5Genesis

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

[H2020 - Grant Agreement No. 815178]

Deliverable D6.1

Trials and experimentation - cycle 1

Editors M. Emmelmann, F. Eichhorn, C. Fuhrhop (Fraunhofer)

Contributors Fraunhofer (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 (NEM), NCSR Demokritos (NCSRD), University of Malaga (UMA), Cosmote (COS), Universitat Politecnica de Valencia (UPV), Atos (ATOS)

Version 1.0

Date July 31st, 2019

Distribution PUBLIC (PU), Annex 4 is CONFIDENTIAL (CO)

List of Authors

Fraunhofer	FOKUS-Fraunhofer Gesellschaft e.V., Institute for Open Communication Systems
M. Emmelmann, M. Monachesi, R. Shrestha, F. Eichhorn, C.Fuhrhop , T. Briedigkeit	
IHP	Innovations for High Performance microelectronics/ *Leibniz-Institut für innovative Mikroelektronik*
J. Gutiérrez, M. Ehrig, N. Maletic, E. Grass	
HUB	Humboldt University of Berlin
S. Dietzel	
KAU	Karlstads Universitet
A. Brunstrom, M. Rajiullah	
SRL	Simula Research Laboratory AS
G. Caso, C. Griwodz, Ö. Alay	
REL	RUNEL NGMT LTD
I. Koffmann	
ECM	EURECOM
P. Matzakos, F. Kaltenberger	
INF	INFOLYSiS
V. Koumaras, C. Sakkas	
INT	INTEL
V. Frascolla	
UNIS	University of Surrey
S.Vahid	
FOG	Fogus Innovations & Services
D. Tsolkas, N. Passas	
NEM	NEMERGENT SOLUTIONS SL
E. Alonso	
NCSRD	NATIONAL CENTER FOR SCIENTIFIC RESEARCH “DEMOKRITOS”
G. Xilouris, M. Christopoulou, H. Koumaras, T. Sarlas, T. Anagnostopoulos	
UMA	UNIVERSITY OF MALAGA
A. Díaz, I. González, B. García, P. Merino	
COS	COSMOTE KINITES TILEPIKOINONIES AE
I. Mesogiti, F. Setaki, E. Theodoropoulou	
UPV	Universitat Politecnica de Valencia
J. Suárez de Puga	
ATOS	ATOS SPAIN SA
E. Jimeno, J. Melián	

Disclaimer

Copyright

The 5GENESIS Consortium consists of the following partners:

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	Release of D6.1	M. Emmelmann, F. Eichhorn, C. Fuhrhop	29.JUL.19

LIST OF ACRONYMS

Acronym	Meaning
5G	5-th Generation of cellular mobile communications
5G NR	5G New Radio
5G-PPP	5G Public-Private Partnership
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
ARM	Advanced RISC Machine
AWGN	Additive white Gaussian noise
BLER	Block Level Error Rate
BS	Base Station
CS	Circuit Switch
COTS	Commercial-Off-The-Self
DUT	Device Under Test
DRAN	Distributed Radio Access Network
DWDM	Dense Wavelength Division Multiplexing
EC	Energy Consumption
EE	Energy Efficiency
ELCM	Experiment Life Cycle Manager
eMBB	Enhanced Mobile Broadband - 5G Generic Service
EMS	Element Management System
E2E	End-to-end
EPC	Evolved Packet Core
E-UTRAN	Evolved Terrestrial Radio Access Network
FCAPS	Fault, Configuration, Accounting, Performance and Security
FPGA	Field Programmable Gate Array
gNB	(Next) generation node B
GDPR	General Data Protection Regulation
HEVC	High Efficiency Video Coding
IoT	Internet of Things
IaaS	Infrastructure as a Service
KPI	Key Performance Indicator
LBO	Local Break-Out
LTE	Long Term Evolution
MANO	Management & Orchestration
MEC	Multi-access Edge Computing
MIMO	Multiple Input Multiple Output
MCPTT	Mission critical push-to-talk

MCS	Mission critical services
MME	Mobility Management Entity
mmWave	Millimeter Wave
MN	Mobile Network
NB-IoT	Narrow Band – Internet of Things
NLOS	Non Line of Sight
NS	Network Service
NSA	Non-Stand-Alone
OAI	Over the Air Integration
OSS	Operational Support Services
OTA	Over-The-Air
PL	Packet Loss
PFCP	Packet Forwarding Control Protocol
PS	Packet Switched
PTMP	Point-to-Multi-Point
QoS	Quality of Service
RAT	Radio Access Technology
REST	Representational State Transfer
RRH	Remote Radio Head
RTP	Real-time protocol
RTSP	Real-time Streaming Protocol
SA	Stand-Alone
SDK	Software Development Kit
SDN	Software Defined Networking
SUT	System Under Test
TaaS	Testing as a Service
TAP	Test Automation Platform
TR	Time Ratio
TRxP	Transmission Reception Point
UE	User Equipment
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

Executive Summary

This deliverable describes the trials and experimentation results from the first integration cycle of 5GENESIS. Upcoming versions of this deliverable will describe the trials and experimentation results from the second integration cycle (D6.2, M21) and the third integration cycle (D6.3, M36).

After defining the KPIs to be validated and the metrics to be measured, the core of the document describes in detail the selected fourteen primary test cases and testing procedures.

Finally, the deliverable presents the measured results of the experiments performed at the five platforms (Malaga, Athens, Limassol, Surrey, and Berlin) run by the 5GENESIS consortium.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1. Purpose of the document

During the last years, standardisation bodies, industry alliances and regulatory bodies have put a lot of effort into defining the services 5G networks shall deliver. Additionally, they have defined the Key Performance Indicators (KPIs) [1] and target values as part of a quantitative assessment. All these efforts have led to a number of results that are indicative objectives to be met by operational 5G deployments [2].

In this context, the aim of the 5GENESIS project is to evaluate various 5G equipment and network deployments (such as those comprising the five 5GENESIS platforms), towards the achievement of the KPIs' targeted values with respect to those expected in commercial 5G network deployments. Additionally, this assessment will allow to identify the critical parameters that can impair the achievement of those targets values in future, commercial 5G deployments.

For the purpose of avoiding multiplication of work, and depending on the specific technical characteristics of each 5GENESIS platform, work related to KPIs evaluation (investigation of critical factors, and testing) has been divided between the five platforms.

This deliverable describes the trials and experimentation results from the first integration cycle of 5GENESIS. Upcoming versions of this deliverable will describe the trials and experimentation results from the second integration cycle (D6.2, M21) and the third third integration cycle (D6.3, M36). To better depict the progress conducted, it is expected that those documents will maintain the same structure as this deliverable.

1.2. Structure of the document

This document is structured in twelve sections and four annexes. A brief description of each sections follows.

Section 2 describes the measurement concept and methodology used, and defines a number of terms used in later sections of the document. It provides the statistical background for the post processing of measurements, common to all test cases.

Section 3 provides an overview of the KPIs provided by 5G-PPP, describes the methodology used and introduced the general results template used for all experiments.

Section 4 gives the specific definition of an abstraction of those 5G-PPP KPIs, which, in the framework of 5GENESIS, they are known as "Metrics".

Section 5 provides a detailed description of every individual test, including information about the target KPI, the methodology, the calculation process and output, the potential complementary measurements, the pre-conditions required, as well as the applicability and test case sequence.

Section 6 gives a short overview of the traffic profiles used.

The subsequent *sections 7 to 11* give descriptions on the experiments performed at each of the five platforms used in 5GENESIS (Malaga, Athens, Limassol, Surrey and Berlin) and the results obtained.

Section 12 includes the conclusions.

The annexes provide background information on energy efficiency specifications, the assessment of energy efficiency of Cloud RAN (CRAN) networks, as well as a reference to confidencial measuraments conducted in pre-commercial 5G equipment.

1.3. Target Audience

The primary target audience of this first WP6 deliverable encompasses industry and standardization stakeholders, allowing them to validate the 5G KPIs, based on the description of the test cases and the subsequent experimentation results from the first integration cycle, providing the joint evaluation of the results obtained from the experiments in the different platforms.

As the approach is based on industry best practices, this deliverable is best suited for industry stakeholders, although not limited to them.

Other stakeholders that can benefit from the document include:

- Standardisation organizations
  Where the test cases can form the basis of test suites.
- European Commission
  To evaluate the conduction and results of 5G experimentation.
- 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. MEASUREMENT CONCEPTS AND METHODOLOGY

2.1. Measurement system vs. device and system under test

Any solid interpretation of performance measurements requires a thorough description of the full measurement environment encountered while performing the test. Such a description, in general, distinguishes between the system or equipment evaluated and the testing or measurement environment used to conduct the evaluation. As such, the following definitions hold:

- *Measurement system*: One or more measurement devices and any other necessary system elements interconnected to perform a complete measurement from the first operation to the end result [1] [3].
- *Device under test (DUT*): The device to be placed in a test fixture (measurement system) and tested [1]. Usually, a single device being tested [3].
- *System under test (SUT)*: A system of devices, i.e., a specific combination of DUTs, being tested at the same time [3]. A SUT may – especially for virtualized network environments or software – include the computer system hardware and software on which the implementation under test operates [1].

In general, when reporting results of a performance evaluation – let it be a full end-to-end (E2E) 5G KPI evaluation or the KPI evaluation for a specific (sub) system of a 5G deployment – *an appropriate description of the full measurement system should be included. Such documentation of the measurement system should – if applicable – include the view of a potential virtualized network environment and placement of the virtualized network functions and the underlying physical components,* as especially the latter may impact the performance results observed for a given SUT or DUT.

The following example illustrates the interaction and influence of the virtualized network view of a measurement setup with its underlying physical view.

Figure 1 and Figure 2 illustrate exactly the same measurement system, which is used to assess the performance of a virtualized 5G Core. In the first measurement setup, the measurement probes as well as the SUT, i.e. the 5G Core, are all placed in the virtualized network environment on the same compute and storage (bare metal) device, whereas in the second setup, the two probes and the SUT are each placed at different devices. From the virtualized network point of view, both measurement systems have exactly the same properties: all components are directly attached to the same network and the test data flow goes from the first probe to the 5G Core and from there to the second probe. Still, exactly the same KPI evaluation (e.g. measuring the delay or throughput of a connection via the SUT (5G core) may result in completely different results.

The reason for that difference is the different physical architecture of the underlying physical infrastructure. Whereas in the first setup results are mainly impaired by the performance of the “compute and storage node (1)” (the data flow is only internal to the latter bare metal component), the data flow in the second setup has to go from the “compute and storage node (1)” over a leaf-switch to the “compute and storage node (2)”, and then over the leaf-switch, a spine-switch, to another leaf-switch and, finally, to the “compute and storage node (3)”. Given the properties of the physical infrastructure, results are impaired by the number of intermediate hops, by the compute capacity of all three "compute and storage" nodes, as well as by the link capacity. Note that the 1 Gbps capacity of the last switch and link is by nature limiting a potential throughput measurement to 1 Gbps. As such, a description of the measurement setup should always include both, the virtualized network view, and the bare-metal, physical infrastructure view.

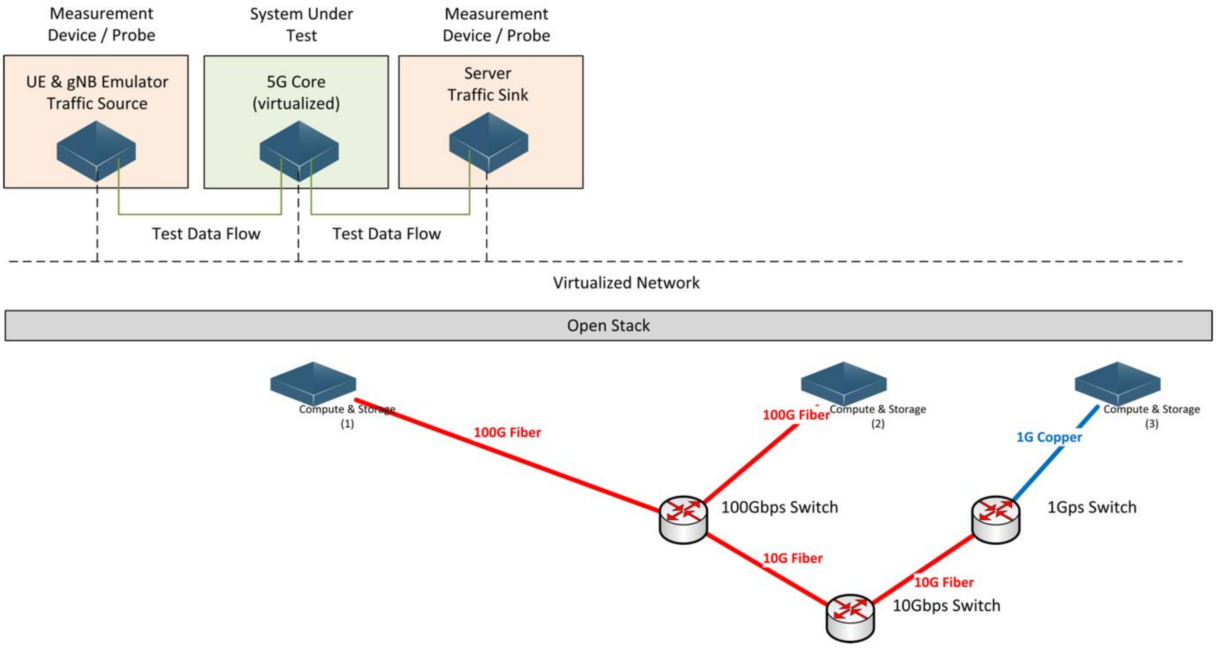


Figure 1 Example measurement system: clustered virtualized components.

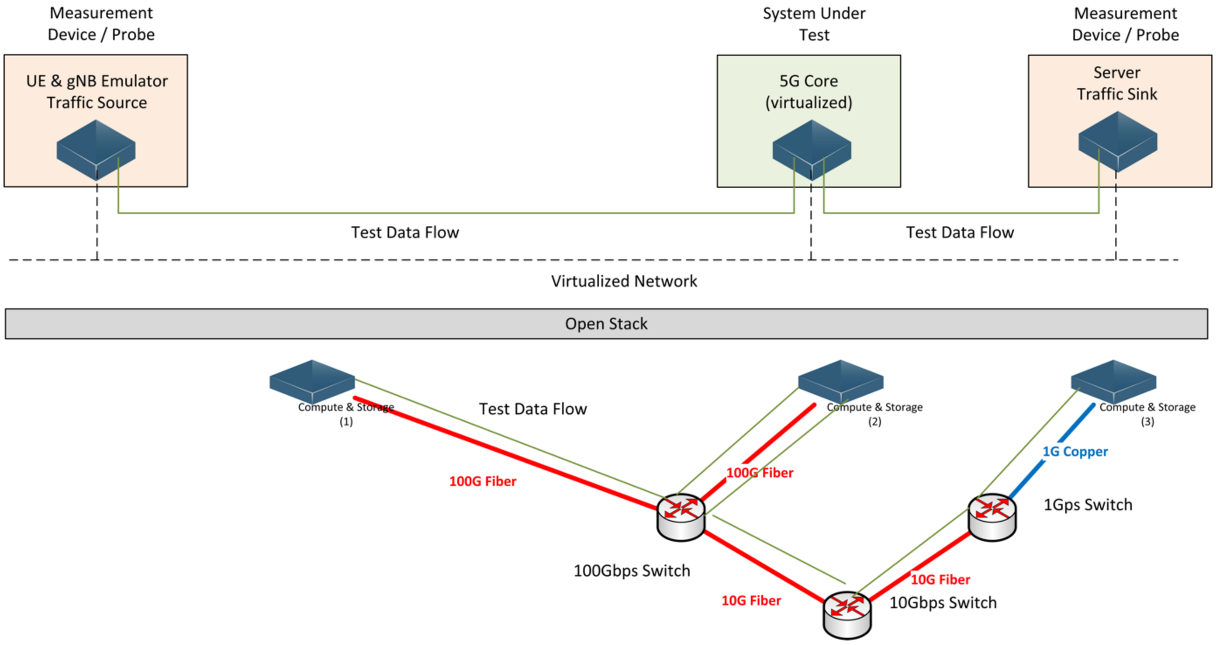


Figure 2 Example measurement system: distributed virtualized components.

2.2. Accuracy vs. precision

One of the main objectives of this document and, generally, of the work carried out in the framework of WP6, is to descibe a well-defined strategy to conduct measurement campaigns and to report the obtained measurement results. In this context, it is essential to distinguish between *accuracy* and *precision* when reporting measurement results.

The IEEE Standard 100 [1] defines *accuracy* as "the quality of freedom from mistake or error, that is, of conformity to truth or to a rule", and *precision* as "the quality of coherence or repeatability of measurement data, customarily expressed in terms of the standard deviation of the extended set of measurement results from a well-defined (adequately specified) measurement process in a state of statistical control".

Figure 3 illustrates the relation between accuracy and precision. Therein, the "bull's eye" represent the "golden, well know true value" of a parameter to assess. The "dots" represent the actual values measured by the experiment. Thus, accuracy – also noted as "trueness" – assesses how close a measurement is to the correct, i.e. "true", value for that measurement. The precision of a measurement system – also sometimes referred to as a gauge of repeatability or reproducibility – represents how close the agreement is between repeated measurements (i.e. repeated under the same conditions). Ideally, a measurement system (or device) is both accurate and precise, leading to measurements all close to and tightly clustered around the true value.

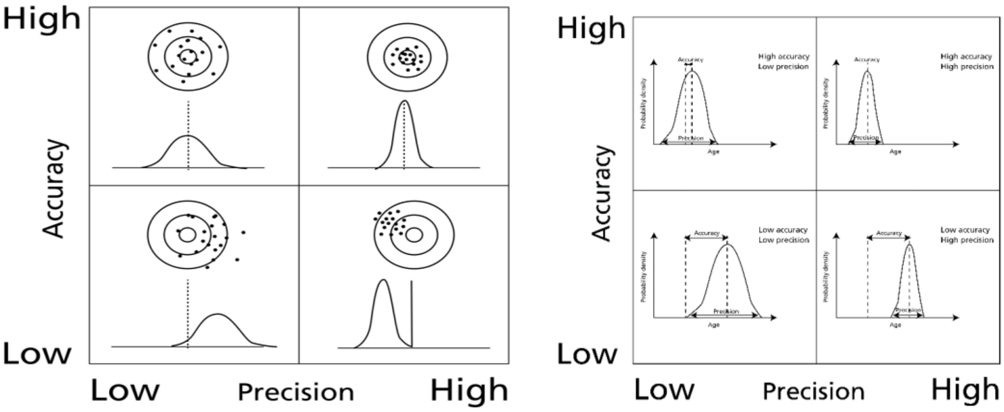


Figure 3 Relation between accuracy and precision.

In general, it is always a challenge to measure a KPI to assess a SUT together with the necessary consideration of accuracy and precision of the gained results. The precision of the measurement results can easily be quantified, e.g. by reporting standard deviation or confidence values. However, assessing the accuracy requires the knowledge of the "true value" which is most of the time unknown (as the goal of the measurement is the quantification of the latter).

