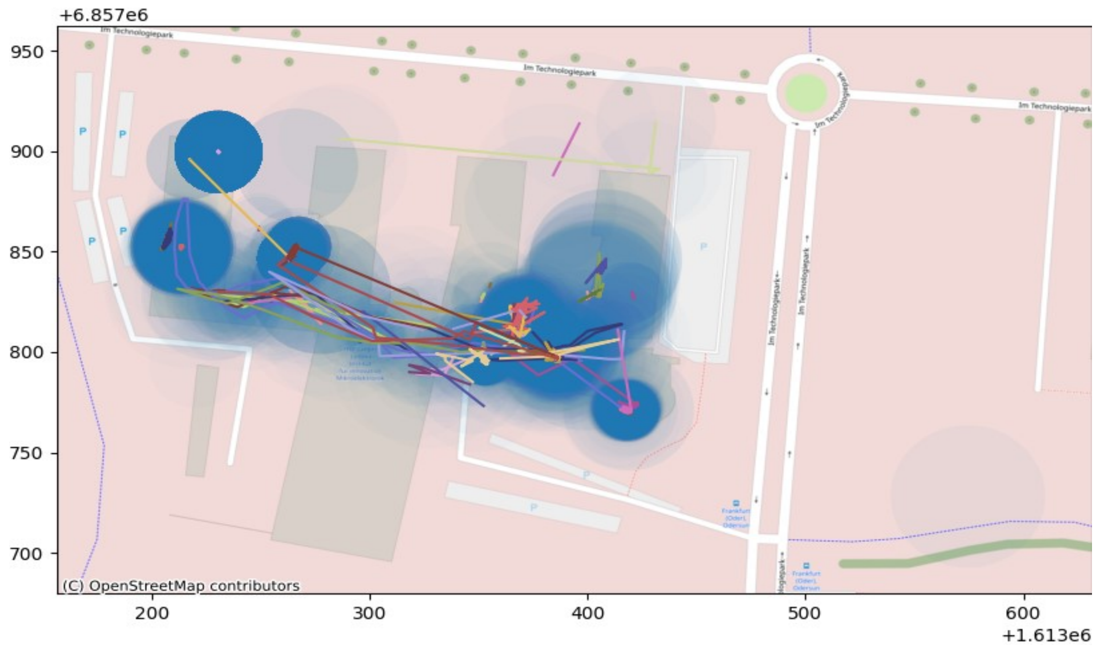

Figure 10-2 Trajectories measured during data collection at the IHP after removing outlayers and data points with an error of more than 40m. Blue circles represent the accuracy of each data point.

Before the dataset can be used for further analysis and experiments it has to be filtered. We removed all out-layers outside the immediate vicinity of the IHP main building and removed all data points with an accuracy larger than 40 m. This leaves 136 trajectories with 7,420 data points total. The median accuracy is 14.9 m and median number of data points per trajectory is 409 data points (mean 1032.83, standard deviation 1107.3). The mean traveled distance of trajectories from the filtered dataset is 356m with standard deviation 2259 m and median 17 m.

10.3.2. Generation of Synthetic Dataset

Two algorithms were tested for generating synthetic location data, DPT and AdaTrace. For fairness, we used $\varepsilon=1$ for all algorithms. AdaTrace is available at [A38]. The source code for DPT was available on the personal page of the main author at Duke University up until mid of 2021. The page has since been removed. We can therefore not provide an active link to the code.

10.3.2.1. DPT

For DPT, various types of parameter settings were tested. A pre-selection of parameters was done based on results on the Geolife dataset. The following parameters were analyzed:

- Sanitation: This parameter defines how the prefix tree is pruned. We test the two options suggested by the authors: '*geometric_adaptiveprune:0.1*' and '*geometric_fixedprune:0.0*'.
- Sampling: This parameter defines, weather direction are used when sampling is so, how the corresponding window is defined. The parameters used in the evaluation are: '*dir:1.20:15*' and '*nodir*'.
- Speed Arrays: This parameter sets the different grid resolutions for the algorithm given as an array in meters. The switch to a larger grid is assumed by the authors to be a switch to a different mode of transport. Since we are looking at a very small scale area, we select various small grid resolutions that can be seen as the differentiation of sitting, walking and running. We tested these parameters: '*5,15,30*', '*5,15,30,60*' and '*10,20,30,40*'.
- Model Depth: Additionally the depth of the model has to be passed when providing an array of speed values. We use the suggested default value *3*.
- Model Selection: There is also the option to have the algorithm learn the speed array. For this purpose, a part of the privacy budget can be allocated. We tested these two parameters: *0.01* and *0.1*.
- We unexpectedly found that the speed array default value from the DPT config file is taken into account even when a parameter for model selection parameter is passed to learn the speed array. We therefore tested two default values for the config file: *'100,200,400,800,1600,3200'* (the default set in the DPT code) and '*5,15,30*'.

10.3.2.2. AdaTrace

The main advancement from DPStar to AdaTrace is the built in defense against various attack vectors. The code allows to turn these defenses on and off. In the evaluation we test both settings.

10.3.3. Evaluation of Datasets

In the following we will take a look at the results of our evaluation. For each parameter configuration of each algorithm, 10 synthetic datasets are generated.

We consider the following metrics:

- Distribution EMD: Difference between synthetic data and ground data using the EMD. For each configuration the mean over all 10 runs and the standard deviation are reported.
- Diameter Error: As defined in Section 10.1 of Annex 2. For each configuration the mean over all 10 runs and the standard deviation are reported.
- Trip Distribution Error: As defined Section in 10.1 of Annex 2. For each configuration the mean over all 10 runs and the standard deviation are reported.
- Query Answering Error: As defined Section in 10.1 of Annex 2. For each dataset the relative error for 500 different queries is calculated. We report mean and standard deviation combined for all 10 runs for each configuration.
- Frequent patterns: For each dataset we calculate the top 100 frequent patters. We determine the percentage of frequent patterns that also existed in the original dataset (ignoring ranks). We use the different grid sizes 5m, 10m and 20m which are applied before frequent patterns are calculated. For each configuration and grid size we calculate mean, standard deviation and the mean over the F1 scores over all 10 runs. We investigated several algorithms for frequent patter mining. The commonly known apriori algorithm does not recognize repetitions of items in the itemset, but instead ignores them. It is therefore not suitable as we are looking at sequences which might return to earlier locations. The PrefixSpan algorithm suffered from the problem, that the collected dataset has many long trajectories which stay at certain locations for a long time. As a result the set of frequent sequential patterns is crowded with patterns of different length containing the same information. As a result we decided to remove sequential repetitions of locations from all trajectories. Additionally, we opt to used the BIDE algorithm for mining closed patters. The algorithms aims at finding a set of patterns for which no super-pattern with a larger support exists.