In addition, the assessment of a 5G E2E KPI highly depends on the considered SUT, which in particular does not only include specific 5G components such as radio access and packet core, but also the underlying testbed infrastructure of a specific experimentation site, which includes characteristics of the network connectivity, switching capacity, and virtualizaiton aspects specifically found at a facility. As such, comparing results of a 5G KPI assessment conducted at different 5GENESIS facilities, or even among diffent ICT-17 testbeds (i.e., 5GENESIS, 5G-EVE, or

5G-VINNI) needs a proper methodology. One approach is proposed, consisting of the so called "calibration tests" and of the quantification of the precision of reported results. Such approach is followed by 5GENESIS and is described in the following section.

2.3. System Validation

The 5GENESIS facility comprises of five platforms and one portable demonstrator, each one with different infrastructure deployments, system capabilities, deployed services and measurement tools. The 5GENESIS project aims *at* facilitating a unified facility under the umbrella of a Coordination layer. The components of the Coordination layer, which are instantiated within each platform, are responsible for the KPI validation and use case demonstration. The 5GENESIS experimentation methodology employs two ways to ensure the validation of proper operation of the testing infrastructure and probe elements prior to execution of a test case. The first one, namely calibration, is more thorough and is defined specifically via several separate test cases. The second one, namely operational validation, is expected to require prior manual validation of the proper operation of the testing infrastructure.

2.3.1. Calibration tests

For any thorough system evaluation, literature suggest a measurement methodology in which two out of three evaluation procedures – namely analysis, simulation, and measurement – are independently employed [4]. In the case that only measurements are applied, the involved toolchain undergoes a *calibration test* in which the SUT is a stripped-down, well known component having known properties. Ideally, such a calibration test is conducted for each KPI to be evaluated and used for each measurement system (i.e. toolchain and involved infrastructure of the testbed). If such calibration measurements of the same (simplified) system under test produce comparable results for a given KPI when measured via different measurement systems, also preferably by different persons at different platforms, results may be assumed to be accurate within the precision limits of the given calibration measurement.

Calibriaton tests do not necessarily represent a scenario that characterizes a real-world use-case, but can be seen in general as an extremely simplified experiment, which allows to obtain a baseline performance of the underlying SUT. This allows, in particular, to provide a thorough interpretation of the outcome of any 5G E2E KPI evaluation. For example, a measurement of the achievable throughput might result for one experiment conducted on one testbed a value of 15 Gbps, while the same experiment conducted at another facility results in 10 Gbps even though both experiments involve the same 5G new radio components and the same 5G packet core. An adjunct set of calibration tests is cabable of quantizing the characteristics of the underlying infrastructure possibly showing that one platform is limited to 10 Gbps data throughput regardless of deployed 5G components, whereas the other is capable of handling 100 Gbps data. Thus, even though both platforms can report that the given 5G KPI is met, the calibration tests allow to provide an interpretation of the measurement data to state that the 15 Gbps throughput limit characterizes the 5G components, as the underlying system does not impose any performance limitiations.

As such, a poper set of calibration tests does not only involve the assessment of the underlying testbed infrastructure, which can potentially impact the outcome of a 5G E2E KPI assessment, but also includes a very simplified testcase for any 5G E2E KPI, which every testbed can easily run to have comparable results among testbeds, regardless of further experiments that characterize the KPI for a specifc use-case or vertical application.

The following example illustrates in detail such methodology for such a calibration measurement for assessing the round-trip time (RTT) imposed by a SUT.

The measurement system consists of two instruments, called “client” and “server”, which communicate over the SUT. Two different measurement systems are used: for the first measurement system, the client starts a “ping test” towards the server to probe the RTT and the results of the “ping” are directly taken out from the ping-application. For the second measurement system, ping is also used to trigger sending packets from the client to the server, but a packet capture tool is used to record the time between the ICMP-request emitted at the client and the received ICMP-response in order to calculate the RTT.

For the calibration measurement, the SUT is “empty”, i.e. client and server are *directly* connected, e.g., via an Ethernet cable. Results for the calibration show that independent of the used measurement system, the measured average RTT is 0.6 ms. Individual measurements are all in the interval of [0.3 ms; 0.9 ms]. Knowing that the SUT is “empty”, the measured RTT is caused by the overhead introduced by the measurement system itself. Thus, the calibration measurement shows that the measurement system has an *accuracy of 1 ms* in the worst case and produces precise results within a ±3 ms interval. Note: two independent measurement systems produced the same result for the “empty” SUT.

Now assume that the same measurement system is used to assess the RTT of a, e.g., router of firewall between client and server. Results show, e.g., a measured RTT of 8 ms (±0.5 ms). Considering the accuracy obtained by the calibration measurement, one can derive that the “true” RTT introduced by the SUT is within [7 ms; 8 ms] ±0.5 ms precision interval. As accuracy is in general reported in percent, the results has a 12.5% accuracy (1 ms /8 ms).

For completeness, it should be mentioned that a calibration measurement might not be necessary if the “true value” is known. This is for example the case for assessing location accuracy if the SUT (more precisely a device under test) is placed at a well-known location, for which the “true” position is known via GPS coordinates. In that case, the accuracy of the location information reported by the device may be derived by comparing it to the “true” GPS value (though, of course, the error in accuracy for the GPS information itself should be considered as well).

2.3.2. Operational validation

In general, the operational validation can be considered as a stripped-down calibration test, which only validates the pure operation of the fascility. For example, a simple ping test might be conducted via two components in order to verify that they are both “reachable”. In contrast to a thorough calibration test, operational validation does not quantify the performance of one ore several components of the underlying system.

2.4. Post-processing of measurements

To analyze and validate 5G KPIs, 5GENESIS targets the automatic execution of a large amount of experiments. Given a KPI, its evaluation is carried out within several test cases, which differ in terms of infrastructure configurations and network conditions; each test case contains several iterations of a single test. The test is repeated over a statistically significant number of iterations. Overall, this allows to get a precise picture of the KPI, and to understand its behaviour across heterogeneous scenarios, which are given by the different test cases.

A full yet concise picture of the results of trials and test cases is thus needed; for this reason, the collected measurements are post-processed, and relevant statistical indicators are evaluated and reported as final outcomes.

The same methodology for the evaluation of the statistical indicators is applied to all the 5G KPIs targeted by 5GENESIS, presented in detail in Sections 4 and 5, and for this reason the methodology is presented here in a general form.

The following terminology is applied in this document:

- Experiment: A set of one or multiple test cases.
- Test case: A description of the procedure on how to evaluate a metric. A test case contains several iterations (replica) of a single test. Based on executing several replicas, a test case allows to quantify the precision of the reported result(s).
- Test: A set of one or more measurements that result in a statistical quantization of a metric, i.e., a test is a single replica/iteration.
- Iteration / replica: One execution of a test as described in the test case.

2.4.1. Sample vs. population and Running independent replicas

The main goal in assessing the performance (here 5G E2E KPI) of a 5G system is to quantify the universal behaviour of the SUT. In theory, such SUT can be described as a stochastic process, which is unkown, i.e. its parameters cannot be stated. To accurately characterize this process, one would have to consider an infinite number of drawings or measurements (population representing the process). Such an infinite number of measurements would allow to fully characterize the process, e.g. via the mean μ (of the populaton) [4].

In practice, it is impossible to conduct an unlimited set of measurements. Instead, a test with a finite number of measurements (samples from the population) is conducted and the statistical characteristics of such a test, in this example the sample mean $\bar{x}$, is likely to be different form μ. Even conducting several, independent replicas of the test will result in different values of the sample mean. It is important to distinguish between the two, i.e. population mean vs. sample mean, as the former is a fixed value whereas the latter is a statistic random variable. Following the law of large numbers, the mean of several sample means, i.e. $\bar{\bar{x}} = \frac{1}{i}\sum_i \bar{x}_i$, from a large number of tests should be close to μ, and will tend to become closer as more trials are performed. Besides, $\overline{x_i}$ follow a normal distribution, which allows to quantify the precision of $\bar{\bar{x}}$ by stating confidence intervals even for a limited number of tests using the Student-T distribution characteristics. Figure 5 illustrates this methodology, which in literature is also refered to as “batch means / independent replica analysis” [4] [80].

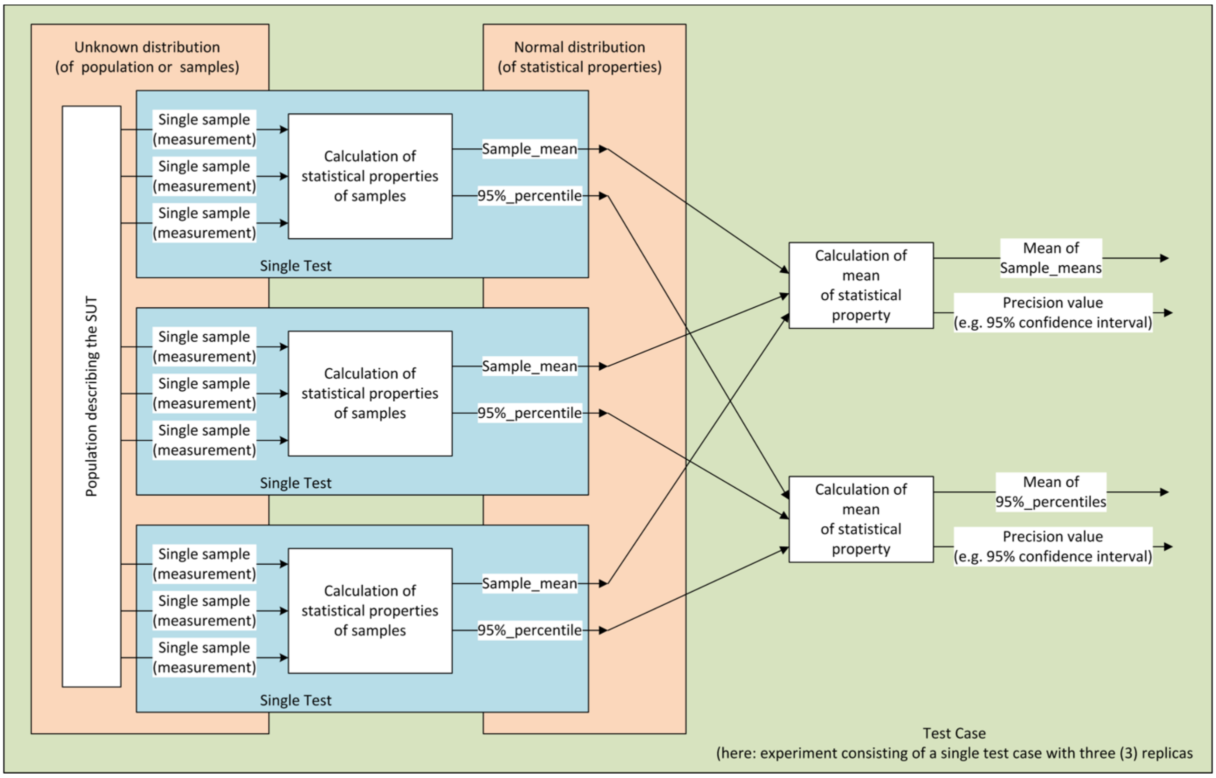


Figure 4 Example of batch means / independent replica analysis.

2.4.2. Calculation process for reported KPIs

As highlighted above, a single test case, focused on the evaluation of a KPI x in a predefined scenario (e.g., the evaluation of the throughput under specific network conditions), is repeated for a number of I iterations. Then, within the i^{th} iterration ($i = 1, \dots, I$), a number of N samples of the KPI are collected. A single KPI sample collected during the i^{th} iteration in the following is referred to as $x_{i,n}$ (with $n = 1, \dots, N$), while the entire set of samples collected during the same iteration is denoted by the vector $\boldsymbol{x}_i$. The statistical indicators for each iteration, are then computed, as reported in Table 1.

Table 1 Statistical indicators for a single iteration.

Indicator	Notation	Formula
Average (Mean)	$\bar{x}_i$	$\frac{1}{N}\sum_n x_{i,n}$
Standard deviation	σ_{x_i}	$\sqrt{\frac{1}{N}\sum_n (x_{i,n} - \bar{x}_i)^2}$
Median	x_i^{med}	$\begin{cases} \left(\frac{N+1}{2}\right)\text{th} & (N \text{ odd}) \\ \frac{\left(\frac{N}{2}\right)\text{th} + \left(\frac{N+1}{2}\right)\text{th}}{2} & (N \text{ even}) \end{cases}$ (Samples in ascending order, $(a)\text{th}$ indicates the sample at the a^{th} position in the ordered vector)

p%-Percentile $0 < p \leq 100$	$x_i^{p\%}$	$\left(\left\lceil\frac{p}{100}\times N\right\rceil\right)\text{th}$ (Samples in ascending order, $(a)\text{th}$ indicates the sample at the a^{th} position in the ordered vector, $\lceil a \rceil$ indicates the ceiling operator (the least integer $\geq a$))))
Minimum	$x_i^{\min}$	$\min(\boldsymbol{x}_i)$
Maximum	$x_i^{\max}$	$\max(\boldsymbol{x}_i)$

The statistical indicators computed for each iteration are then used to compute the statistical indicators of the test case, for which the I iterations were executed. This is done by averaging the indicators for each iteration over the amount of iterations. Denoting as x_i^{stat} the generic statistical indicator for the i^{th} iteration, the corresponding value for the test case, x^{stat}, is then obtained as follows:

$$x^{\text{stat}} = \frac{1}{I}\sum_{i} x_i^{\text{stat}}$$

Moreover, since each statistical indicator of the test case is computed as an average over a limited amount of I samples, a t% Confidence Interval (CI) can be adopted to denote the precision of the provided outcome. In particular, the 95% CI is widely used, and defines an interval containing the true value of the sampled indicator, i.e., x^{stat}, with 95% probability. The CI is usually evaluated using a Student-T distribution (in particular when the number of samples is low) with a number of degrees of freedom, denoted as v, equal to the number of available samples minus one, resulting in $v = I - 1$ in the present case [4]. The following indication, for each statistical indicator of the test case, can be then given as final outcome:

$$x^{\text{stat}} \pm t_{.95}\frac{\sigma_{x_i^{stat}}}{\sqrt{I}}$$

where:

- $t_{.95}$ is the so-called t value (or t score), which depends on the CI being evaluated (95 % in this case) and v, and can be derived from tabular approximations of the Student-T distribution;
- $\sigma_{x_i^{stat}}$ is the standard deviation of the vector $\boldsymbol{x}_i^{stat}$, containing the outcomes x_i^{stat} of the statistical indicator under analysis for each iteration, which are used to derive the corresponding indicator x^{stat} of the test case;
- $\frac{\sigma_{x_i^{stat}}}{\sqrt{I}}$ is the so-called standard error.

3. INTRODUCTION OF KPIS AND EXPERIMENTATION METHODOLOGY

3.1. Introduction to 5G-PPP KPIs

To date, standardisation bodies such as 3GPP, ETSI and ITU, as well as industry alliances and regulatory bodies have put a lot of effort in defining the services with the required Quality of Service (QoS) to be delivered by 5G networks, as well as the 5G network deployments' features and capabilities required for this purpose. To this end, various KPIs and target values have been defined to assess the 5G infrastructure (user and network equipment capabilities) [65][66][67][68][69][70], the services (network services and application services delivered over 5G infrastructures) [71] and 5G network deployments' quality [72].

As part of the strategy of the European Commission w.r.t. collaborative funded research projects, the research results are intended to shape 5G standards, to validate relevant spectrum identification and to support a global 5G vision [8]. At this stage of development (ICT-17), the resulting infrastructures will be used to validate the technological options in a full system context, and to extract results regarding their capacity to deliver future, commercial 5G network deployments with performance meeting the aforementioned KPIs targets.

To this end, significant advances have been achieved in previous 5G-PPP phases [9], where the KPIs and corresponding evaluation procedures proposed in the collaborative work so far can be used to harmonize evaluation results coming from different sources. The overall goal is to facilitate a fair assessment and comparability of the different technical concepts considered for 5G.

The recent 5G-PPP Test, Measurement and KPIs Validation Working Group White Paper [2] provides a unified vision on the Test and Measurement topics for 5G, allowing for common procedures and terminology and provides substantiated answers to more high-level relevant questions.

At the same time, in the context of 5G-PPP activities (projects and collaborative works) the infrastructure and services' KPIs defined by standardization bodies and industry alliances have been compiled. From this exercise, the high-level, operational, 5G network deployment KPIs have been derived [73]. These 5G-network deployment KPIs essentially reflect the network service delivery objectives and requirements expected from operational 5G network deployments. They can be translated as requirements to be fulfilled by the network operators as stakeholders undertaking the role of delivering the network deployments, either enforced by regulation or by the market they address, irrespective of the underlying technological specifics of the system that is deployed.

All the abovementioned efforts have led to a number of target results, being the definition of the KPIs and objectives to be met by operational 5G deployments, as shown in Table 2.

Table 2 5G-PPP KPIs and Target Values for Network Deployments.

Capacity	
Target	1. Absorb 1 Tbps in the equivalent to a smart office (10 Tbps/km^2) 2. Reach a peak data rate between 1 and 10 Gbps for specific deployment scenarios and use cases 3. Deployment and operation of 10 small cells per km^2, and support of 10 Gbps per Remote Radio Head (RRH) in access domain
Ubiquity	
Target	>99,9% spatial availability (with satellite/terrestrial aggregation)
Speed	
Target	1. Stationary and urban pedestrian ≤ 5 km/h 2. Urban vehicular ≤ 30 km/h 3. Vehicular high speed ≤ 300 km/h
Latency	
Target	1. ≤10 ms E2E (data plane) 2. 2 ms on the air interface (radio interface)
Reliability	
Target	>99,999%
Density of users	
Target	Between 10.000 and 1.000.000 devices per km^2 for specific use cases
Location accuracy	
Target	One meter (1m) in 99% of the cases
Energy efficiency	
Target	>50% reduction in energy consumption (EC) in comparison to already available technology (for specific network components)
Service creation time	
Target	Decrease of service creation time by at least one order of magnitude, compared to 4G. Clear improvement of the level of automation of service related processes, i.e. activating group communications in Mission critical services (MCS)
Network management CAPEX/OPEX	
Target	>50% decrease in network management CAPEX/OPEX, as assessed by feedback from operators

The aforementioned KPIs are of target for operational 5G deployments, which will be extensively deployed during the coming years, supported by key ICT European players. The aim of 5G experimental deployments, such as those 5GENESIS fosters, is to investigate to which extent the currently available equipment can achieve the performance that is expected for operational 5G networks. In this process, 5GENESIS will identify the shortcomings of the current technology and set the path to tackle them.

In this regard, 5GENESIS has dedicated effort to define an experimentation methodology, included in deliverable D2.3 [12], which addresses these (or aspects of these) KPIs homogeneously, irrespectively of the underlying system specifics. This methodology is revised in Section

3.2 and brings forward the concept of "Metric", which represents a high-level definition of the target measurement parameter(s). A detailed definition of the 5GENESIS metrics is included in Section 4.

3.2. Experimentation Methodology

In this deliverable we revisit and refine the 5GENESIS experimentation methodology, which was initially described in deliverable D2.3 [12]. This section brings forward a more mature methodology, which has been designed to facilitate the execution of a series of tests and to allow for the validation of the 5G KPIs and the verification of 5G technologies with an E2E approach.

The key concepts of the 5GENESIS experimentation methodology were the following:

- the experiment descriptor, which contains all the information required by the platforms to run the experiments.
- the test cases, which defines the KPI targeted during the experiment, the procedure and the measurements that have to be collected in order to validate the KPI.
- the experimentation scenarios, which details the E2E conditions for running the experiments, such as the mobility and the location of the User Equipment (UE).
- slice configurations, which detail the E2E resources allocated for the execution of the experiments.

Within the 5GENESIS experimentation methodology, the term "Metric" refers to a generic high-level definition of a target quality factor (attribute) to be evaluated, i.e., a definition independent of the underlaying system, the reference protocol layer, or the tool used for the measurement. A metric is the umbrella for the definition of more specific KPIs. The list of metrics considered by the project are defined in Section 4. The template for defining a Metric is shown in Table 3.

Table 3 Metric template

	Metric Name	*-ID number-*
1	Metric Definition *Here goes the definition of the metric, which refers to a generic quality factor independent of the underlying system and of the layer in which we are measuring it*	

3.2.1. Experiment descriptor template

The Experiment Descriptor template has been updated with a new field called "List of Target Metric(s). This field includes the list of Metrics targeted in the experiment.

Moreover, there are other minor changes in the Experiment Descriptor template:

- Network Services (NS) descriptions have been included as part of the slice description (see row **List of Slice Configurations to be established**).
- The parameters related on the definition of custom experiments are now part of a common section called "**Secondary input required for custom experiments**".

The updated Experiment descriptor is shown in Table 4.

Table 4 Experiment descriptor template.

	Experiment Descriptor	*-ID number-*	
#	Description of the fields to be completed	Input Values	Importance
1	**Experiment details** Information required to uniquely identify the experiment. Note 1: A Security Manager is used for editing with safety and privacy data related to the experimenter) Note 2: Each experiment shall include all the combinations of the target metrics/test cases/scenarios/slice configurations listed in the following fields of this form. (one target metric linked to one test case, for a specific scenario and a slice configuration is the minimum requirement for a complete experiment).	*Experiment ID* *Owner ID* *Organization ID* *Platform ID* *Type of experiment*	Mandatory
2	**List of the Target Metric(s)** Selection of the metrics (identified by IDs) targeted by the experiment. (see the **Metric Template**)	*Metric ID1* ..	Mandatory
3	**List of Test Case(s) to be executed** Selection of the test cases (identified by IDs) to be used in the experiment. Note: A test case includes KPI-associated Information (KPI definition, measurement methodology, complementary monitoring needed, etc) linked to a metric from the list in the field above. (see the **Test Case Template**)	*Test Case ID1* *Test Case ID2* … *Test Case IDi* …	Mandatory
4	**List of Scenarios to be considered** Selection of the Scenarios (identified by IDs) for which the test cases (selected in the previous field) will be executed. Note: A scenario includes information related to all the parameters that affect the values of the KPIs to be measured, network deployment and environment conditions, etc. (see the **Scenario Description Template**)	*Scenario ID1* *Scenario ID2* … *Scenario IDi* …	Mandatory
5	**List of Slice Configurations to be established** Definition of the Slice templates (identified by IDs) that are required for the experiment(s).	*Slice ID*: *NSD ID 1* *Radio Conf.* *Extra parameters*	Mandatory

	(see the **Slice Configuration Template**)	…	
		Slice IDi Config	
		…	
		Slice IDi Conf	
		…	
	Traffic Description Template (at least one traffic source or service type should be specified)	*Traffic sources*	Optional
		Service Type	Optional
6	**Secondary input** **required for custom experiments**	*UEs identification*	Mandatory (unattended experiments)
		Application under test	Mandatory (unattended experiments)
		Intermediate reporting of KPIs and Time between intermediate reports	Optional

3.2.2. Test case template

The test case specifies the conditions of the SUT, the procedure to execute the tests, collect the measurements and compute the KPIs.

The test case template introduced in deliverable D2.3 has been updated by renaming some of the fields, adjusting their content and adding new fields. In particular, the field "**Test procedure**" has been renamed to "**Methodology**" and the sequence of actions to be ran during the execution of the test case has been moved to a new field named "**Test case sequence**". The "**Methodology**" includes the declaration of the required number of iterations, the monitoring time, the monitoring frequency, etc. The field "KPI computation procedure" and the field "Test case output" has been merged in a new field called "Calculation process and output". Finally, three new fields have been added:

- **Complementary measurements**. The measurements specified in this field are not the main target of the test case, but can be useful when interpreting of the outputs of the test case.
- **Pre-conditions**. To ensure that the test cases are executed in the same conditions, this field specifies the conditions that need to be met by the SUT before the execution of the test case.

Applicability. To verify whether the test case is applicable to the SUT, this field includes the list of features and capabilities that should be supported by the SUT when executing the test case.

Table 5 provides the final test case template used in this deliverable to specify the test cases in Section 5.

Table 5 Test case template

<table>
<tr><th colspan="2">Test Case Template</th><th>*-ID number-*</th><th>*-Related Metric ID-*</th></tr>
<tr><td>#</td><td colspan="3">Description of the fields to be completed</td></tr>
<tr><td>1</td><td colspan="3">Description of the target KPI

Here goes the definition of the target KPI. Each test case targets only one KPI (main KPI). However, secondary measurements from complementary KPIs can be added as well (see field 4 in this template). The definition of the main KPI specializes the related target metric (the ID of the related target metric is declared in the first row of this template). More precisely, the definition of the main KPI declares at least the reference points from which the measurement(s) will be performed, the underlay system, the reference protocol stack level etc...</td></tr>
<tr><td>2</td><td colspan="3">Methodology

Here the acceptable values for the monitoring time, the iterations required, the monitoring frequency, etc., are declared. The reference to the calibration test is taken from the test case. This is to facilitate the comparison between measurements.</td></tr>
<tr><td>3</td><td colspan="3">*Calculation process and output*

Here goes information related to the calculation process required. This is information may include details related to the underlay system. Here goes also the Units of the metric, and potentially a request for first order statistics (Min, Max, etc.)</td></tr>
<tr><td>4</td><td colspan="3">Complementary measurements

A secondary list of KPIs useful to interpret the values of the target KPI. Getting these measurements is not mandatory for the test case.</td></tr>
<tr><td>5</td><td colspan="3">Pre-conditions

Any requirement that needs to be done before *execution* of this test case. A list of test specific pre-conditions that need to be met by the SUT including information about **equipment configuration**, **traffic descriptor** i.e., precise description of the initial state of the SUT required to start executing the test sequence</td></tr>
<tr><td>6</td><td colspan="3">Applicability

A list of features and capabilities which are required to be supported by the SUT in order to execute this test (e.g., if this list contains an optional feature to be supported, then the test is optional)</td></tr>
<tr><td>7</td><td colspan="3">Test Case Sequence

Specializes the measurement process (methodology) of the metric for the selected underlay system. Measurements points and measurement procedure specification.</td></tr>
</table>

3.2.3. Scenario

The “Scenario” concept was introduced in deliverable D2.3. However, the template was not provided. This deliverable provides a detailed scenario template. The current version of the scenario template includes radio configuration parameters. In future deliverables the scenario template will be updated with parameters from the rest of the components of an E2E network. The parameters that are part of the definiton of the scenario are different from those specified by the slice. The parameters defined in the scenario establish the working point of the network and the location and mobility conditions of the UE.

The scenario template is meant to be a guideline for the definition of network scenarios to reproduce realistic conditions in which to perform the test cases. The list of parameters shown in Table 6 stems from a deep investigation of the radio parameters that could affect the performance of the KPIs under test . Depending of the platform, these parameters could be configurable or not. The configuration of these parameters is not mandatory.

Table 6 Scenario template.

Scenario Description Template		*-ID number-*
#	Description of the fields to be completed	
1	Radio access technology *4G,5G*	
2	Standalone / Non-Standalone (if applicable)	
3	Cell Power	
4	Frequency band: *Sub-6 GHz* *mmWave*	
5	Maximum bandwidth per component carrier *50 MHz, 100 MHz, 200 MHz, 400 MHz*	
6	Sub-carrier spacing *Sub 6 GHz: 15 kHz, 30 kHz, 60 kHz* *mmWave: 60 kHz, 120 kHz, 240 kHz, 480 kHz*	
7	Number of component carriers *Maximum number of CC = 16 (5G)* *Maximum number of CC = 5 (4G)*	
8	CP *Cyclic Prefix: normal, extended*	
9	Massive MIMO *Number of antennas on NodeB*	
10	MIMO schemes (codeword and number of layers) *The number of codewords per PDSCH assignment per UE* ○ *1 codeword for 1 to 4-layer transmission* ○ *2 codewords for 5 to 8-layer transmission.* *DL DMRS based spatial multiplexing (SU-MIMO/MU-MIMO) is supported* ○ *At least, the 8 orthogonal DL DMRS ports are supported for SU-MIMO* ○ *Maximum 12 orthogonal DL DMRS ports are supported for MU-MIMO*	
11	Modulation schemes	

	Downlink: QPSK, 16 QAM, 64 QAM, 256 QAM *UplinK: QPSK, 16 QAM, 64 QAM, 256 QAM*
12	Duplex mode *FDD, TDD*
13	TDD uplink/downlink pattern (if applicable) 0.5 ms, 0.625 ms, 1 ms, 1.25 ms, 2 ms, 2.5ms, 5 ms, 10 ms
14	Contention based random access procedure/contention free (if applicable)
15	User location and speed

3.2.4. Traffic description

A traffic template shown in Table 7 has been specified to define the traffic profiles used during the experiments.

Table 7 Traffic description template

Traffic Description Template		*ID number-*
#	Description of the fields to be completed	
1	Traffic sources *Here goes the description of the traffic sources that emulate the traffic from real applications or reproduce background traffic conditions*	
2	Service Type (optional) *Here goes a description of the service provided while the KPI is measured*	

3.3. Results gathering template

The final report after the execution of a set of related test cases in a particular scenario and slicing configuration shall follow the structure indicated in Table 8.

Table 8 Template for the final report of results.

Test Case ID				
General description of the test				
Purpose				
Executed by	Partner:		Date:	
Involved Partner(s)				
Scenario				
Slicing configuration				
Components involved *(e.g. HW components, SW components)*				

Metric(s) under study (*Refer to those in Section 4*)	
Additional tools involved	
Primary measurement results *(those included in the test case definition)*	
Complementary measurement results	

The fields in the final report of results are as follows:

Test Case ID(s):	A list of test cases that are executed in the same environment und the same conditions. In particular, such a set of test cases may include several tests to obtain different statistical properties of a given metric under study, e.g. test case on Average RTT, test case on 95%ile RTT
General description:	A verbal description of the executed test, highlighting its goal
Executed by (Partner):	The primary partner (or platform operator) executing the test
Executed (Date):	The date, when the test was executed
Involved Partner(s):	A list of partners who contributed to the test or were directly involved in its execution
Scenario:	A description of the scenario or experimental set-up underlying the test
Slicing configuration:	A description of the network slice used to execute the test
Metric(s) under study:	A list of metric names (according to Section 4) for which test cases are executed. This list may include apart from the primary metric, which is inherently given by the list of test case(s), additional, secondary metrics, which were optinally gathered during the execution of the test
Additional tools involved:	A list of tools, essential to the execution of the test, which are not described in the test case specification
Primary measurement results:	The results for the primary metric(s) covered by the listed test case(s). The stated results are to follow the format defined in the test case description, i.e. in general there is a single value for a KPI (e.g. average delay) in combination with a confidence value for that result
Secondary measurement results:	Results obtained for additional (optional) metrics

4. DEFINITION OF METRICS

As mentioned in Section 3.1, 5G-PPP has defined a set of high level KPIs to assess the quality and capability of future, commercial 5G network deployments to meet the envisioned services' QoS. 5GENESIS experimentation activities will revolve around these KPIs, with the aim to deliver the facilities along with the test methodology and procedures to conduct the experiments and to extract and process the obtained results. The overall purpose is to assess aspects that affect these high level KPIs. To maintain a mapping between the numerous test-procedures and methodologies of 5GENESIS and the 5G-PPP KPIs they address (fully or partially), those KPIs have been abstracted as generic Metrics, which are summarized in this section. Figure 7 sketches the mapping of the high-level KPIs to those obtained as the outcomes of the 5GENESIS test cases.

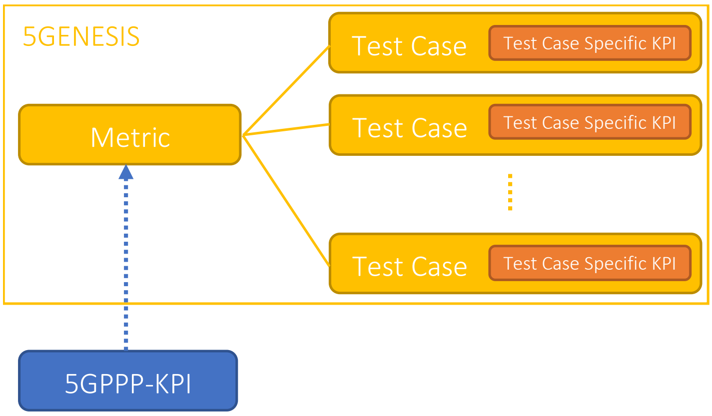


Figure 5 Mapping of KPI abstractions within the 5GENESIS framework.

4.1. Baseline metrics

4.1.1. Capacity

"Capacity 5G-PPP KPIs" refers to several aspects of the offered 5G network deployment capacity, namely:

1. The *offered network capacity per geographical area* (as defined by [4]), with the requirement to be equivalent to the total offered traffic to be served per geographic area unit. This is practically an operational 5G network deployment requirement reflected as KPI measurable in large-scale commercial deployments. This aspect is tightly related more to the network capacity planning and dimensioning rather than the underlying network technology. Of course, a number of network technology-related capabilities aspects and functionalities may influence the degree to which this capacity KPI and aspect is achieved.
2. The *peak user data rate* for specific deployment scenarios and use cases; reflecting the data rate requirements of a number of data-intensive applications [47] to be met by a single access network node equipment. This aspect is related to equipment-related capabilities and performance aspects.
3. The *minimum capacity of a single access network node,* along with

4. The *maximum number of access network nodes to be deployed per geographical area;* stemming from the previous market-imposed performance requirements, and being related to equipment-related capabilities and performance aspects.

Therefore, it becomes prominent that there is no single test to evaluate the generic 5G-PPP KPI capacity; instead aspects of the KPI can be evaluated through a number of equipment and deployment-related tests. For the purposes of having a common reference of these tests addressing capacity aspects (which can be used to evaluate the high level KPI, after processing), we define below the generic "capacity metric".

Metric	*Capacity*
Metric Definition *"Capacity":* Amount of data traffic to be offered or served, per a specific unit; the latter being a primary physical measurement unit (namely time and space) or/and a specific technology equipment component (namely user equipment, access network node, cluster of access network nodes with specifically defined characteristics, etc.).	

4.1.2. Density of users

"Density of users 5G-PPP KPI" is defined as the number of devices (i.e. users) per unit area, to be supported by an operational or commercial 5G deployment. The density of users depends highly on the specific functionality or service that is considered as simultaneously offered to them at a given reliability target. In general terms, this KPI and its set target value, has stemmed from the high-level objective to serve UE and the foreseen high density of IoT devices from a single operational network infrastructure [47]. In practice, the achievability of this KPI depends on a number of deployment-specific factors, namely the core network dimensioning, the access network deployment (including the number of next generation Node Bs (gNB), coverage planning end dimensioning, etc.), as well as on the equipment dimensioning capabilities.

Therefore, it becomes prominent that there is no single test to evaluate the generic "density of users 5G-PPP KPI"; instead aspects of the KPI can be evaluated through a number of equipment and deployment-related tests. For the purposes of having a common reference of these tests addressing user and device density aspects (which can be used to evaluate the high level KPI, after processing), we define below the generic "Density of users metric".

Metric	*Density of users*
Metric Definition *"Density of users"* Metric Definition: the maximum number of devices (i.e. users) per unit. This unit represents a physical measurement unit (namely time window and space/area) and/or a specific network component, which can be supported by the system, with a specific service that is simultaneously offered to all users at a given reliability.	

4.1.3. Energy efficiency

Metric	*Energy efficiency*
Metric Definition	
"Energy Efficiency (EE)" refers to minimization of energy used (consumption) in service delivery. More specifically, the network EE refers to minimization of the Radio Access Network (RAN) EC in relation to the traffic capacity provided, whilst device EE is the capability to minimise the power consumed by the device modem in relation to the traffic characteristics."	

4.1.4. Latency

"Latency 5G-PPP KPI" is considered (1) in the framework of a control plane as the time it takes to transfer a given piece of information from the end-user device (UE, IoT device, etc.) up to the 5G Core Network node(s) responsible for the network access control and service provisioning; and (2) in the framework of a user plane as the time it takes to transfer a given piece of information from the end-user device (UE, IoT device, etc.) up to the end providing the data service or application.

In general terms, these KPIs and their set target value, stem from the high level objective to serve highly interactive [70][68] and mission critical services [66][67][70]. In practice, the achievability of this KPI depends on a number of deployment-specific factors related to the network equipment control plane processing capabilities, the network topology with regard to the placement of the application serving nodes and, of course, the end-device location (within the serving network). Given the distributed network topology envisioned in 5G networks and user mobility, latency does not constitute a single feature/value throughout a network deployment, most probably also varying over time (depending on QoS provisioning policies).

Therefore, it becomes prominent that there is no single test to evaluate the generic "latency 5G-PPP KPI"; instead aspects of the KPI can be evaluated through a number of equipment and deployment-related tests. For the purposes of having a common reference of these tests addressing latency aspects, we define below the generic "latency metric".

Metric	*Latency*
Metric Definition	
"Latency": The time it takes to transfer a given piece of information from a source to a destination, from the moment it is transmitted by the source to the moment it is received at the destination (in this link direction only) over the SUT.	

4.1.5. Round-Trip-Time

Similar to Latency, the "Round-Trip-Time 3GPP KPI" is considered (especially for user plane) as the time it takes to transfer a given piece of information from the end-user device (UE, IoT device etc.) (herein transmitting node) up to the end providing the data service or application (herein receiving node), to process the piece of data at the receiving node, and to transfer an acknowledgement status back to the transmitting node [70].

As also clarified in 3GPP specifications [70], this generic performance indicator does not assume correct reception of either the piece of data or the acknowledgement status, while the nodes need to be defined. In practice, the definition of this KPI and its target value depends on a number of deployment-specific factors related to the network equipment processing capabilities, the network topology with regard to the placement of the application serving nodes, and of course the definition of the "processing" that takes place. A simple ping session can be considered as the minimal processing of a packet, thus roughly providing the sum of uplink and downlink latency.