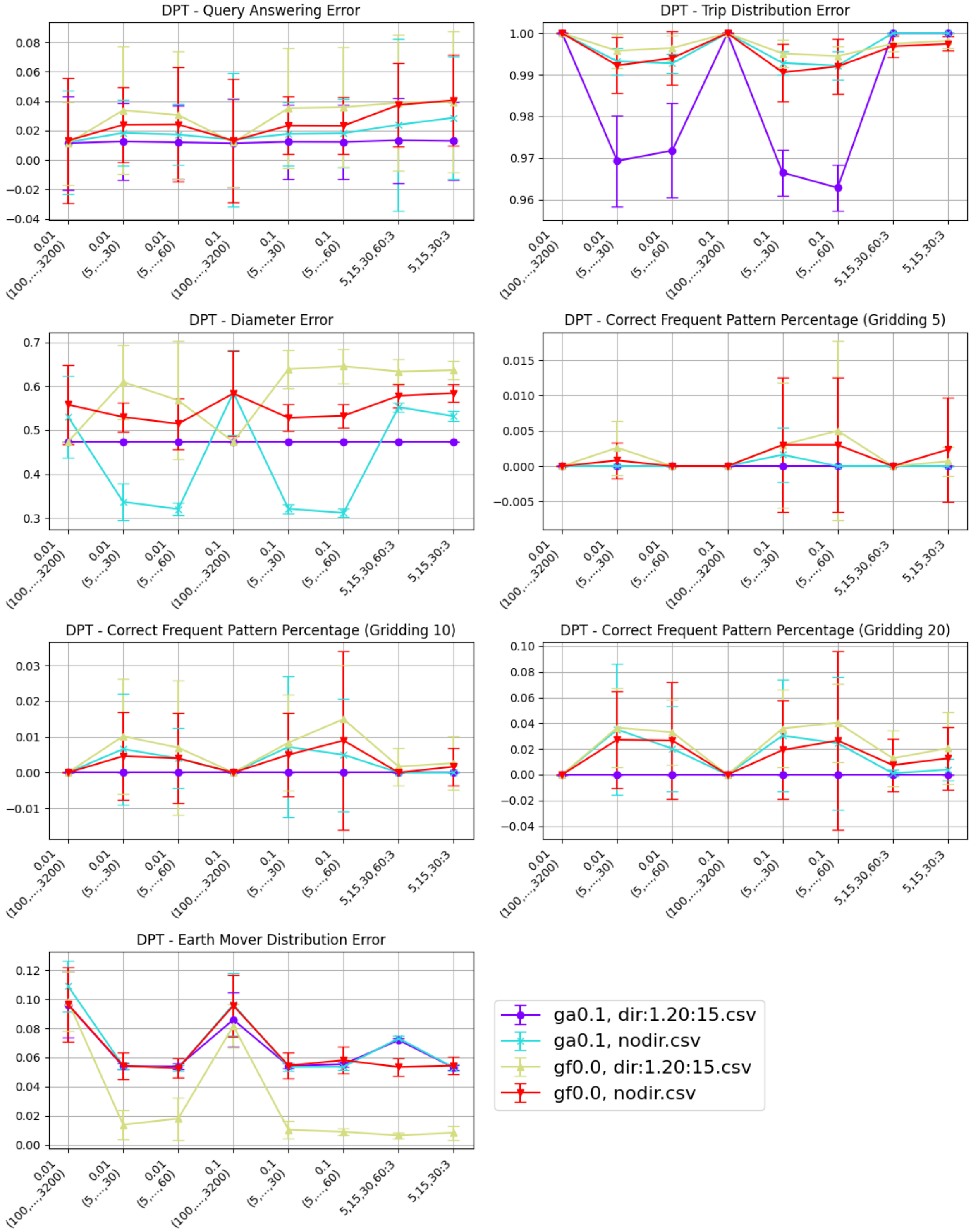

Figure 10-3 Metrics for various parameter combinations of DPT. X axis are parameters for model selection with default value or the speed array with model depth used instead. Y axis represents the corresponding error. Geometric adaptive pruning is shortened to 'ga', geometric fixed pruning is shortened to 'gf'.

10.3.3.1. DPT Parameter Exploration

To find the most suitable parameters for DPT we compare results for the parameter combinations discussed in 4.2.1. Figure 10-3 presents results for our metrics.

For almost all metrics, manually setting speed Arrays did not provide an improvement. Looking at frequent patterns it is easy to see that automatically learning the speed array performs better, especially when a larger portion of the privacy budget is spent on this task. Using smaller or for our use case more realistic values for the default speed array did also improve metrics. This is particularly visible when comparing the frequent pattern values for a default of *'100,200,400,800,1600,3200'* against a default of *'5,15,30,60'*.

Query answering, trip distribution and diameter error are better for adaptive pruning than for fixed pruning. This is not the case however for frequent patterns and the overall distribution.

Also surprising is the fact, that adding direction information does not always improve the data quality.

For these reasons we selected two DPT configurations to be compared with AdaTrace:

- DPT C1: '*geometric_adaptiveprune:0.1',* model selection value *'0.1'* with default *'5,15,30,60', 'dir:1.20:15'*
- DPT C2: '*geometric_fixedprune:0.0*', model selection value *'0.1'* with default *'5,15,30,60', 'dir:1.20:15'*

This evaluation has shown that DPT is highly parametrizable and that adaptive pruning is not always the best solution.