Therefore, RTT does not constitute a single feature/value throughout a network deployment, and it is most probably also varying over time (depending on QoS provisioning policies). For the purposes of having a common reference of these tests addressing latency aspects, we define below the generic "RTT metric".

Metric Name	*Round-Trip-Time*
Metric Definition *"Round-Trip-Time":* Time it takes to transfer a given piece of data between two nodes, to process the piece of data at the receiving node, and to transfer an acknowledgement status back to the transmitting node, measured from the moment the piece of data is transmitted to the moment the acknowledgement status is received.	

4.1.6. Delay

"Delay 3GPP KPI" is considered as the time it takes to transfer a given piece of data between two nodes, measured from the moment it is transmitted to the moment it is received [70]. The difference with Latency in 3GPP terminology is that Latency refers to correct reception of the piece of information, while delay does not assume correct reception.

As also clarified in 3GPP specifications [70], this generic performance indicator does not assume correct reception of the piece of data, while the nodes need to be defined. In practice, the definition of this KPI and its target value depends on a number of deployment-specific factors as in the Latency and RTT case, thus delay does not constitute a single feature/value throughout a network deployment, and it is most probably also varying depending on network and processing conditions. For the purposes of having a common reference of these tests addressing delay aspects, we define below the generic "Delay metric".

Metric	*Delay*
Metric Definition *"Delay"*: Delay is the time it takes to transfer a given piece of data between two nodes, measured from the moment it is transmitted to the moment it is received; irrespectively of whether it is received correctly or not.	

4.1.7. Location accuracy

"Location Accuracy 5G-PPP KPI" is defined as the distance between the measured position of a device (UE, IoT device, etc.) and its true physical position, acquired in a number of cases.

Location accuracy depends on various factors such as: the equipment used, the type of measurements used (received and processed), the positioning algorithm(s), the mobile network (MN) deployment, the network traffic conditions and the user location with regard to the MN footprint – especially if positioning relies solely on MN measurements-, environmental conditions and so on. Thus, the results may vary significantly when taken under different spatio-temporal cases.

Therefore, it becomes prominent that there is no single test to evaluate the generic "location accuracy 5G-PPP KPI"; instead, it can be evaluated through a number of equipment and deployment-related tests. For the purposes of having a common reference of these tests, we define below the generic "Location Accuracy metric".

Metric	*Location accuracy*
Metric Definition *"Location"*: The distance between the measured position of an UE and its true position at a specific Cartesian coordinate system.	

4.1.8. Reliability

Reliability is a service level agreement relevant KPI, highly related to the definition of the "service" to be provided, and to the definition of a number of operational, spatio-temporal conditions of the service provisioning. In terms of network reliability KPI spans from data reliability (i.e. low data error rate/probability), to network availability in terms of coverage and resources. Therefore this KPI is usually relevant (thus measured) in operational or commercial network deployments, while the target set by 5G-PPP refers to covering the requirement for minimum service availability/connectivity for critical communication services [67][70] – thus to specific network slice/service provisioning (under spatio-temporal conditions to be set by the regulator of each country).

In test environments, reliability tests are more related to testing of functionalities, SW/HW units/configurations/deployment principles ensuring reliability. Therefore, also in this case, tests (and targeted results) assessing the "reliability" KPI and its achievability can be versatile; for the purpose of having a common reference of these tests within 5GENESIS, we define below the generic "Reliability metric".

Metric	*Reliability*
Metric Definition *"Reliability"*: the probability that an item (i.e. the SUT or DUT) will perform its intended function for a specified interval under stated conditions [1].	

4.1.9. Service creation time

Service Creation Time 5G-PPP KPI refers to the time required for the provision of a network service over a network physical and virtual infrastructure. Due to the versatility of network implementations and deployments favoured by 5G network specifications, this time may include various physical network domains and virtual components, depending on the requested service template. Thus, it may vary on the basis of requested service and specific network deployment (e.g. depending on Multi-access Edge Computing (MEC) availability, SDN technologies/infrastructure, scale of deployment from service provisioning end to service consuming end, etc.). For instance, Service creation time may include the time required to provision, deploy, configure, and activate the underlying communication infrastructure, i.e. a network slice, including all physical and virtual components; or it may not include the latter time in case a service is created on-top of an existing communication infrastructure / network slice (such condition/assumption need to be clarified in the relevant tests).

Practically, 5G-PPP efforts have focused on identifying the possible (mandatory or optional) parts/segments involved in the E2E service creation (such as network provisioning elements, MEC, NFVI elements, SDN and WAN domains of various technological, etc.) along with the time delay they introduce in the total "service creation time", prior to setting global timing targets. In the context of 5GENESIS, tests will span from simple to more complex network services creation in the various test facilities, and measurements can be obtained at the level of the Slice Management network element and/or at each specific domain (i.e. NFV, WAN, EMS). For the purpose of having a common reference of these tests within 5GENESIS, we define below the generic "Service Creation Time metric".

Metric	*Service Creation Time*
Metric definition *"Service creation time": The time required for the provision, deployment, configuration and activation of a full E2E communication service over a network slice, including all the physical and virtual components that are entailed in the Communication Service descriptors.*	

4.1.10. Speed

An inherent feature of MNs is their capability to provide network services to devices on the move. The QoS of the provisioned services however is highly affected by the mobile devices' velocity. The continuous increase in transportation means' velocity along with the advent of new applications and services, necessitate the provisioning of high QoS network services to mobile devices moving at continuously higher speeds. However, the achievability of the 5G-PPP KPI in terms of mobility speed, is a function of the network and application service, the network deployment and the specific location of the mobile device within the network footprint. Therefore, a number of tests are needed to evaluate it under different conditions. We define the generic "Speed metric" as follows.

Metric	*Speed*
Metric Definition *"Speed"*: The velocity of a UE to which a specific network service can be provisioned with a specific QoS.	

4.1.11. Throughput

Throughput is a KPI characterizing the capability of data transfer of a network or connectivity link under specific network conditions. This KPI is a factor of a vast number of parameters, conditions, and defined SUT; especially for 5G networks it is a slice-dependent performance indicator. Therefore, for the purpose of having a common reference of the tests measuring throughput in the context of 5GENESIS, the homonymous metric has been defined as follows.

Metric	*Throughput*
Metric Definition *"Throughput"*: data (payload) successfully transferred within a given time period from a data source to a data sink.	

4.1.12. Ubiquity

"Ubiquity", as a 5G-PPP KPI, refers to the capability of an operational or commercial network deployment to deliver to a number of users/devices, or/and to a number/extent of locations a specific service with a specific QoS (e.g., data rate, latency, PL rate) with a specified reliability, under specific conditions. "Ubiquity" is often used as a term by regulatory bodies to refer to this set of service provisioning characteristics/extent, and is defined per case (regulator/country/area, etc.). Such definitions of "Ubiquity KPI" can be found in [74][75][76][77][78].

It becomes prominent that ubiquity is a pure deployment-related KPI and that there is no single test to evaluate the generic 5G-PPP KPI ubiquity. Instead aspects of the KPI such as QoS vs. coverage for specific service, QoS vs number of users, coverage at hotspots with specific access network nodes' deployment, etc., can be evaluated through a number of equipment and deployment-related tests. For the purpose of having a common reference of these tests addressing such aspects, we define below the generic "ubiquity metric".

Metric	*Ubiquity*
Metric Definition *"Ubiquity":* The capability of an operational/commercial network deployment to deliver to a number of users/devices, or/and to a number/extent of locations a specific service with a specific QoS (e.g., data rate, latency, PL rate) and with specified reliability, under specific operational conditions.	

4.1.13. Mission critical push-to-talk (MCPTT)

Metric	*MCPTT*
Even though *MCPTT* is not a metric per se, it is defined as the metric umbrella that gathers all the 3GPP standardized MCPTT (voice communication signalling and data) delay between the transmission and the reception of a data packet over the SUT.	

4.1.14. Network management CAPEX/OPEX

One of the high-level objectives of the market driving the "5G networks" advancements is the reduction of the costs generally associated with the provisioning of telecommunication services at a specific QoS. This goal is determined by a number of techno-economic factors, which can be listed as (not limited to): network equipment (referring to various network physical/logical/functional segments) costs (CAPEX/OPEX related), network deployment specificities (including area topology, scale of deployment (associated to economies of scale), existing infrastructure), market specificities (affecting among others, network deployment and services requirements), service/QoS provisioning policies followed by network/service operators, and so on. Therefore, the CAPEX/OPEX reduction is a KPI to be evaluated by means of a techno-economic study; thus it is out of the scopy of technical test cases.

4.2. Application Level Metrics

4.2.1. Video streaming jitter

Metric	VideoStreamJitter
Metric Definition *"Video Stream Jitter":* The mean deviation of the difference in packet spacing at the receiver compared to the sender for a pair of packets	

5. TARGETED METRICS AND TEST CASES

This section gives the detailed descriptions of the test cases in sufficient detail to allow conducting experiments with measurements that can be meaningfully compared with future measurement results. The test cases were designed to provide numeric measurement results (where applicable), fine grained testing and automatic testability.

Numeric measurement results

Where possible, test should provide information about best/worst/average results of a test, as well as stating the precision of the measurement results by giving 95% confidence interval.

While KPIs in most cases require just a single value for comparison with a threshold level (e.g. latency < 10 ms E2E), the project benefits from having numeric measurements available to:

- Determine the margin with which KPI successes are achieved.
- Allow the derivation of combined KPIs (e.g. have different latency requirements for stationary and high-speed receivers).

Based on the numeric results, a simple pass/fail metric on KPIs will be subsequently derived in the project.

Fine grained testing

Tests are designed, where possible, to allow the determination of the contribution of individual elements of the platform to the resulting metric. For this, the test cases include calibration tests, employing minimal configurations (e.g., sender and receiver being on the same physical host or running on an "empty" SUT). Purpose of such tests is to provide the ability to distinguish which component in a platform is the most critical one, in achieving a specific KPI.

Automatic testing

All tests should be capable of being run automatically, without requiring user configuration or input to allow subsequent integration into an automated test suite, which then, combined with testing for KPI achievement, can later be utilized for testing, development and potentially certification of 5G components. Table 8 shows the planning of assessing specific KPIs at each 5GENESIS platform during the project lifetime..

Table 8 Assessment of KPIs per platform.

Metric	Malaga Platform	Athens Platform	Limassol Platform	Surrey Platform	Berlin Platform
Latency	x	x	x	x	
Capacity	x				
Ubiquity			x	x	
Speed	x	x			x
Reliability	x		x	x	x
Density of users	x			x	x
Location accuracy	x				
Energy efficiency				x	
Service creation time	x	x	x	x	x

The numbering of the test cases is according to: TC-<Metric-Acronym>-<num>, where

- <Metric-Acronym> refers to the acronym of the metric under consideration within the test case, i.e.:
  <Metric-Acronym> := <Latency> | Cap | Ubi | Rel | Den | Loc | Ene | Ser | Net
    - <Latency> := Rtt | Del
    - Rtt – a test case related to RTT
    - Del – a test case related to delay
    - Cap – a test case related to capacity
    - Ubi – a test case related to ubiquity
    - Rel – a test case related to reliability
    - Den – a test case related to density of users
    - Loc – a test case related to location accuracy
    - Ene - a test case related to energy efficiency
    - Ser – a test case related to service creation time
    - Net – a test case related to Network management CAPEX/OPEX
- <num> refers to the enumeration / numbering of test cases for the metric <Metric-Acronym>. <num> is represented with three digits.

5.1. Baseline tests

5.1.1. Capacity tests

The general metric of capacity described in Section 4.1.1 refers to operational network deployments. Instead, considering the 5GENESIS Platforms' deployments, the Capacity KPI can be measured as "*the total access network capacity of a single gNB over its corresponding radio-coverage area*". The capacity is measured for both the uplink and the downlink. The coverage area will be calculated based on the 5G UE' sensitivity (average values or based on the radio equipment available at each platform).

At this point, it shall be noted that access network node capacity is a function of the available bandwidth and the average spectral efficiency. This is, the number of correctly received bits over a certain period of time divided by the channel bandwidth of a specific band divided by the number of Transmission-Reception Points (TRxPs), and it is measured in bit/s/Hz/TRxP [8].

Therefore, the evaluation of this KPI (with this metric) will include analytical methods or system-level simulations for the definition of the actual target per deployment, as described in [5]-[7] (and the comparison of the measured values from 5GENESIS Platforms against them), and the extrapolation of values measured from the 5GENESIS Platforms to other deployment configurations.

With similar restrictions and influencing factors, the Capacity KPI can be also measured by "the total access network capacity of a max. number of gNB that can be deployed over the corresponding total radio-coverage area".

<table>
<tr><th colspan="2">Test Case</th><th>TC-Cap-001</th><th>Capacity</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Area traffic capacity
The KPI refers to the total traffic throughput served per geographic area (Mbps/m^2)[1] and is a measure of how much traffic a network can carry per unit area. It depends on the site density, bandwidth and average spectral efficiency.</td></tr>
<tr><td>2</td><td colspan="3">Methodology
In the ITU-R R M.2410-0 Report the area traffic capacity is calculated taking into account the number of correctly received bits, i.e. the number of bits contained in the SDUs delivered to Layer 3, over a certain period of time. This can be derived for a particular use case (or deployment scenario) considering one frequency band and one Transmission Reception Point (TRxP) layer, based on the (i) achievable average spectral efficiency; (ii) network deployment, e.g., TRxP (site) density, and (iii) bandwidth.
The TRxP is an antenna array with one or more antenna elements available to the network located at a specific geographical location for a specific area[2].
Let W denote the channel bandwidth and ρ the TRxP density (TRxP/m^2). The Area Traffic capacity C_{area} is related to the average spectral efficiency SE_{avg} through the following equation:
$$C_{area} = \varrho(\frac{TRxP}{m^2}) \times W(Hz) \times SE_{avg}(\mathrm{Bps/Hz/TRxP})$$
TRxP density ρ is the Number of TRxPs divided by the Area (m^2) over which the experimenter calculates the traffic capacity.
The average spectral efficiency is the aggregate throughput of all users (the number of correctly received bits, i.e. the number of bits contained in the SDU delivered to Layer 3, over a certain period of time) divided by the channel bandwidth of a specific band divided by the number of TRxPs. It is measured in bits/s/Hz/TRxP.
The channel bandwidth for this purpose is defined as the effective bandwidth normalized appropriately considering the uplink/downlink ratio. The effective bandwidth is defined as:
$$BW_{eff} = BW \times TR$$
where BW is the occupied channel bandwidth and TR is the time ratio of the link.
In FDD systems, TR equals 1, while in TDD systems it depends on the downlink/uplink configuration.
Let $R_i(T)$ denote the number of correctly received bits by user i (downlink) or from user i (uplink) in a system comprising a user population of N users and M TRxPs. Furthermore, let W denote the channel bandwidth and T the time over which the data bits are received. The average spectral efficiency is defined according to the following equation:
$$SE_{avg} = \frac{\sum_{i=1}^{N} R\_i(T)}{T \cdot BW_{eff} \cdot M}$$</td></tr>
</table>

---

[1] ITU-R M.2410-0 Minimum requirements related to technical performance for IMT-2020 radio interface(s), 2017

[2] 3GPP TR 38.913 version 14.2.0 Release 14 5G; Study on Scenarios and Requirements for Next Generation Access Technologies

Based on the definitions above, the calculation of the area traffic capacity shall comprise the following steps:

1) Measure the aggregate throughput of all users on the PDCP Layer of the eNB. If this is not possible, then measure the aggregate throughput on the S1-U interface of the EPC.
2) Use the aggregate throughput in order to calculate the Average Spectral Efficiency.
3) Use the Average Spectral Efficiency to calculate the Area Traffic Capacity.

Two cases are perceived for the creation of traffic in order to experimentally evaluate the area traffic capacity: (i) full buffer where traffic is generated and injected in the system using UDP protocol (ii) non-full buffer when TCP protocol is used for the generation of traffic.

The simplified approach to be followed involves only downlink capacity. A single TRxP is defined in open space conditions. For the evaluation, a full 5G deployment is used, comprising a 5G UE, a 5G NR gNB and a 5G Core. A traffic generator capable of producing UDP or TCP traffic is used behind the N6 interface of the 5G system and a traffic sink that collects traffic is used at the UE side.

The generated traffic profile will be according to TD-002 and the measurement duration shall be 120s per iteration out of a total of 5 iterations.

3

Calculation process and output

The calculation process includes the following steps:

Let R̅(T) be the average aggregate throughput measured over a time interval T, and R_n(T) be the average aggregate throughput measured at iteration n over a time interval T.

The reported Average Aggregate Throughput on the PDCP Layer on the eNB/gNB shall be calculated as follows:

$$Average\ Aggregate\ Throughput\ \bar{R}(T) = \frac{1}{N}\sum_{n=1}^{N} R\_n(T)$$

This value is used to calculate the Average Spectral Efficiency, leading to the Area Traffic Capacity.

The necessary calculations are shown in the following table:

Parameters	Formula
Effective Bandwidth (Hz)	$BW_{eff} = BW \times TR$
Average Aggregate Throughput (Mbps)	$\bar{R}(T) = \frac{1}{N}\sum_{n=1}^{N} R\_n(T)$
Average Spectral Efficiency(bit/s/Hz/TRxP)	$SE_{avg} = \frac{\bar{R}(T)}{BW_{eff} \times Number\ of\ Area\ TRxPs}$
Area (m^2)	<Depends on the Area geometry>
Site density (TRxP/m^2)	$\rho = \frac{Number\ of\ \mathrm{Area}\ TRxPs}{Area}$

<table>
<tr><td></td><td><table><tr><td>Estimated average area traffic capacity (Mbps/m^2)</td><td>$$C_{avg} = \rho \times W \times SE\_avg$$</td></tr></table></td></tr>
<tr><td>4</td><td>Complementary measurements

<i>The experimenter should record the following complementary metrics during the measurement:</i>

<i>At the UE:</i>
<i>DL RSRP</i>
<i>DL RSRQ</i>
<i>SINR</i>
<i>At the eNB/gNB:</i>
<i>Uplink Signal Strength</i>
<i>CQI reported from UE</i></td></tr>
<tr><td>5</td><td><i>Pre-conditions</i>

Prior to the beginning of the tests case:
The experimenter will define the area of interest and the UE locations for performing the measurements.
The UE will be placed in various locations within the cell, including locations at the edge of the area (lowest possible SNR).</td></tr>
<tr><td>6</td><td>Applicability

N/A</td></tr>
<tr><td>7</td><td>Test Case Sequence
<i>Start recording the complementary metrics.</i>
<i>Initiate the traffic generator in the Downlink, using UDP traffic.</i>
<i>Set the duration of the measurement to 120s.</i>
<i>Upon completing the measurement, record the average throughput reported on the PDCP layer on the eNB.</i>
<i>Stop recording the complementary metrics.</i>
<i>Repeat Steps 1-5 in case of multiple iterations.</i>
<i>Repeat steps 1-6 for TCP traffic</i>
<i>Compute the area traffic capacity for each traffic case (UDP/TCP), as defined in the section “Calculation Process and Output”.</i></td></tr>
</table>

5.1.1.1. Capacity calibration tests

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Cap-002</i></th><th><i>Capacity Calibration</i></th></tr>
<tr><td>1</td><td colspan="3"><i>Target KPI</i>
<i>Capacity</i>
The Capacity calibration test aims at verifying the proper operation of the MN, before performing the measurements.</td></tr>
<tr><td>2</td><td colspan="3"><i>Methodology</i>
The calibration procedure of the Capacity KPI consists of verifying data connectivity between the UE and the N6 interface of the 5G Network.
The generated traffic profile shall be according to TD-002.
The measurement probe at the UE will issue the ICMP ECHO_REQUEST.
One iteration will include 100 consecutive requests.
The total number of iterations is set to 5.
The verification of the data connectivity shall be performed at the edge of the area of interest (lowest possible SNR).:</td></tr>
<tr><td>3</td><td colspan="3"><i>Calculation process and output</i>
The experimenter will calculate the average ICMP round trip.
The required output should be reported, as follows:
<table>
<tr><th colspan="2">RTT(ms)</th></tr>
<tr><td>Mean</td><td>$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$</td></tr>
</table></td></tr>
<tr><td>4</td><td colspan="3"><i>Complementary measurements</i>
<i>The experimenter will record the following complementary metrics during the measurement:</i>
<i>At the UE:</i>
<i>DL RSRP</i>
<i>DL RSRQ</i>
<i>SINR</i>
<i>PL Rate</i>
<i>At the eNB/gNB:</i>
<i>Uplink Signal Strength</i>
<i>CQI reported from UE</i>
For each one of these metrics, use the following methodology:
Average per iteration
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Average</td></tr>
</table>

	$$avg = \frac{1}{i}\sum_{i} avg_i$$
5	*Pre-conditions* Prior to the beginning of the tests case: - The experimenter will define the area of interest and the UE locations for performing the measurements. - In case of multiple UEs, they will be placed in various locations within the cell, including locations at the edge of the area (lowest possible SNR).
6	*Applicability* N/A
7	*Test Case Sequence* 1. *Start the monitoring probes.* 2. *Begin pinging from the UE to the N6 interface using the defined ICMP Traffic Profile.* 3. *Record the RTT.* 4. *Stop monitoring probes.* 5. *Calculate and record the average RTT and the averages of the complementary metrics, as described in “Complementary Metrics” per iteration.* 6. *Repeat Steps 1-5 for each one of the 5 iterations.* 7. *Compute the RTT as described in “Calculation Process and Output”.* 8. *Assess whether RTT and PL are within acceptable limits.*

5.1.2. Density of user tests

The fact that “density” is in general understood as “something that is measured” in relation to the area the measurement relates to – i.e. here, for density of users, giving results in the SI unit 1/m^2) – imposes a special challenge for a test case: While the number of users served by a 5G-core over one or multiple base stations (BS) can be easily obtained as an operational parameter out of the packet core, quantizing the area the measurement relates to is not easily assessible. The coverage area of a single BS or small cell could be quantized, e.g.:

- by measuring the signal reception power at a several locations and then mapping them to an assumed availability of a service with specified QoS Class Identifier (QCI), thus obtaining the coverage area,
- via the theoretical coverage via propagation models or “best known results” from deployment experience,
- by quantizing the area that the users under concern within a measurement campaign reside in, acknowledging the fact that the actual coverage of the related radio cell may be larger.

To address this complexity within the measurement campaigns, 5GENESIS decided to define a set of test cases, each addressing either the number of devices served by a packet core, by a single or multiple gNBs, the number of gNBs deployable in a given region, or gauging the lower bound of geographical coverage for a given deployment. The resulting measurements hence allow to directly assess the density of users – in case the coverage area under concern is measured during an experiment - or to derive the density of users - in case the coverage area under concern is estimated. Table 9 provides examples for KPIs, which can be directly assessed in a test case and primarily measured.

Table 9: Assessing the density of users.

Metric	*Density of Users*
Number of devices that can be registered/simultaneously served (with traffic of specific QCI) per 5G-core element	#units
Max. number of devices that can be registered/simultaneously served (with traffic of specific QCI) in a 5G-Core solution (even if this requires dimensioning of one or more components of it)	#units
Max. number of devices that can be registered/simultaneously served (with traffic of specific QCI) by a single gNB	#units
Max. number of devices that can be registered/simultaneously served (with traffic of specific QCI) by the total number of gNBs that can be connected to a 5G-Core solution	#units
Max. number of gNB that can be deployed over the corresponding total radio-coverage area.	#units/m^2
Max. number of devices that can be registered/simultaneously served in a 5G-Core providing network service (via gNBs or n3GPP XS technologies) for a well-known geographical region.	#units/m^2

5.1.2.1. Density of user calibration tests

Maximum number of devices registered per Packet-core

<table>
<tr><th colspan="2">Test Case</th><th>*TC-Den-001*</th><th>*Density of Users*</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Maximum number of devices registered per Packet Core
This KPI refers to the maximum number of devices (UEs), which a packet core can support. It hence provides information on the upper bound of the density of users per unit area as regardless of how the coverage area is optimized within a given system, the core cannot support more users than obtained via this measurement.
This test case isolates the potential impact of the behaviour of the UEs or BSs on the results; the SUT is the pure packet core and the measurement tools used to communicate with the core replicate the interface between the packet core and a BS.
Source A → Test tool / instrument emulating UE, gNB behaviour towards the packet core (N:1/N:2 interface)
Destination B → AMF of packet core
Underlying system → 5G packet core</td></tr>
<tr><td>2</td><td colspan="3">Methodology
To measure the maximum number of devices that can register per packet core, a set of consecutive experiments are run. In each experiment the number of UEs registering at the packet core is</td></tr>
</table>

<table>
<tr><td></td><td>constantly increased, up to the point at which the registration process fails or takes longer than a predefined, set upper time limit. Then, the maximum number of devices that can register is given by the maximum of the number of UEs in the set of experiments for which the registration is successful and completed before the predefined, set upper time limit.

Hence, a single iteration consists of several consecutive steps. They are characterized by an increasing number of UE registering to the packet core. Several replica of an iteration shall be conducted to gain confidence values for reported results.

The test case shall include the consecutive execution of several iterations according to the following properties.

Duration of a single iteration → given by termination criteria, i.e.: terminate iteration when (a) at least one registration attempt within a single step of the iteration fails or (b) when the time required to conduct all scheduled registrations within a step exceed a given threshold.
Increase between consecutive steps of number of UEs attempting to register → 50
Operations per second (target attachment rate) → R = 100 Hz
Timeout value for a step within an iteration → 1.5 * max(100 ms*U , U/R), where U is the number of UEs to attach / register within the given step, and R is the attachment rate.
Number of replica (iterations) → At least 25 iterations</td></tr>
<tr><td>3</td><td>*Calculation process and output*

The required output should be calculated according to the following methodology:

Let max_i be the maximum number of devices that can be registered as observed in the i^{th} iteration; and let x_(i,n) be the number of UEs in the i^{th} step within a trial as such that all registration succeed and as such that all registration complete the timeout for that trial; and let x_(i+1,n) be the number of UEs in the $i+1^{th}$ step within a trial as such that at least one registration fails or as such that the registrations take more than the timeout to complete, then max_i is given by:

$$max_i = x_{i,n}$$

Then, the overall (reported) maximum number of devices that can be registered max shall be calculated as the average of all max_i:

$$max = \frac{1}{i}\sum_i max_i$$

For the overall reported maximum number of devices that can register max, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

The output should be provided as:

<table>
<tr><td colspan="3">number of devices that can be registered at a packet core</td></tr>
<tr><td rowspan="2">max</td><td colspan="2">95% confidence interval for Mean</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table></td></tr>
</table>

4	Complementary measurements Average duration of UE registration time
5	Pre-conditions A deployed and working 5G packet core (SUT) to which the testing tool (emulating UE and gNB behaviour) may connect to. No registered UEs in the system.
6	Applicability This test case applies for all scenarios that evaluate the performance of a packet core.
7	Test Case Sequence 1. Assure the precondition is met, i.e. there are no registered UE in the system 2. Set the test tool to emulate U number of UE, which attempt to register at the packet core, to 50 (Initial value). 3. set the test tool to run consecutive registration requests of the UE towards the packet core at a rate of R=100 Hz. 4. Report the time to complete step (2). 5. if (a) all registrations succeeded and (b) if the recoded time is less than the set timeout value. i. Increase by 50 the number of UE in the test tool, which attempt to register at the packet core, and ii. Deregister all UE at the packet core and assure that no UE is registered in the core. iii. Repeat this test sequence from step (3) onwards 6. Deregister all UE and terminate the iteration. 7. Repeat steps 1 to 6 for each one of the 25 iterations. 8. Compute the KPIs as defined in section "Calculation process and output".

Maximum number of active devices per Packet-core

Test Case	*TC-Den-002*	*Density of Users*
1	Target KPI **Maximum number of active devices per Packet Core** This KPI refers to the maximum number of active devices (UE), which a packet core can support. It hence provides information on the upper bound of the density of users per unit area under the constraint of considering only active users, as regardless of how the coverage area is optimized within a given system, the core cannot support more simultaneously active users than obtained via this measurement. This test case is designed to only focus on the maximum number of active users a packet core can support, and not as a stress test in terms of experiencing a high signalling rate of UE going from idle into active mode. This test case isolates potential impact of the behaviour of UE or BS on the results; the system under test is the pure packet core and the measurement tools used to communicate with the core replicate the interface between the packet core and a BS. Source A → Test tool / instrument emulating UE, gNB behaviour towards the packet core (N:1/N:2 interface)	

	Destination B → AMF of packet core Underlying system → 5G packet core
2	Methodology To measure the maximum number of active devices per packet core, a set of consecutive experiments are run, as such that in each experiment the number of UE going from idle into active mode is constantly increased. The number of UE switching into active mode is increased up to the point at which the signalling fails or takes longer than a predefined, set upper time limit. Then, the maximum number of active devices that a packet core can support is given by the maximum number of UE in the set of experiments for which the transition from idle into active mode is successful and completes before the predefined, set upper time limit. Hence, a single iteration of the experiment consists of several consecutive steps being characterized by an increasing number of UE requesting the packet core to switch from idle into active mode. Several replica of an iteration shall be conducted to gain confidence values for reported results. The test case shall include the consecutive execution of several iterations according to the following properties. • Duration of a single iteration → given by termination criteria, i.e.: terminate iteration when (a) at least one UE is denied to switch into active mode, or the signalling fails within a single step of the iteration or (b) when the time required to conduct all scheduled registrations within a step exceed a given threshold. • Increase between consecutive steps of number of UEs attempting to switch into active mode → 50 • Operations per second (target attachment rate) → R = 100 Hz • Timeout value for a step within an iteration → 1.5 * max(100ms*U , U/R), where U is the number of UEs to attach / register within the given step, and R is the rate at which UEs request to switch into active mode. • Number of replica (iterations) → At least 25 iterations
3	*Calculation process and output* The required output should be calculated according to the following methodology: Let max_i be the maximum number of devices that can be simultaneously active as observed in the ith iteration; and let x_(i,n) be the number of UE in the ith step within a trial as such that all UE successfully transfer into active mode and as such that all UE activations complete the timeout for that trial; and let x_(i+1,n) be the number of UE in the i+1th step within a trial as such that at least one request of a UE to switch into active mode fails or as such that the registrations take more than the timeout to complete, then *max_i* is given by: $$max_i = x_{i,n}$$ Then, the overall (reported) maximum number of active devices that a packet core can support *max* shall be calculated as the average of all *max_i*: $$max = \frac{1}{i}\sum_{i} max_i$$

<table>
<tr><td></td><td>
<p>For the overall reported maximum number of active devices supported by a packet core, <i>max</i>, the 95 % confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.</p>
<p>Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].</p>
<p>The output should be provided as:</p>
<table>
<tr><th colspan="3">number of active devices that can be supported by a packet core</th></tr>
<tr><td rowspan="2">max</td><td colspan="2">95% confidence interval for Mean</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>
<p>Complementary measurements</p>
Average duration of UE activation time
</td></tr>
<tr><td>5</td><td>
<p>Pre-conditions</p>
<p>A deployed and working 5G packet core (SUT) to which the testing tool (emulating UE and gNB behaviour) may connect to.</p>
<p>A number of UE are registered at packet core, but are idle. This number shall be set to the maximum number of registered UE by a packet core as identified by test case <i>TC-Den-001</i>.</p>
<p>Zero active UE in the packet core.</p>
</td></tr>
<tr><td>6</td><td>
<p>Applicability</p>
<p>This test case applies for all scenarios that evaluate the performance of a packet core.</p>
</td></tr>
<tr><td>7</td><td>
<p>Test Case Sequence</p>

Assure the precondition is met, i.e. there is no active UE in the system and there are a number of registered UE in the system, where the number is the maximum number of registered UE
Set the test tool to emulate U number of UE, which attempt to transition into active mode, to 50 (Initial value)
set the test tool to run consecutive activation requests of the UE towards the packet core at a rate of R=100Hz
Report the time to complete step (2)
if (a) all activations succeeded and (b) if the recoded time is less than the set timeout value.

Increase by 50 the number of UE in the test tool, which attempt to become active, and
Deactivate all UEs at the packet core and assure that the preconditions are met
Repeat this test sequence from step (3) onwards

Deregister all UE and terminate the iteration
Repeat steps 1 to 6 for each one of the 25 iterations
Compute the KPIs as defined in section “Calculation process and output”

</td></tr>
</table>

5.1.3. Energy efficiency tests

5.1.3.1. RAN Energy efficiency tests

Maximum (peak) energy efficiency

<table>
<tr><th colspan="2">Test Case</th><th>TC-Ene-001</th><th>Energy Efficiency</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
RAN Energy Efficiency
This KPI aims to measure the RAN EE and represents the efficiency with which each Joule of energy is used to transmit information. This KPI expression represents the data volume of the BSs under consideration, divided by the total EC of the BS sites (including the support infrastructure).</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For the evaluation, a full 5G deployment composed of a 5G UE (or UE emulator), a 5G NR gNB and 5G core is used. A traffic generator capable of producing UDP or TCP traffic is used behind the N6 interface of 5G system and a traffic sink that collects traffic is used at the UE side.
The generated traffic uses full length IP packets (i.e. 1500 bytes) and the baseline measurement duration is 24 hours (experiment repeated for 7 days). The data volume is collected via NMS counters and EC through measurements (using watt meters or from utility provider).
The EE can be measure based on the following KPIs:
• $KPI_{EE\text{-}capacity}$ (used to measure EE, on the basis of EC in relation to capacity),
• $KPI_{EE\text{-}site}$ (used to measure the EE of the support infrastructures of the site as compared to the consumption of the BS(s) of the site).
The KPIs are applicable to all stages of network utilization. However, it has to be recognized that as the BS utilization increases:
• $KPI_{EE\text{-}capacity}$ will increase, since the BS equipment operates more efficiently at higher load levels
• $KPI_{EE\text{-}site}$ will increase.
The $KPI_{EE\text{-}capacity}$ ($EE_{MN,DV}$), expressed in bit/J, is defined as the ratio between the Data Volume (DV_{MN}) and the EC (EC_{MN}), in the MN:
$$EE_{MN,DV} = \frac{DV_{MN}}{EC_{MN}}$$
As the KPI is measured in unit of “bits/Joule” [equivalent to bits/Watt hour (Wh)] it represents the efficiency with which each Joule of energy is used to transmit information. This KPI expression represents the data volume of the BS over the backhaul network divided by the total EC of the BS site (including the support infrastructure). The EC_{MN} includes EC of each BS of the BS site as well as that of the support infrastructure of the BS site, during the measurement period. This KPI is used for MNs handling high data volumes, in particular in dense-urban, urban areas (i.e. capacity-limited deployments).
The $KPI_{EE\text{-}site,}$ denoted as SEE, expressed in “Wh”, is an additional network KPI describing the EC of the telecom equipment with reference to the total EC:
$$SEE = \frac{EC_{BSs}}{EC_{BSs}+EC_{SI}}$$</td></tr>
</table>

	EC_{BSs} represents EC of BSs under test site, and EC_{SI} represents EC of supporting infrastructure, during the measurement period.
	The SEE metric provides an INDICATION of SEE in terms of how big a fraction of total energy is used for actual telecom equipment (telecommunication service delivery). In other words, it provides the EC overhead incurred due to the BS site support infrastructure/equipment.
	The KPI definitions follow the recommendations in ITU-T L.1331 [42] and ETSI Standard ES 203 2283 [34], (which are technically equivalent), and describe the EC and MN EE measurements in operational networks.
	Note: the data vol. and EE measurements will be performed on a sub-network i.e. a selection of BS sites which constitutes *the partial MN under test*. The ETSI ES 203 228 [34] section 7, defines a method to extrapolate the measured EE KPIs of **the partial MN under test** to the operator's whole RAN.
	The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
	• Duration of a single replica (iteration) → at least 15 minutes
	  o Note, that there will be a max. of 24*60 / 15 = 96 iterations over a 24-hour period, therefore the min. # iterations over a single 24-hour period is set to 96/8 = 12.
	• Number of replica (iterations) → At least 12 x 7 = 84
	  o Note that "experimentation" period (for final results reporting) covers a period of 7-days.

3	*Calculation process and output* **EC Measurement** The MN EC (EC_{MN}) is the sum of the EC of equipment included in the MN under investigation. The network EC is measured according to the assessment process defined in section 6 of [33] such that individual metrics are provided per RAT and per MNO. **The overall EC of the partial MN under test** is measured as follows: $EC_{MN} = \sum_i (\sum_k EC_{BS_{i,k}} + EC_{SI_i}) + \sum_j EC_{BH_j} + \sum_l EC_{RC_l}$ where: • EC_{MN} is EC, in the MN under test, and is measured in Watts/hr.[4] (Wh = Joule), over the period of measurement T. • BS refers to the BSs in the MN under measurement. • BH is the backhauling providing connection to the BSs in the MN under measurement. • SI is the site infrastructure (Rectifier, battery losses, climate equipment, TMA, tower illumination, etc.). • RC is the control node(s), including all infrastructure of the RC site. • *i* is an index spanning over the number of sites. • *j* an index spanning over the number of BH equipment connected to the *i* sites. • *k* is the index spanning over the number of BSs in the *i*-th site.

---

[3] ITU-T L.1331/1330 and ETSI Standard ES 203 228 describes EC and MN EE measurements **in operational networks**, whilst power consumption and EE measurements of **individual MN elements** are described in several standards (e.g. ETSI ES 202 706 [44] for radio base stations).

[4] 1 Wh = 3·6 KJ

- l is the index spanning over the control nodes of the MN.