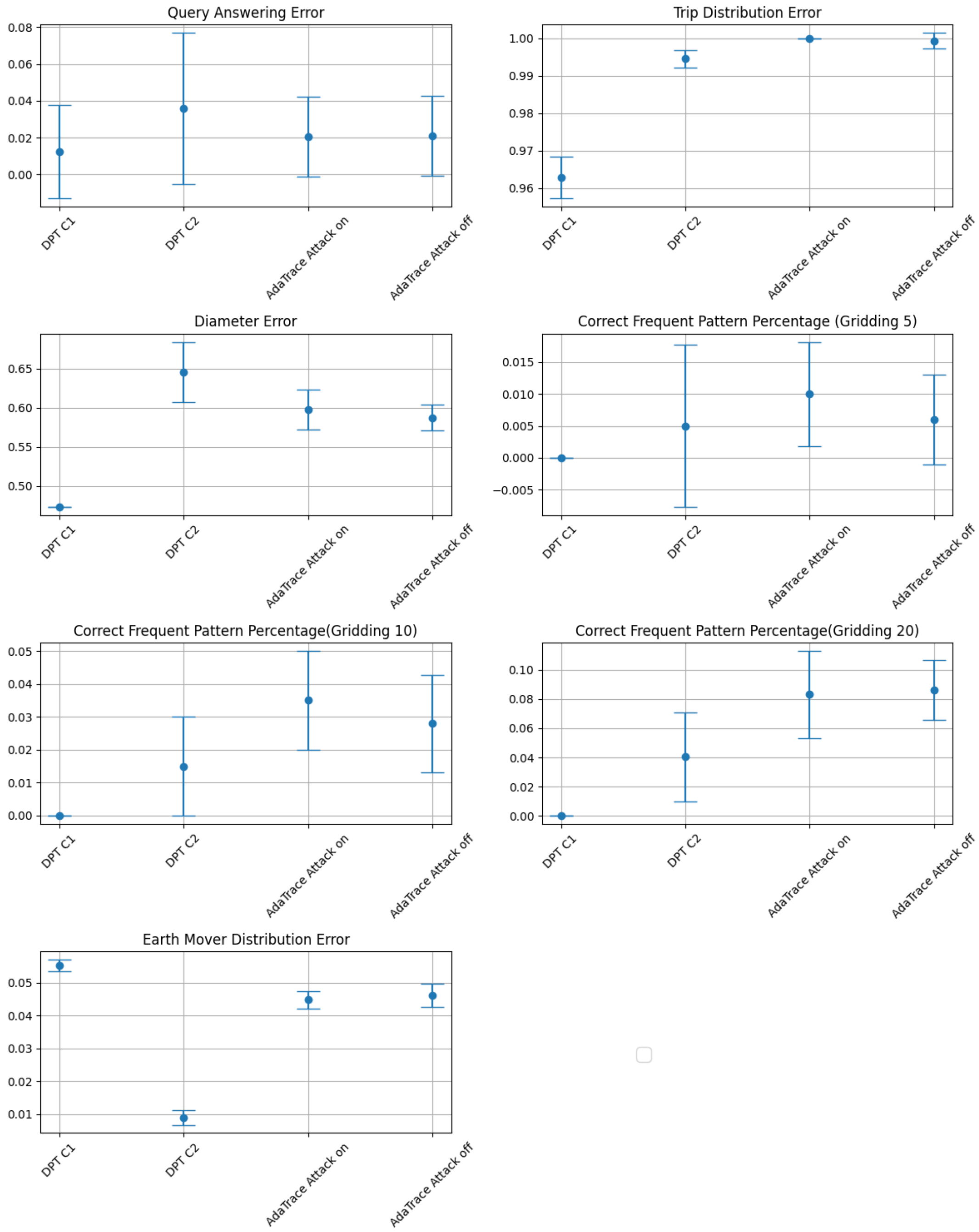

Figure 10-4 Comparison of AdaTrace and DPT on various metrics.

10.3.3.2. Comparing DPT and AdaTrace

To figure out which algorithm performs better on our data we compare both algorithms for two different configurations. See Figure 10-4 for the overview of results.

Query answering is equally good for all configurations, although DPT C2 has a larger standard deviation compared to the others. Regarding diameter error and trip distribution DPT C1 performs best. Ada Trace does not seem to preserve trip distribution. For the diameter error it ensures better results than DPT C2. For frequent patterns AdaTrace produces better results than DPT. DPT C1 does not seem to preserve any frequent patterns. For the comparison of the point distributions DPT C2 was the closest to the ground distribution. Both configurations of AdaTrace performed better than DPT C1.