The EC of the various segments e.g. BS, BH, CR etc. can be measured by means of metering information provided by utility suppliers, COTS tools e.g. smart-meter plugs [79] or by MN integrated measurement systems. EC_{MN} is measured in unit of Wh (Watt Hours). Power consumption and EE measurements of individual MN elements are described in several standards e.g., ITU-T L.1310 [40] for radio base stations and ITU-T L.1320 [62] for power and cooling equipment. When a MN integrated measurement system according to ETSI ES 202 336-12 [52] is available, it should be used in addition to the utility provided EC allowing a more precise estimation of the consumption per RAT and per MNO [79].

Data Volume (DV) Measurement

The DV_{MN} shall be measured using network counters for data volume related to the aggregated traffic in the set of BSs considered in the MN under test.

For packet switched services, DV_{MN} is defined as the data volume delivered by the equipment of the partial MN under investigation during the time frame T of the EC assessment. The assessment process defined in section 6 shall be used:

$$DV_{MN-PS} = \sum_{i,k} DV_{BS_{i,k}-PS}$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T. *i* and *k* are defined in formula (1).

For circuit switched services[5] like voice, DV_{MN-CS} is defined as the data volume delivered by the equipment of the MN under investigation during the time frame T of the EC assessment:

$$DV_{MN-CS} = \sum_{i,k} DV_{BS_{i,k}-CS}$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T. *i* and *k* are like in formula (1).

The overall data volume is computed as follows:

$$DV_{MN} = DV_{MN-PS} + DV_{MN-CS}$$

DV_{MN} can be derived based on standard counters defined in ETSI TS 132 425 [34] for LTE (or 3GPP equivalent: TS 32.425), multiplying by the measurement duration T. DV_{MN} is computed in unit of bit.

For packet switch (PS) traffic, the data volume is considered as the overall amount of data transferred to and from the users present in the MN under test. Data volume shall be measured in an aggregated way per each RAT present in the MN and shall be measured referring to counters derived from vendor O&M systems.

For Circuit Switch (CS) traffic, the data volume is considered as the number of minutes of communications during the time T multiplied by the data rate of the corresponding service and the call success rate[6]. The call success rate is equal to 1 minus the sum of blocking and dropping rates, i.e.:

---

[5] Note that "circuit switched", refers to all voice, interactive services and video services managed by the MNOs, including CS voice and real-time video services delivered through dedicated bearers.

[6] Note that for CS traffic (e.g. VoLTE) in LTE RAT, there are no measurements defined in TS 32.425. for calculation of CS traffic refer to Table 4.4.3.2-2, in [51].

$$\mathit{Call\ Success\ Rate} = (1 - \mathit{dropping\ rate}) \times 100\ [\%] \tag{5}$$

The dropping includes the intra-cell call failure (rate of dropping calls due to all the causes not related to handover) and the handover failure:

$$1 - \mathbf{dropping\ rate} = (1 - \mathbf{intracell\ failure\ rate})(1 - \mathbf{handover\ failure\ rate}) \tag{6}$$

In order to include reliability in the measurement the aggregated data volume shall be provided together with the 95th percentile of the cumulative distribution, for each RAT in the MN.

For data reporting, templates available in ANNEX A in ETSI ES 203 228 [33] (or equivalents i.e. ANNEX I of Rec. ITU-T L.1331 [41] or in 3GPP TR 32.856 [51]) are used.

OUTPUTs:

1. The $KPI_{EE\text{-}capacity}$ (denoted as $EE_{MN,DV}$ below), expressed unit of “bits/Joule” [equivalent to bits/Watt hour (Wh)]., is calculated as the ratio between the Data Volume (DV_{MN}) and the Energy Consumption (EC_{MN}), in the mobile network (MN), during the 7-day measurement period:

$$EE_{MN,DV} = DV_{MN} / EC_{MN} \tag{7}$$

2. The $KPI_{EE\text{-}site}$ (denoted as SEE) expressed in unit of “Wh” [Watt hour], is calculated as the ratio of the energy consumption of the telecom equipment to the total energy consumption:

$$SEE = EC_{BSs}/(EC_{BSs} + EC_{SI}) \tag{8}$$

where, EC_{BSs} represents energy consumption of BSs under test site, and EC_{SI} represents energy consumption of supporting infrastructure, during the 7-day measurement period.

3. The **peak** (observed over the measurement period) values of $EE_{MN,DV}$ and SEE will be reported.

The required peak output should be calculated according to the following methodology:

Let EEmax_i be the maximum EE measured in the ith iteration, and x_(i,n) be the measured EE for each sinle 15-min. iteration:.

$$EEmax_i = \max(x_{i,n})$$

Then, the (reported) maximum EE (EEpeak) shall be calculated as follows:

$$EEpeak = \frac{1}{i}\sum_{i} EEmax_i$$

For the reported maximum EE (EEpeak), the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.

Peak RAN EE (EEpeak) output should be provided as:

EE [b/j]		
Max	95% confidence interval for Max	
	Lower bound	Upper bound

4 Complementary measurements

The following complementary metrics will be available & logged during the measurement period:

At the UE:

- *DL RSRP*

	- *DL RSRQ* - *SINR* *At the eNB/gNB:* - *Uplink Signal Strength* - *CQI reported from UE*
5	*Pre-conditions* The experimenter will define the following parameters of their setup: - Technology/Band - Transmission Mode (TDD/FDD) - Receive/Transmit Frequencies - Number of Antennas - Channel Bandwidth - UE Category - Service/slice configuration (eMBB, URLLC, etc.) Prior to the beginning of the tests: - The experimenter will define the (designated) area of interest, size of “partial network under test” [unit: # sites], number of UEs, and the UE locations (random deployment) - The UEs will be placed in various locations within the cell, including locations at the edge of the area (lowest SNR). There will be only one or more UE (or UE emulator) successfully connected to the network. Measurements shall be performed without any energy savings features to evaluate basic RAN energy efficiency.
6	Applicability The UE and eNB/gNB should provide monitoring of the Complementary metrics.
7	Test Case Sequence 1. Place UEs (operating in NSA mode) in a test location corresponding to good coverage. 2. Start recording the complementary metrics. 3. Initiate the traffic generator in the Downlink & Uplink, using UDP traffic. 4. Set the period of the measurement to 24 hrs. 5. Stop recording the complementary metrics at end of measurement period. 6. Upon completing the measurement, record the DV_{MN} based on NMS counters & eq. (1). 7. Upon completing the measurement, record the EC_{MN} from utility meters/watt meters & apply eq. (4). 8. Repeat Steps 1-6 for period of 7 days. 9. Compute the weekly EE KPI based on eq. (7) and (8). Extend to yearly, using extrapolation method in [x]. 10. Report the MN EE assessment results for 4G and 5G deployments separately, using the KPIs definition provided in section "Calculation process and output" as well as the templates provided in Annex 2.

Average (expected mean) energy efficiency

Test Case		TC-Ene-002	*Energy Efficiency*
1	Target KPI: **Energy Efficiency** This KPI aims to measure the RAN EE, and represents the efficiency with which each Joule of energy is used to transmit information. This KPI expression represents the data volume of the BSs under consideration, divided by the total EE of the BS sites (including the support infrastructure).		
2	Methodology For the evaluation, a full 5G deployment composed of a 5G UE (or UE emulator), a 5G NR gNB and 5G core is used. A traffic generator capable of producing UDP or TCP traffic is used behind the N6 interface of 5G system and a traffic sink that collects traffic is used at the UE side. The generated traffic is using full length IP packets (i.e. 1500bytes) and the baseline measurement duration is 24 hours (experiment repeated for 7-days). The data volume will be collected via NMS counters and energy consumption through measurements (using watt meters or from utility provider). The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 15 minutes o Note, that there will be a max. of 24*60 / 15 = 96 iterations over a 24-hour period, therefore the min. # iterations over a single 24-hour period is set to 96/8 = 12. • Number of replica (iterations) → At least 12 x 7 = 84 o Note that "experimentation" period (for final results reporting) covers a period of 7-days.		
3	*Calculation process and output* **EC Measurement** The MN EC (EC_{MN}) is the sum of the EC of equipment included in the MN under investigation. The network ECZ is measured according to the assessment process defined in section 6 of [33] such that individual metrics are provided per RAT and per MNO. **The overall EC of the partial MN under test** is measured as follows: $EC_{MN} = \sum_i(\sum_k EC_{BS_{i,k}} + EC_{SI_i}) + \sum_j EC_{BH_j} + \sum_l EC_{RC_l}$ (1) where: • EC_{MN} is EC, in the MN under test, and is measured in Watts/hr.[7] (Wh = Joule), over the period of measurement T. • BS refers to the BSs in the MN under measurement. • BH is the backhauling providing connection to the BSs in the MN under measurement. • SI is the site infrastructure (Rectifier, battery losses, climate equipment, TMA, tower illumination, etc.). • RC is the control node(s), including all infrastructure of the RC site. • *i* is an index spanning over the number of sites. • *j* an index spanning over the number of BH equipment connected to the *i* sites.		

---

[7] 1 Wh = 3·6 KJ

- *k* is the index spanning over the number of BSs in the *i*-th site.
- l is the index spanning over the control nodes of the MN.

The EC of the various segments e.g. BS, BH, CR etc. can be measured by means of metering information provided by utility suppliers or by mobile network integrated measurement systems. EC_{MN} is measured in unit of Wh (Watt Hours). Power consumption and EE measurements of individual MN elements are described in several standards e.g., ITU-T L.1310 [40] for radio base stations and ITU-T L.1320 [62] for power and cooling equipment. When a mobile network integrated measurement system according to ETSI ES 202 336-12 [52] is available, it should be used in addition to the utility provided EC allowing a more precise estimation of the consumption per RAT and per MNO.

DV Measurement

The DV_{MN} shall be measured using network counters for data volume related to the aggregated traffic in the set of BSs considered in the MN under test.

For PS services, DV_{MN} is defined as the DV delivered by the equipment of the partial MN under investigation during the time frame T of the EC assessment. The assessment process defined in section 6 shall be used:

$$DV_{MN-PS} = \sum_{i,k} DV_{BS_{i,k}-PS} \tag{2}$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T. *i* and *k* are defined in formula (1).

For CS services[8] like voice, $DV_{MN\text{-}CS}$ is defined as the DV delivered by the equipment of the MN under investigation during the time frame T of the EC assessment:

$$DV_{MN-CS} = \sum_{i,k} DV_{BS_{i,k}-CS} \tag{3}$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T. *i* and *k* are like in formula (1).

The overall data volume is computed as follows:

$$DV_{MN} = DV_{MN-PS} + DV_{MN-CS} \tag{4}$$

DV_{MN} can be derived based on standard counters defined in ETSI TS 132 425 [34] for LTE (or 3GPP equivalent: TS 32.425), multiplying by the measurement duration T. DV_{MN} is computed in unit of bit.

For PS traffic, the DV is considered as the overall amount of data transferred to and from the users present in the MN under test. DV shall be measured in an aggregated way per each RAT present in the MN and shall be measured referring to counters derived from vendor O&M systems.

For CS traffic, the DV is considered as the number of minutes of communications during the time T multiplied by the data rate of the corresponding service and the call success rate[9]. The call success rate is equal to 1 minus the sum of blocking and dropping rates, i.e.:

$$Call\ Success\ Rate = (1 - dropping\ rate) \times 100\ [\%] \tag{5}$$

The dropping includes the intra-cell call failure (rate of dropping calls due to all the causes not related to handover) and the handover failure:

---

[8] Note that "circuit switched", refers to all voice, interactive services and video services managed by the MNOs, including CS voice and real-time video services delivered through dedicated bearers.

[9] Note that for CS traffic in LTE RAT, there are no measurements defined in TS 32.425. for calculation of CS traffic refer to Table 4.4.3.2-2, in [51].

$$1-\text{dropping rate}=(1-\text{intracell failure rate})(1-\text{handover failure rate}) \tag{6}$$

In order to include reliability in the measurement the aggregated DV shall be provided together with the 95th percentile of the cumulative distribution, for each RAT in the MN.

For data reporting, templates available in ANNEX A in ETSI ES 203 228 [33] (or equivalents i.e. ANNEX I of Rec. ITU-T L.1331 [41] or in 3GPP TR 32.856 [51]) are used.

OUTPUTs:

4. The $KPI_{EE\text{-}capacity}$ ($EE_{MN,DV}$), expressed unit of "bits/Joule" [equivalent to bits/Watt hour (Wh)]., is calculated as the ratio between the DV (DV_{MN}) and the EC (EC_{MN}), in the MN, during the 7-day measurement period:

$$EE_{MN,\,DV} = DV_{MN} / EC_{MN} \tag{7}$$

5. The $KPI_{EE\text{-}site}$ (denoted as SEE) expressed in unit of "Wh" [Watt hour], is calculated as the ratio of the EC of the telecom equipment to the total EC:

$$SEE = EC_{BSs}/(EC_{BSs} + EC_{SI}) \tag{8}$$

Where, EC_{BSs} represents EC of BSs under test site, and EC_{SI} represents EC of supporting infrastructure, during the 7-day measurement period.

6. The **mean** (observed over the measurement period) values of $EE_{MN,DV}$ and SEE,will be reported.

The required output should be calculated according to the following methodology:

Let EEavg_i be the calculated average EE for the i^{th} iteration, and x_(i,n) be the measured EE for each iteration:

$$EEavg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average EE*mean* shall be calculated as the average of all x_i

$$EEmean = \frac{1}{i}\sum_{i} EEavg_i$$

For the overall average EE (EEmean), the 95 % confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

EEmean output should be provided as:

EE [b/j]		
Mean	95% confidence interval for Mean	
	Lower bound	Upper bound

4 Complementary measurements

The following complementary metrics will be available and logged during the measurement period:

At the UE:

- *DL RSRP*
- *DL RSRQ*
- *SINR*

	At the eNB/gNB: • *Uplink Signal Strength* • *CQI reported from UE*
5	*Pre-conditions* The experimenter will define the following parameters of each setup: • Technology/Band • Transmission Mode (TDD/FDD) • Receive/Transmit Frequencies • Number of Antennas • Channel Bandwidth • UE Category • Service/slice configuration (eMBB, URLLC etc.) Prior to the beginning of the tests: • The experimenter will define the (designated) area of interest, size of “partial network under test” [unit: # sites], number of UE, and the UE locations (random deployment) • The UE will be placed in various locations within the cell, including locations at the edge of the area (lowest SNR). • There will be only one or more UEs (or UE emulator) successfully attached to the network.
6	Applicability The UE and eNB/gNB should provide monitoring of the Complementary metrics.
7	Test Case Sequence *1. Start recording the complementary metrics.* *2. Initiate the traffic generator in the Downlink and Uplink, using UDP traffic.* *3. Set the period of the measurement to 24 hrs.* *4. Stop recording the complementary metrics at end of measurement period.* *5. Upon completing the measurement, record the* DV_{MN} *based on NMS counters & eq. (1).* *6. Upon completing the measurement, record the* EC_{MN} *& apply eq. (4).* *7. Repeat Steps 1-6 for period of 7 days.* *8. Compute the weekly EE KPI based on eq. (7) and (8). Extend to yearly, using extrapolation method.* *9. Report the MN EE assessment results for 4G and 5G deployments separately, using the KPIs definition provided in section "Calculation process and output" as well as the templates provided in Annex 2*

5.1.3.2. UE Energy efficiency tests

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Ene-003</i></th><th><i>Energy Efficiency</i></th></tr>
<tr><td>1</td><td colspan="3">Target KPI
UE Energy Efficiency
This KPI aims to measure the UE EE, and represents the efficiency with which each Joule of energy is used to transmit information. This KPI expression represents the data volume (MBs of data transfer) of the UE under consideration, divided by the total energy consumed.</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For the evaluation, a full 5G deployment composed of a 5G UE, a 5G NR gNB and 5G core is used. A traffic generator capable of producing FTP traffic is used behind the N6 interface of 5G system and a traffic sink that collects traffic is used at the UE side.</td></tr>
<tr><td>3</td><td colspan="3"><i>Calculation process and output</i>
EC Measurement
The UE EC (EC_{UE}) is the refers to the EC of the US under investigation. The UE EC is measured using power meters or a Keysight N6705 power analyser to collect energy measurements from UE where a phone is powered directly through the Power Analyzer, rather than through its internal battery, which is disconnected. EC_{UE} is measured in Watts/hr.[10] (Wh = Joule), over the period of measurement T.
DV Measurement
The DV_{UE} shall be measured using network counters for data volume related to the aggregated traffic in the UE under test.
DV_{UE} can be derived based on standard counters defined in ETSI TS 132 425 [34] for LTE (or 3GPP equivalent: TS 32.425), multiplying by the measurement duration T. DV_{UE} is computed in unit of bit.
OUTPUTs:
1. The $KPI_{UE\text{-}EE}$ (EE_{UE}), expressed unit of “bits/Joule” [equivalent to bits/Watt hour (Wh)]., is calculated as the ratio between the DV (DV_{UE}) and the EC (EC_{UE}), during the 7-day measurement period:
$EE_{UE} = DV_{UE} / EC_{UE}$ (1)
2. The mean (of the observed) values of EE_{UE} will be reported.
The required output should be calculated according to the following methodology:
Let EEavg_i be the calculated average EE for the ith iteration, and x_(i,n) be the measured EE for each iteration:
$EEavg_i = \frac{1}{n}\sum_{n} x_{i,n}$
Then, the overall (reported) average EE<i>mean</i> shall be calculated as the average of all x_i</td></tr>
</table>

---

[10] 1 Wh = 3·6 KJ

<table>
<tr><td></td><td>

$$EEmean = \frac{1}{i}\sum_{i} EEavg_i$$

For the overall average EE (EEmean), the 95% confidence interval shall be reported using the Student-T-distribution for *v* = *i* - *1* degrees of freedom to denote the precision of the experiment.
The EEmean output should be provided as:
<table>
<tr><td colspan="3">EE [b/j]</td></tr>
<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>
</table>
</td></tr>
<tr><td>4</td><td>Complementary measurements
The following complementary metrics will be available & logged during the measurement period:
At the UE:
• *DL RSRP*
• *DL RSRQ*
• *SINR*
At the eNB/gNB:
• *Uplink Signal Strength*
• *CQI reported from UE*</td></tr>
<tr><td>5</td><td>*Pre-conditions*
The experimenter will define the following parameters of their setup:
• Technology/Band
• Transmission Mode (TDD/FDD)
• Receive/Transmit Frequencies
• Number of Antennas
• Channel Bandwidth
• UE Category
• Service/slice configuration (eMBB, URLLC etc.)
Prior to the beginning of the tests:
• The UEs will be placed in various locations within the cell, including locations at the edge of the area (lowest SNR).
• There will be only one or more UEs successfully connected to the network.</td></tr>
<tr><td>6</td><td>Applicability
The UE and eNB/gNB should provide monitoring of the Complementary metrics.</td></tr>
<tr><td>7</td><td>Test Case Sequence
1. Place UEs (operating in NSA mode) in a test location corresponding to good coverage.
2. Configure the UE to download a test file size = 2 MB.
3. Start recording the complementary metrics.
4. Initiate the traffic generator in the Downlink with FTP traffic.
5. After the download is complete, allow the UE to go to inactive mode</td></tr>
</table>

	6. Repeat step 13 to 15 i.e. repeat the procedure 20 times. 7. Stop recording the complementary metrics. 8. Upon completing the measurement, record the DV_{UE}. 9. Upon completing the measurement, record the EC_{UE}. 10. After measuring the total energy consumed and computing the average per MB of data transfer, apply eq. (1). *11.* Report UE EE assessment results for 4G and 5G deployments separately, using the KPIs definition provided in section "Calculation process and output" as well as the templates provided in Annex 2.

5.1.4. Latency tests

5.1.4.1. Latency calibration tests

Average (expected mean) Latency calibration test

Test Case		*TC-Lat-001*	*Latency*
1	Target KPI ***Average (expected mean) Latency Calibration*** The Latency calibration tests aims at assessing the measurement capabilities of the measurement system employed for further Latency tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer		
2	Methodology For measuring Latency, a packet stream is emitted from a source and received by a data sink (destination). The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination. The time between sending the ICMP ECHO_REQUEST at the sender and the time at which the destination receives the ICMP ECHO_REQUEST is captured. The clocks used for creating the two time-stamps have to by synchronized. While the means on how to achieve this synchronization are out-of-scope of the definition of this test case, such means might include synchronizing local clocks via a GPS signal, the use of the Precision Time Protocol (PTP, IEEE 1588), or having source and destination physically collocated in order to employ a single packet capture system with a single clock for recording the sending times and reception times of the ICMP ECHO_REQUEST packet.		

The ICMP ECHO_RESPONSE packets sent from the destination to the source are either discarded in the evaluation of the metric or may be used to simultaneously assess to latency of the reverse (downstream) latency.

The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002.

The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

- Duration of a single replica (iteration) → at least 2 minutes
    - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25

3

Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) Latency:

Let avg_i be the calculated average Latency for the i^{th} replica (iteration), and x_(i,n) be the measured Latency for each packet (i.e. ICMP ECHO_REQUEST) within the replica (iteration).

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average Latency *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average Latency *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

Latency [ms]		
Mean	95% confidence interval for Mean	
	Lower bound	Upper bound

4

Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test
  For each one of these KPIs uses the following methodology

<table>
<tr><td></td><td>Average per iteration
$$avg_i = \frac{1}{n}\sum_n x_{i,n}$$
Average
$$avg = \frac{1}{i}\sum_i avg_i$$</td></tr>
<tr><td>5</td><td>Pre-conditions
• The scenario has been configured. In case of network slicing the slice must be activated.
• The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
• Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.</td></tr>
<tr><td>6</td><td>Applicability
The measurement probes need to be capable of injecting ICMP traffic in the system.</td></tr>
<tr><td>7</td><td>Test Case Sequence
1. Start monitoring probes.
2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined.
3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received.
4. Stop monitoring probes
5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”.
6. Repeat steps 1 to 5 for each one of the 25 iterations
7. Compute the KPIs as defined in section “Calculation process and output”</td></tr>
</table>

Minimum Latency calibration test

Test Case		TC-Lat-002	Latency
1	Target KPI ***Minimum Latency Calibration*** The Latency calibration tests aims at assessing the measurement capabilities of the measurement system employed for further Latency tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an "empty" system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer		
2	Methodology For measuring Latency, a packet stream is emitted from a source and received by a data sink (destination). The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination. The time between sending the ICMP ECHO_REQUEST at the sender and the time at which the destination receives the ICMP ECHO_REQUEST is captured. The clocks used for creating the two time-stamps have to by synchronized. While the means on how to achieve this synchronization are out-of-scope of the definition of this test case, such means might include synchronizing local clocks via a GPS signal, the use of the Precision Time Protocol (PTP, IEEE 1588), or having source and destination physically collocated in order to employ a single packet capture system with a single clock for recording the sending times and reception times of the ICMP ECHO_REQUEST packet. The ICMP ECHO_RESPONSE packets sent from the destination to the source are either discarded in the evaluation of the metric or may be used to simultaneously assess to latency of the reverse (downstream) latency. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single iteration → at least 2 minutes o Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25		

<table>
<tr><td>3</td><td>

Calculation process and output

The required output should be calculated according to the following methodology:

Min Latency:

Let Latency_min_i be the minimum Latency measured in the ith replica (iteration), and x_(i,n) be the measured Latency for each packet within the replica (iteration).

$$Latency\_min_i = \min(x_{i,n})$$

Then, the (reported) minimum Latency (Latency_min) shall be calculated as follows:

$$Latency\_min = \frac{1}{i}\sum_{i} Latency\_min_i$$

For the reported minimum Latency ($Latency\_$), the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.

The RTT output should be provided as:

<table>
<tr><td colspan="3">Latency [ms]</td></tr>
<tr><td rowspan="2">Min</td><td colspan="2">95% confidence interval for Min</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>

</td></tr>
<tr><td>4</td><td>

Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test
  For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$

</td></tr>
<tr><td>5</td><td>

Pre-conditions

- The scenario has been configured. In case of network slicing the slice must be activated.
- The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
- Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.

</td></tr>
</table>

6	Applicability The measurement probes need to be capable of injecting ICMP traffic in the system.
7	Test Case Sequence 1. Start monitoring probes. 2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

Maximum Latency calibration test

Test Case		*TC-Lat-003*	*Latency*
1	Target KPI ***Maximum Latency Calibration*** The Latency calibration tests aims at assessing the measurement capabilities of the measurement system employed for further Latency tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of a an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer		
2	Methodology For measuring Latency, a packet stream is emitted from a source and received by a data sink (destination). The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination. The time between sending the ICMP ECHO_REQUEST at the sender and the time at which the destination receives the ICMP ECHO_REQUEST is captured. The clocks used for creating the two time-stamps have to by synchronized. While the means on how to achieve this synchronization are out-of-scope of the definition of this test case, such means might include synchronizing local clocks via a GPS signal, the use of the Precision Time Protocol (PTP, IEEE 1588), or having source and destination		

physically collocated in order to employ a single packet capture system with a single clock for recording the sending times and reception times of the ICMP ECHO_REQUEST packet.

The ICMP ECHO_RESPONSE packets sent from the destination to the source are either discarded in the evaluation of the metric or may be used to simultaneously assess to latency of the reverse (downstream) latency.

The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002.

The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

- Duration of a single replica (iteration) → at least 2 minutes
    - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25

3

Calculation process and output

The required output should be calculated according to the following methodology:

Max Latency:

Let Latency_max_i be the maximum Latency measured in the ith replica (iteration), and x_(i,n) be the measured Latency for each packet within the replica (iteration).

$$Latency\_max_i = \max(x_{i,n})$$

Then, the (reported) maximum Latency (Latency_max) shall be calculated as follows:

$$Latency\_max = \frac{1}{i}\sum_i Latency\_max_i$$

For the reported maximum Latency ($Latency\_max$), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

The RTT output should be provided as:

Latency [ms]		
Max	95% confidence interval for Min	
	Lower bound	Upper bound

4

Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test
  For each one of these KPIs uses the following methodology

Average per iteration

	Average $avg_i = \frac{1}{n}\sum_{n} x_{i,n}$ $avg = \frac{1}{i}\sum_{i} avg_i$
5	Pre-conditions • The scenario has been configured. In case of network slicing the slice must be active. • The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE. • Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	Applicability The measurement probes need to be capable of injecting ICMP traffic in the system.
7	Test Case Sequence 1. Start monitoring probes. 2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5%-percentile Latency calibration test

Test Case	*TC-Lat-004*	*Latency*
1	Target KPI ***5%-percentile Latency Calibration*** The Latency calibration tests aims at assessing the measurement capabilities of the measurement system employed for further Latency tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of a an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client	

	Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer
2	Methodology For measuring Latency, a packet stream is emitted from a source and received by a data sink (destination). The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination. The time between sending the ICMP ECHO_REQUEST at the sender and the time at which the destination receives the ICMP ECHO_REQUEST is captured. The clocks used for creating the two time-stamps have to by synchronized. While the means on how to achieve this synchronization are out-of-scope of the definition of this test case, such means might include synchronizing local clocks via a GPS signal, the use of the Precision Time Protocol (PTP, IEEE 1588), or having source and destination physically collocated in order to employ a single packet capture system with a single clock for recording the sending times and reception times of the ICMP ECHO_REQUEST packet. The ICMP ECHO_RESPONSE packets sent from the destination to the source are either discarded in the evaluation of the metric or may be used to simultaneously assess to latency of the reverse (downstream) latency. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes o Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25
3	*Calculation process and output* The required output should be calculated according to the following methodology: **5%-Percentile Latency**: Let Latency_5p_i be the 5%-percentile Latency measured in the ith replica (iteration), and x_(i,n) be the measured Latency for each packet within the replica (iteration). $Latency_{5p_i} = 5\%\ \text{percentile}(x_{i,n})$ Then, the (reported) 5%-percentile Latency (Latency_5p) shall be calculated as follows: $Latency\_5p = \frac{1}{i}\sum_{i} Latency\_5p_i$ For the reported 5%-percentile Latency ($Latency\_5p$), the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment. The RTT output should be provided as: Latency [ms]

<table>
<tr><td></td><td><table><tr><td rowspan="2">5% Per-cen-tile</td><td colspan="2">95% confidence interval for Min</td></tr><tr><td>Lower bound</td><td>Upper bound</td></tr></table></td></tr>
<tr><td>4</td><td>Complementary measurements

Ping Success Rate
PL Rate
Average RSRP (Reference Signal Received Power) if available
Average RSRQ (Reference Signal Received Quality) if available
In case of intentional background traffic measure the average IP throughput during the test
For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$
</td></tr>
<tr><td>5</td><td>Pre-conditions

The scenario has been configured. In case of network slicing the slice must be active.
The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.

</td></tr>
<tr><td>6</td><td>Applicability
The measurement probes need to be capable of injecting ICMP traffic in the system.</td></tr>
<tr><td>7</td><td>Test Case Sequence

Start monitoring probes.
Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined.
Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received.
Stop monitoring probes
Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”.
Repeat steps 1 to 5 for each one of the 25 iterations
Compute the KPIs as defined in section “Calculation process and output”

</td></tr>
</table>

95%-percentile Latency calibration test

Test Case	*TC-Lat-005*	*Latency*

1	Target KPI *95%-percentile Latency Calibration* The Latency calibration tests aims at assessing the measurement capabilities of the measurement system employed for further Latency tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of a an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer
2	Methodology For measuring Latency, a packet stream is emitted from a source and received by a data sink (destination). The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination. The time between sending the ICMP ECHO_REQUEST at the sender and the time at which the destination receives the ICMP ECHO_REQUEST is captured. The clocks used for creating the two time-stamps have to by synchronized. While the means on how to achieve this synchronization are out-of-scope of the definition of this test case, such means might include synchronizing local clocks via a GPS signal, the use of the Precision Time Protocol (PTP, IEEE 1588), or having source and destination physically collocated in order to employ a single packet capture system with a single clock for recording the sending times and reception times of the ICMP ECHO_REQUEST packet. The ICMP ECHO_RESPONSE packets sent from the destination to the source are either discarded in the evaluation of the metric or may be used to simultaneously assess to latency of the reverse (downstream) latency. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes ○ Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25
3	*Calculation process and output* The required output should be calculated according to the following methodology: **95%-Percentile Latency**:

<table>
<tr>
<td></td>
<td>
<p>Let Latency_95p_i be the 95%-percentile Latency measured in the ith replica (iteration), and x_(i,n) be the measured Latency for each packet within the replica (iteration).</p>

$$Latency_{95p_i} = 95\%\ \mathrm{percentile}(x_{i,n})$$

<p>Then, the (reported) 95%-percentile Latency (Latency_95p) shall be calculated as follows:</p>

$$Latency\_95p = \frac{1}{i}\sum_{i} Latency\_95p_i$$

<p>For the reported 95%-percentile Latency ($Latency\_95p$), the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.</p>
<p>The RTT output should be provided as:</p>
<table>
<tr><td colspan="3">Latency [ms]</td></tr>
<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>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
<p>Complementary measurements</p>

Ping Success Rate
PL Rate
Average RSRP (Reference Signal Received Power) if available
Average RSRQ (Reference Signal Received Quality) if available
In case of intentional background traffic measure the average IP throughput during the test
For each one of these KPIs uses the following methodology

<p>Average per iteration</p>

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

<p>Average</p>

$$avg = \frac{1}{i}\sum_{i} avg_i$$
</td>
</tr>
<tr>
<td>5</td>
<td>
<p>Pre-conditions</p>

The scenario has been configured. In case of network slicing the slice must be activated.
The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.

</td>
</tr>
<tr>
<td>6</td>
<td>
<p>Applicability</p>
<p>The measurement probes need to be capable of injecting ICMP traffic in the system.</p>
</td>
</tr>
<tr>
<td>7</td>
<td>
<p>Test Case Sequence</p>

Start monitoring probes.

</td>
</tr>
</table>

	2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5.1.4.2. E2E Application Layer Latency tests

Average (expected mean) E2E Application Layer Latency tests

<table>
<tr><th colspan="2">Test Case</th><th>TC-Lat-006</th><th>Latency</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Average (expected mean) E2E Application Layer Latency
This test aims at measure the mean between a total of time of delays between the transmission and the reception of a data packet at application level, from a source and to the destination, and its processing.
This latency is measured between a transmitter source and a receiver, in one direction each time, upstream or downstream. It is measured at the application layer.
This test employs real application-oriented data packet to measure the latency metric within a real environment. Thus, the traffic profile of this test case is application dependent, and has different payload size and headers, thus, must be specified in the instantiation of the test case.
Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects.
Source of packets →application acting as client
Destination of packets → application acting as server
Underlying SUT → The deployed infrastructure of the specific platform, with the components related to a particular use case.
Measurement conducted at layer → Application layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring E2E Application Layer Latency, a packet is emitted from a source and received by the application. The times of emitting the packet at the data source and receiving the data packet at the data sink are to be captured and the Latency is calculated as the difference between the two events.
For this test case, the traffic profile is application-based so, the data packet size and headers depending on the application used by the scenario. Moreover, the times for sending and receiving the packet are measured at the application level. Thus, this latency includes the de-encapsulation of the data within the packet from the receiver.
The traffic source sends an Application data packet towards the destinations. The time between sending this packet at the sender and the time at which the destination receives the packet is captured. The clocks used for creating the two timestamps have to be previously well synchronized,</td></tr>
</table>

<table>
<tr>
<td></td>
<td>
preferably with an external source of time, and through the desired mechanism (GPS, PTP, NTP, etc.)

The test case shall include the consecutive execution of several iterations according to the following properties.

- Duration of a single iteration → at least 1 minute, or the delivery of 50 packets.
- A number of replica (iterations) → At least 25.
</td>
</tr>
<tr>
<td>3</td>
<td>
Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) E2E Application Layer Latency:

Let avg_i be the calculated average Latency for the ith iteration, and x_(i,n) be the measured Latency for each packet, with a specific Application traffic profile, within the iteration.

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average Latency *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average Latency *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The Latency output should be provided as:

<table>
<tr><th colspan="3">Latency [ms]</th></tr>
<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>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
Complementary measurements

Other interesting measurements to be computed with the test results could be:

- PLRate

$$PLR = \frac{Num.of\ lost\ packets}{Num.of\ packets\ received + Num.of\ lost\ packets}$$

- Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

- Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$
</td>
</tr>
</table>

5	*Pre-conditions* Prior to the beginning of the tests: • All components belonging to the Use Case Application has to be in place, running and connecting to each other. Thus, the source has to reach the destination within a given time. Moreover, in the case of network slicing the slice must be activated. • The source has to generate the specified Application packet in a period interval. This data must arrive at least once to the destination, and the information within this data packet has to arrive uncorrupted. • Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	*Applicability* • The SUT must support/handle specific application type traffic. • The measurement probes need to be capable of injecting Application traffic in the system, or to be able to measure the traffic already exchanged by the SUT.
7	*Test Case Sequence* 1. Start the monitoring module to record the departure and arrival of the packets. 2. Log the information in a structured manner, this information includes timestamps, data information and other extra information that could be relevant for the probes. 3. Stop the monitoring module. 4. Calculate and record the average latency and the other parameters specified in this test case. 5. Replicate steps 1 to 5 the number of times specified in this test case. 6. Compute the total of KPI values defined in this test case.

5.1.5. Round-trip-time tests

5.1.5.1. RTT calibration tests

Average (expected mean) RTT calibration test

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Rtt-001</i></th><th><i>Round-Trip-time</i></th></tr>
<tr><td>1</td><td colspan="3">Target KPI
<i>Average (expected mean) Round-Trip-Time Calibration</i>
The RTT calibration tests aims at assessing the measurement capabilities of the measurement system employed for further RTT tests.
The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” system under test. The calibration test is conducted at the network layer.
Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects.
Source of ping packets →measurement probe acting as client
Destination of ping packets → measurement probe acting as server</td></tr>
</table>

<table>
<tr>
<td></td>
<td>

Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT

Measurement conducted at layer → Network layer

</td>
</tr>
<tr>
<td>2</td>
<td>

Methodology

For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events.

For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs "ping" as a tool to generate the traffic stream.

The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002.