While not always providing the best results, Ada Trace gives the best overall utility for the selected metrics. Having attack defenses turned on had very little effect on the results.

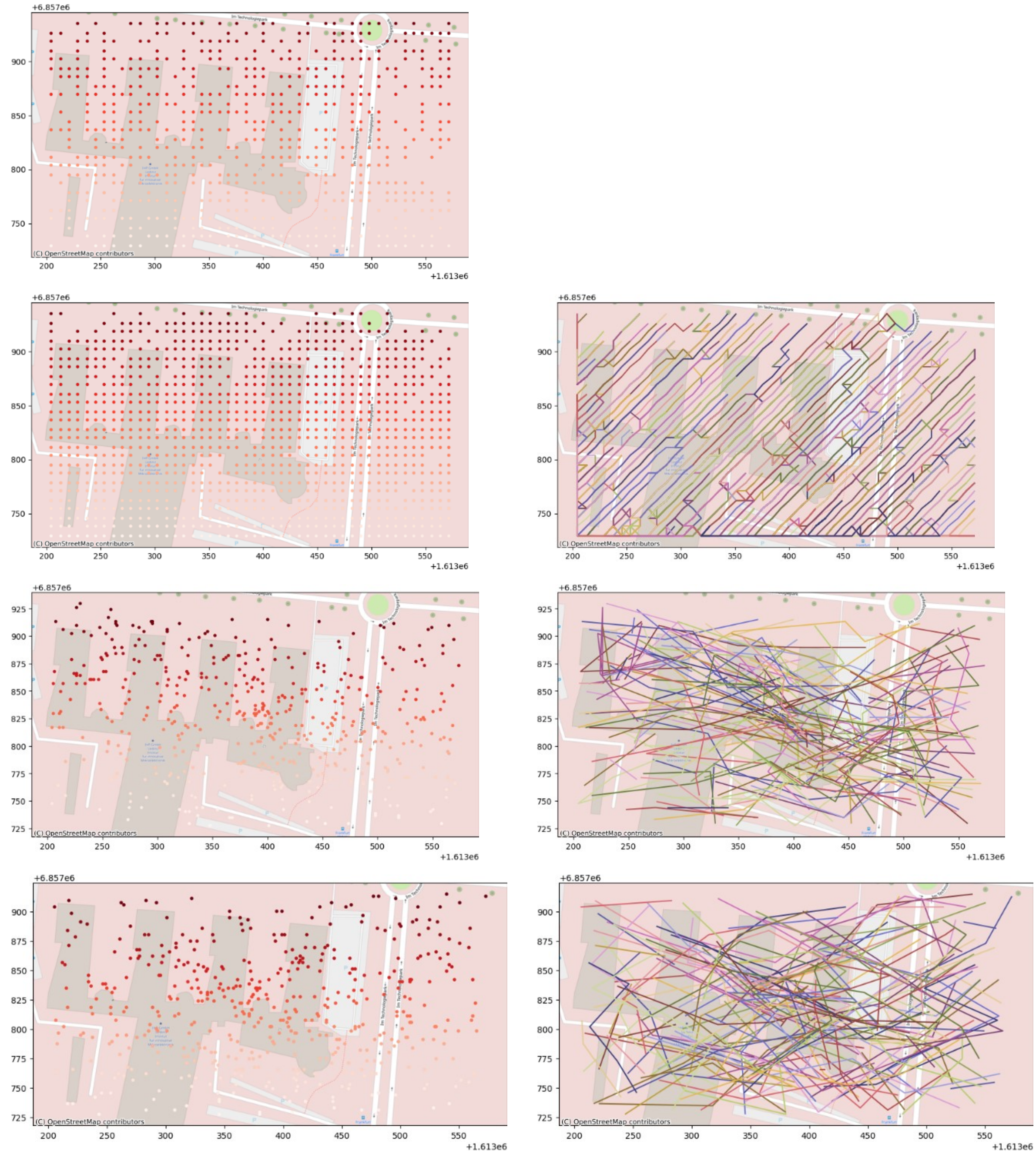

Figure 10-5 A qualitative overview of the synthetic data generated by the different configurations of AdaTrace an DPT. On the left the point distribution, a darker color indicates a larger number of data points at the same location. On the right side, are the generated trajectories plotted. From the top: DPT C1, DPT C2, AdaTrace with attacks off, AdaTrace with attacks on. No trajectory plot for DPT C1 was included as this configuration only created trajectories consisting of a single data point.

In Figure 10-5 we present a qualitative overview over the generated data. We can see that both algorithms, but in particular DPT, generate data where there initially was none. AdaTrace seems to be better at keeping clusters. While still very random, AdaTrace trajectories seem more realistic than those generated by DPT.

Looking at the above evaluations it is also important to highlight that differentially private algorithms perform better the more data is available. This is due to the fact that less noise has to be added to hide the contribution of a single person. The assumption is that for a larger dataset all tested algorithms would have performed better. As not enough data was available it is difficult to say how much more data would be necessary to preserve good utility with minimal error.

10.4. Conclusion

This chapter looked at the usage of synthetic data for releasing private datasets to the public. The Markov model-based algorithms DPT and AdaTrace were selected for evaluation. Results were compared for seven different metrics and generated data was also qualitatively analyzed.

We have seen that while there is much research on synthetic data a long way is still ahead, especially when it comes to providing utility. There is always a subset of metrics that are considered relevant for one use case while on a different dataset other metrics are more important or relevant. Synthetic data forces the data publisher to select features which are preserved. This makes open data exploration more complicated and limit the usefulness of data. On the other hand it allows access to datasets which would not have been published otherwise.

When considering the context of user data collected at experiments such as the Festival of Lights or the IHP experiments, the combination of location data paired with quality of experience data is also interesting. Anonymizing such pairs of data using differential privacy while preserving as much utility as feasible is an interesting possible research direction.

10.5. References

[A1] KITCHIN, Rob. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.

[A2] European Commission. Data protection in the EU. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en, Accessed: 2021-12-17.

[A3] NARAYANAN, Arvind; SHMATIKOV, Vitaly. Robust de-anonymization of large sparse datasets. In: *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 2008. S. 111-125.

[A4] BERLEE, Anna. Using NYC Taxi Data to identify Muslim taxi drivers. *The Interdisciplinary Internet Institute*, 2015.

[A5] TOCKAR, Anthony. Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset. 2014. https://agkn.wordpress.com/author/atockar/, Accessed: 2021-12-15.

[A6] MURRAY-RUST, Peter. Open data in science. *Nature Precedings*, 2008, S. 1-1.

[A7] Kaggle. MNIST. https://www.kaggle.com/c/digit-recognizer, Accessed:2021-15-16.

[A8] DWORK, Cynthia, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 2014, 9. Jg., Nr. 3-4, S. 211-407.

[A9] KOLOURI, Soheil, et al. Sliced Wasserstein auto-encoders. In: *International Conference on Learning Representations*. 2018.

[A10] Stanford InfoLab. Chapter 4 The Earth Mover's Distance (EMD). https://infolab.stanford.edu/pub/cstr/reports/cs/tr/99/1620/CS-TR-99-1620.ch4.pdf, Accessed: 2021-12-06

[A11] GURSOY, Mehmet Emre, et al. Differentially private and utility preserving publication of trajectory data. *IEEE Transactions on Mobile Computing*, 2018, 18. Jg., Nr. 10, S. 2315-2329.

[A12] HE, Xi, et al. DPT: differentially private trajectory synthesis using hierarchical reference systems. *Proceedings of the VLDB Endowment*, 2015, 8. Jg., Nr. 11, S. 1154-1165.

[A13] WANG, Senzhang; CAO, Jiannong; YU, Philip. Deep learning for spatio-temporal data mining: A survey. IEEE transactions on knowledge and data engineering, 2020.

[A14] CHOI, Seongjin; KIM, Jiwon; YEO, Hwasoo. TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part C: Emerging Technologies*, 2021, 128. Jg., S. 103091.

[A15] RAO, Jinmeng, et al. LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection. In: *11th International Conference on Geographic Information Science (GIScience 2021)-Part I*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[A16] YEOM, Samuel, et al. Privacy risk in machine learning: Analyzing the connection to overfitting. In: 2018 IEEE 31st Computer Security Foundations Symposium (CSF). IEEE, 2018. S. 268-282.

[A17] CHEN, Dingfan, et al. Gan-leaks: A taxonomy of membership inference attacks against generative models. In: *Proceedings of the 2020 ACM SIGSAC conference on computer and communications security*. 2020. S. 343-362.

[A18] GONG, Maoguo, et al. A survey on differentially private machine learning. *IEEE Computational Intelligence Magazine*, 2020, 15. Jg., Nr. 2, S. 49-64.

[A19] YU, Lantao, et al. Seqgan: Sequence generative adversarial nets with policy gradient. In: *Proceedings of the AAAI conference on artificial intelligence*. 2017.

[A20] ACS, Gergely, et al. Differentially private mixture of generative neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 31. Jg., Nr. 6, S. 1109-1121.

[A21] CHEN, Si, et al. RNN-DP: A new differential privacy scheme base on Recurrent Neural Network for Dynamic trajectory privacy protection. *Journal of Network and Computer Applications*, 2020, 168. Jg., S. 102736.

[A22] RAHMAN, Md Atiqur, et al. Membership Inference Attack against Differentially Private Deep Learning Model. *Trans. Data Priv.*, 2018, 11. Jg., Nr. 1, S. 61-79.

[A23] JAYARAMAN, Bargav; EVANS, David. Evaluating differentially private machine learning in practice. In: *28th USENIX Security Symposium (USENIX Security 19)*. 2019. S. 1895-1912.

[A24] FAROKHI, Farhad; KAAFAR, Mohamed Ali. Modelling and quantifying membership information leakage in machine learning. *arXiv preprint arXiv:2001.10648*, 2020.

[A25] STADLER, Theresa; OPRISANU, Bristena; TRONCOSO, Carmela. Synthetic Data–Anonymisation Groundhog Day. To appear at: 31st USENIX Security Symposium.

[A26] BOUSQUET, Olivier; LIVNI, Roi; MORAN, Shay. Synthetic Data Generators--Sequential and Private. *Advances in Neural Information Processing Systems*, 2020, 33. Jg.

[A27] FRIGERIO, Lorenzo, et al. Differentially private generative adversarial networks for time series, continuous, and discrete open data. In: *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, Cham, 2019. S. 151-164.

[A28] CHEN, Rui; ACS, Gergely; CASTELLUCCIA, Claude. Differentially private sequential data publication via variable-length n-grams. In: *Proceedings of the 2012 ACM conference on Computer and communications security*. 2012. S. 638-649.

[A29] BINDSCHAEDLER, Vincent; SHOKRI, Reza. Synthesizing plausible privacy-preserving location traces. In: *2016 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2016. S. 546-563.

[A30] ZHENG, Yu, et al. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 2010, 33. Jg., Nr. 2, S. 32-39.

[A31] MOREIRA-MATIAS, Luis, et al. Predicting taxi–passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 2013, 14. Jg., Nr. 3, S. 1393-1402.

[A32] BRINKHOFF, Thomas. A framework for generating network-based moving objects. *GeoInformatica*, 2002, 6. Jg., Nr. 2, S. 153-180.

[A33] GURSOY, Mehmet Emre, et al. Utility-aware synthesis of differentially private and attack-resilient location traces. In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 2018. S. 196-211.

[A34] GHANE, Soheila; KULIK, Lars; RAMAMOHANARAO, Kotagiri. TGM: A generative mechanism for publishing trajectories with differential privacy. *IEEE Internet of Things Journal*, 2019, 7. Jg., Nr. 4, S. 2611-2621.

[A35] AL-HUSSAENI, Khalil, et al. SafePath: Differentially-private publishing of passenger trajectories in transportation systems. *Computer Networks*, 2018, 143. Jg., S. 126-139.

[A36] LI, Meng, et al. Achieving differential privacy of trajectory data publishing in participatory sensing. *Information Sciences*, 2017, 400. Jg., S. 1-13.

[A37] National Coordination Office for Space-Based Positioning, Navigation, and Timing. GPS Accuracy. https://www.gps.gov/systems/gps/performance/accuracy/, Accessed: 2021-11-30

[A38] GURSOY, Mehmet Emre. AdaTrace. https://github.com/git-disl/AdaTrace, Accessed: 2021-12-15.