The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

- Duration of a single replica (iteration) → at least 2 minutes
    - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25

</td>
</tr>
<tr>
<td>3</td>
<td>

Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) RTT:

Let avg_i be the calculated average RTT for the i^{th} replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average RTT *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average RTT *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

<table>
<tr><td colspan="2">RTT [ms]</td></tr>
<tr><td>Mean</td><td>95% confidence interval for Mean</td></tr>
</table>

</td>
</tr>
</table>

<table>
<tr><td></td><td><table><tr><td></td><td>Lower bound</td><td>Upper bound</td></tr></table></td></tr>
<tr><td>4</td><td>Complementary measurements
• Ping Success Rate
• PL Rate
• Average RSRP (Reference Signal Received Power) if available
• Average RSRQ (Reference Signal Received Quality) if available
• In case of intentional background traffic measure the average IP throughput during the test
For each one of these KPIs uses the following methodology
Average per iteration
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Average
$$avg = \frac{1}{i}\sum_{i} avg_i$$</td></tr>
<tr><td>5</td><td>Pre-conditions
• The scenario has been configured. In case of network slicing the slice must be activated.
• The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
• Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.</td></tr>
<tr><td>6</td><td>Applicability
The measurement probes need to be capable of injecting ICMP traffic in the system.</td></tr>
<tr><td>7</td><td>Test Case Sequence
1. Start monitoring probes.
2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined.
3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received.
4. Stop monitoring probes
5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”.
6. Repeat steps 1 to 5 for each one of the 25 iterations
7. Compute the KPIs as defined in section “Calculation process and output”</td></tr>
</table>

Minimum RTT calibration test

<table>
<tr><th colspan="2">Test Case</th><th>*TC-Rtt-002*</th><th>*Round-Trip-time*</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Minimum Round-Trip-Time Calibration
The RTT calibration tests aims at assessing the measurement capabilities of the measurement system employed for further RTT tests.
The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” system under test. The calibration test is conducted at the network layer.
Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects.
Source of ping packets →measurement probe acting as client
Destination of ping packets → measurement probe acting as server
Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT
Measurement conducted at layer → Network layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events.
For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream.
The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002.
The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
- Duration of a single replica (iteration) → at least 2 minutes
    - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25</td></tr>
<tr><td>3</td><td colspan="3">*Calculation process and output*
The required output should be calculated according to the following methodology:
Min RTT:
Let RTTmin_i be the minimum RTT measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).
$$RTTmin_i = \min(x_{i,n})$$</td></tr>
</table>

<table>
<tr>
<td></td>
<td>
Then, the (reported) minimum RTT (RTTmin) shall be calculated as follows:

$$RTTmin = \frac{1}{i}\sum_{i} RTTmin_i$$

For the reported minimum RTT (RTTmin), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

The RTT output should be provided as:

<table>
<tr><th colspan="3">RTT [ms]</th></tr>
<tr><td rowspan="2">Min</td><td colspan="2">95% confidence interval for Min</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$
</td>
</tr>
<tr>
<td>5</td>
<td>
Pre-conditions

- The scenario has been configured. In case of network slicing the slice must be activated.
- The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
- Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
</td>
</tr>
<tr>
<td>6</td>
<td>
Applicability

The measurement probes need to be capable of injecting ICMP traffic in the system.
</td>
</tr>
<tr>
<td>7</td>
<td>
Test Case Sequence

1. Start monitoring probes.
2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined.
3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received.
4. Stop monitoring probes
</td>
</tr>
</table>

	5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

Maximum RTT calibration test

Test Case		*TC-Rtt-003*	*Round-Trip-time*
1	Target KPI ***Minimum Round-Trip-Time Calibration*** The RTT calibration tests aims at assessing the measurement capabilities of the measurement system employed for further RTT tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer		
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes ○ Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25		

<table>
<tr>
<td>3</td>
<td>
<p><i>Calculation process and output</i></p>
<p>The required output should be calculated according to the following methodology:</p>
<p>Max RTT:</p>
<p>Let RTTmax_i be the maximum RTT measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).</p>
<p>$$RTTmax_i = \max(x_{i,n})$$</p>
<p>Then, the (reported) maximum RTT (RTTmax) shall be calculated as follows:</p>
<p>$$RTTmax = \frac{1}{i}\sum_{i} RTTmax_i$$</p>
<p>For the reported maximum RTT (RTTmax), the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.</p>
<p>RTT output should be provided as:</p>
<table>
<tr><td colspan="3">RTT [ms]</td></tr>
<tr><td rowspan="2">Max</td><td colspan="2">95% confidence interval for Max</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
<p>Complementary measurements</p>

Ping Success Rate
PL Rate
Average RSRP (Reference Signal Received Power) if available
Average RSRQ (Reference Signal Received Quality) if available
In case of intentional background traffic measure the average IP throughput during the test
For each one of these KPIs uses the following methodology

<p>Average per iteration</p>
<p>$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$</p>
<p>Average</p>
<p>$$avg = \frac{1}{i}\sum_{i} avg_i$$</p>
</td>
</tr>
<tr>
<td>5</td>
<td>
<p>Pre-conditions</p>

The scenario has been configured. In case of network slicing the slice must be activated.
The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.

</td>
</tr>
<tr>
<td>6</td>
<td>
<p>Applicability</p>
<p>The measurement probes need to be capable of injecting ICMP traffic in the system.</p>
</td>
</tr>
</table>

7	Test Case Sequence 1. Start monitoring probes. 2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5%-percentile RTT calibration test

Test Case		*TC-Rtt-004*	*Round-Trip-time*
1	Target KPI ***5%-Percentile Round-Trip-Time Calibration*** The RTT calibration tests aims at assessing the measurement capabilities of the measurement system employed for further RTT tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer		
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. The test case shall include the consecutive execution of several replica (iterations) according to the following properties.		

<table>
<tr><td></td><td>

- Duration of a single replica (iteration) → at least 2 minutes
  - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25

</td></tr>
<tr><td>3</td><td>

Calculation process and output

The required output should be calculated according to the following methodology:

5%-Percentile RTT:

Let RTT_5p_i be the 5%-percentile RTT measured in the ith replica (iteration), and x_(i,n) be the RTT for each packet within the replica (iteration).

$$RTT_{5p_i} = 5\%\ \text{percentile}(x_{i,n})$$

Then, the (reported) 5%-percentile RTT (RTT_5p) shall be calculated as follows:

$$RTT\_5p = \frac{1}{i}\sum_{i} RTT\_5p_i$$

For the reported 5%-percentile RTT (RTT_5p), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

The RTT output should be provided as:

<table>
<tr><th colspan="3">RTT [ms]</th></tr>
<tr><td rowspan="2">5% Per-centile</td><td colspan="2">95% confidence interval for 5% Percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>

</td></tr>
<tr><td>4</td><td>

Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$

</td></tr>
<tr><td>5</td><td>

Pre-conditions

- The scenario has been configured. In case of network slicing the slice must be activated.
- The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.

</td></tr>
</table>

	• Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	Applicability The measurement probes need to be capable of injecting ICMP traffic in the system.
7	Test Case Sequence 1. Start monitoring probes. 2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

95% -percentile RTT calibration test

Test Case	*TC-Rtt-005*	*Round-Trip-time*
1	Target KPI ***95%-Percentile Round-Trip-Time Calibration*** The RTT calibration tests aims at assessing the measurement capabilities of the measurement system employed for further RTT tests. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” system under test. The calibration test is conducted at the network layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of ping packets →measurement probe acting as client Destination of ping packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Network layer	
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events.	

<table>
<tr><td></td><td>For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream.
The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002.
The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
• Duration of a single replica (iteration) → at least 2 minutes
○ Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
• Number of replica (iterations) → At least 25</td></tr>
<tr><td>3</td><td><i>Calculation process and output</i>
The required output should be calculated according to the following methodology:
95%-Percentile RTT:
Let RTT_95p_i be the 95%-percentile RTT measured in the ith replica (iteration), and x_(i,n) be the RTT for each packet within the replica (iteration).
$$RTT_{95p_i} = 95\% \text{ percentile}(x_{i,n})$$
Then, the (reported) 95%-percentile RTT (RTT_95p) shall be calculated as follows:
$$RTT\_95p = \frac{1}{i}\sum_{i} RTT\_95p_i$$
For the reported 95%-percentile RTT (RTT_95p), the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.
The RTT output should be provided as:
<table>
<tr><td colspan="3">RTT [ms]</td></tr>
<tr><td rowspan="2">95% Percentile</td><td colspan="2">95% confidence interval for 95% Percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>Complementary measurements
• Ping Success Rate
• PL Rate
• Average RSRP (Reference Signal Received Power) if available
• Average RSRQ (Reference Signal Received Quality) if available
• In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology
Average per iteration
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Average</td></tr>
</table>

	$avg = \frac{1}{i} \sum_{i} avg_i$
5	Pre-conditions • The scenario has been configured. In case of network slicing the slice must be activated. • The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE. • Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	Applicability The measurement probes need to be capable of injecting ICMP traffic in the system.
7	Test Case Sequence 1. Start monitoring probes. 2. Using the traffic generator, begin pinging from one probe to the other using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5.1.5.2. E2E network layer RTT test (LTE Rel.14 Core and RAN) tests

Average (expected mean) E2E network layer RTT test (LTE Rel.14 Core and RAN)

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Rtt-006</i></th><th><i>Round-Trip-time</i></th></tr>
<tr><td>1</td><td colspan="3">Target KPI
<i>E2E network layer</i>
<i>Average (expected mean) Round-Trip-Time</i>
This KPI refers to the RTT measured from a client to a server over a MN. The Core and the RAN domain of the underlay network abide by the specifications of 3GPP release 14. It measures the duration from the transmission of the data packet at the UE, to the successful reception at the node of an external server connected directly to the UPF function of the core network (N6 interface), plus the response time back to the UE.
Source A → UE
Destination B → Server node at the N6 interface.
Underlying system → LTE release 14 Core and RAN
Layer → Network layer</td></tr>
</table>

<table>
<tr>
<td>2</td>
<td>Methodology

For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events.

For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream.

The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. Additional traffic profiles may be applied for additional experiments.

The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

• Duration of a single replica (iteration) → at least 2 minutes
o Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
• Number of replica (iterations) → At least 25</td>
</tr>
<tr>
<td>3</td>
<td>Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) RTT:

Let avg_i be the calculated average RTT for the i^{th} replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Then, the overall (reported) average RTT avg shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$
For the overall average RTT avg, the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

<table>
<tr><td colspan="3">RTT [ms]</td></tr>
<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>
</table></td>
</tr>
</table>

4	Complementary measurements - Ping Success Rate - PL Rate - Average RSRP (Reference Signal Received Power) if available - Average RSRQ (Reference Signal Received Quality) if available - In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology Average per iteration $$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$ Average $$avg = \frac{1}{i}\sum_{i} avg_i$$
5	Pre-conditions The scenario has been configured. In case of network slicing, the slice must be activated. The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE. Connect a reachable computer machine (end point) in the N6 interface. Deploy the monitoring probes to collect RSRP, RSRQ and IP throughput measurements. Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	Applicability For networks and devices that support internal traffic generation or routing external IP traffic
7	Test Case Sequence 1. Start monitoring probes to collect RSRP, RSRQ and IP throughput, if applicable. 2. Using the traffic generator, begin pinging from the UE to the end-point using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

Minimum E2E network layer RTT test (LTE Rel.14 Core and RAN)

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Rtt-007</i></th><th><i>Round-Trip-time</i></th></tr>
<tr><td>1</td><td colspan="3">Target KPI
<i>E2E network layer</i>
<i>Minimum Round-Trip-Time</i>
This KPI refers to the RTT measured from a client to a server over a MN. The Core and the RAN domain of the underlay network abide by the specifications of 3GPP release 14. It measures the duration from the transmission of the data packet at the UE, to the successful reception at the node of an external server connected directly to the UPF function of the core network (N6 interface), plus the response time back to the UE.
Source A → UE
Destination B → Server node at the N6 interface.
Underlying system → LTE release 14 Core and RAN
Layer → Network layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events.
For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream.
The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. Additional traffic profiles may be applied for additional experiments.
The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
• Duration of a single replica (iteration) → at least 2 minutes
    ○ Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
• Number of replica (iterations) → At least 25</td></tr>
<tr><td>3</td><td colspan="3"><i>Calculation process and output</i>
The required output should be calculated according to the following methodology:
Min RTT:
Let RTTmin_i be the minimum RTT measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).
$$RTTmin_i = \min(x_{i,n})$$
Then, the (reported) minimum RTT (RTTmin) shall be calculated as follows:
$$RTTmin = \frac{1}{i}\sum_{i} RTTmin_i$$</td></tr>
</table>

<table>
<tr><td></td><td>
For the reported minimum RTT (RTTmin), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.
The RTT output should be provided as:
<table>
<tr><td colspan="3">RTT [ms]</td></tr>
<tr><td rowspan="2">Min</td><td colspan="2">95% confidence interval for Min</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>
Complementary measurements
• Ping Success Rate
• PL Rate
• Average RSRP (Reference Signal Received Power) if available
• Average RSRQ (Reference Signal Received Quality) if available
• In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology
Average per iteration
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Average
$$avg = \frac{1}{i}\sum_{i} avg_i$$
</td></tr>
<tr><td>5</td><td>
Pre-conditions
The scenario has been configured. In case of network slicing the slice must be activated.
The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.
Connect a reachable computer machine (end point) in the N6 interface.
Deploy the monitoring probes to collect RSRP, RSRQ and IP throughput measurements.
Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
</td></tr>
<tr><td>6</td><td>
Applicability
For networks and devices that support internal traffic generation or routing external IP traffic
</td></tr>
<tr><td>7</td><td>
Test Case Sequence
1. Start monitoring probes to collect RSRP, RSRQ and IP throughput, if applicable.
2. Using the traffic generator, begin pinging from the UE to the end-point using one of the ICMP traffic patterns defined.
3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received.
4. Stop monitoring probes
5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”.
</td></tr>
</table>

	6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

Maximum E2E network layer RTT test (LTE Rel.14 Core and RAN)

Test Case		*TC-Rtt-008*	*Round-Trip-time*
1	Target KPI ***E2E network layer*** ***Maximum Round-Trip-Time*** This KPI refers to the RTT measured from a client to a server over a MN. The Core and the RAN domain of the underlay network abide by the specifications of 3GPP release 14. It measures the duration from the transmission of the data packet at the UE, to the successful reception at the node of an external server connected directly to the UPF function of the core network (N6 interface), plus the response time back to the UE. Source A → UE Destination B → Server node at the N6 interface. Underlying system → LTE release 14 Core and RAN Layer → Network layer		
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. Additional traffic profiles may be applied for additional experiments. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes ○ Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25		
3	*Calculation process and output* The required output should be calculated according to the following methodology: **Max RTT**: Let RTTmax_i be the maximum RTT measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration). $RTTmax_i = \max(x_{i,n})$ Then, the (reported) maximum RTT (RTTmax) shall be calculated as follows:		

<table>
<tr>
<td></td>
<td>

$$RTTmax = \frac{1}{i}\sum_{i} RTTmax_i$$

For the reported minimum RTT (RTTmin), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

The RTT output should be provided as:

<table>
<tr><td colspan="3">RTT [ms]</td></tr>
<tr><td rowspan="2">Max</td><td colspan="2">95% confidence interval for Max</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>

</td>
</tr>
<tr>
<td>4</td>
<td>

Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$

</td>
</tr>
<tr>
<td>5</td>
<td>

Pre-conditions

The scenario has been configured. In case of network slicing the slice must be activated.

The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.

Connect a reachable computer machine (end point) in the N6 interface.

Deploy the monitoring probes to collect RSRP, RSRQ and IP throughput measurements.

Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.

</td>
</tr>
<tr>
<td>6</td>
<td>

Applicability

For networks and devices that support internal traffic generation or routing external IP traffic

</td>
</tr>
<tr>
<td>7</td>
<td>

Test Case Sequence

1. Start monitoring probes to collect RSRP, RSRQ and IP throughput, if applicable.
2. Using the traffic generator, begin pinging from the UE to the end-point using one of the ICMP traffic patterns defined.

</td>
</tr>
</table>

	3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5%-percentile E2Enetwork layer RTT test (LTE Rel.14 Core and RAN)

Test Case		*TC-Rtt-009*	*Round-Trip-time*
1	Target KPI ***E2E network layer*** ***5% Percentile Round-Trip-Time*** This KPI refers to the RTT measured from a client to a server over a MN. The Core and the RAN domain of the underlay network abide by the specifications of 3GPP release 14. It measures the duration from the transmission of the data packet at the UE, to the successful reception at the node of an external server connected directly to the UPF function of the core network (N6 interface), plus the response time back to the UE. Source A → UE Destination B → Server node at the N6 interface. Underlying system → LTE release 14 Core and RAN Layer → Network layer		
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. Additional traffic profiles may be applied for additional experiments. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes o Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration). • Number of replica (iterations) → At least 25		
3	*Calculation process and output* The required output should be calculated according to the following methodology: **5% percentile**:		

<table>
<tr><td></td><td>
Let p5_i be the 5% percentile measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).

$$p5_i = 5 percent\ percentile\ for\ all\ x_{i,n}$$

Then, the (reported) 5% percentile shall be calculated as follows:

$$p5 = \frac{1}{i}\sum_i p5_i$$

For the reported 5% percentile, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

<table>
<tr><th colspan="3">RTT [ms]</th></tr>
<tr><td rowspan="2">5% percentile</td><td colspan="2">95% confidence interval for 5% percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>
Complementary measurements

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_n x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_i avg_i$$
</td></tr>
<tr><td>5</td><td>
Pre-conditions

The scenario has been configured. In case of network slicing the slice must be activated.

The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE.

Connect a reachable computer machine (end point) in the N6 interface.

Deploy the monitoring probes to collect RSRP, RSRQ and IP throughput measurements.

Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
</td></tr>
</table>

6	Applicability For networks and devices that support internal traffic generation or routing external IP traffic
7	Test Case Sequence 1. Start monitoring probes to collect RSRP, RSRQ and IP throughput, if applicable. 2. Using the traffic generator, begin pinging from the UE to the end-point using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

95%-percentile E2E network layer RTT test (LTE Rel.14 Core and RAN)

Test Case		*TC-Rtt-010*	*Round-Trip-time*
1	Target KPI ***E2E network layer*** ***95% Percentile Round-Trip-Time*** This KPI refers to the RTT measured from a client to a server over a MN. The Core and the RAN domain of the underlay network abide by the specifications of 3GPP release 14. It measures the duration from the transmission of the data packet at the UE, to the successful reception at the node of an external server connected directly to the UPF function of the core network (N6 interface), plus the response time back to the UE. Source A → UE Destination B → Server node at the N6 interface. Underlying system → LTE release 14 Core and RAN Layer → Network layer		
2	Methodology For measuring RTT, a packet stream is emitted from a source and received by a data sink (destination). The data sink shall acknowledge the correct reception of the data packet back to the sink. The time of emitting the packet at the data source and receiving the acknowledgement at the data source are to be captured and the RTT is calculated as the difference between the two events. For this test case, the traffic source sends ICMP ECHO_REQUEST towards the destination, the latter responding with an ICMP ECHO_RESPONSE. This test case employs “ping” as a tool to generate the traffic stream. The stream of generated ICMP ECHO_REQUESTs shall comply to the following traffic profile as specified in TD-002. Additional traffic profiles may be applied for additional experiments. The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → at least 2 minutes		

  - Note, the duration has to ensure that at least 100 ICMP ECHO_REQUESTs are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25

Calculation process and output

The required output should be calculated according to the following methodology:

95% percentile:

Let p95_i be the 95% percentile measured in the ith replica (iteration), and x_(i,n) be the measured RTT for each packet within the replica (iteration).

$$p95_i = 95percent\ percentile\ for\ all\ x_{i,n}$$

Then, the (reported) 5% percentile shall be calculated as follows:

$$p95 = \frac{1}{i}\sum_{i} p95_i$$

3 For the reported 5% percentile, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

RTT [ms]		
5% percentile	95% confidence interval for 5% percentile	
	Lower bound	Upper bound

Complementary measurements

A secondary list of KPIs useful to interpret the values of the target KPI. Getting these measurements is not mandatory for the test case.

- Ping Success Rate
- PL Rate
- Average RSRP (Reference Signal Received Power) if available
- Average RSRQ (Reference Signal Received Quality) if available
- In case of intentional background traffic measure the average IP throughput during the test
4 For each one of these KPIs uses the following methodology

Average per iteration

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Average

$$avg = \frac{1}{i}\sum_{i} avg_i$$

5	Pre-conditions The scenario has been configured. In case of network slicing the slice must be activated. The traffic generator should support the generation of the traffic pattern defined in the RTT traffic patterns section. Connect the traffic generator to the UE if not internal in the UE. Connect a reachable computer machine (end point) in the N6 interface. Deploy the monitoring probes to collect RSRP, RSRQ and IP throughput measurements. Ensure that unless specifically requested in the scenario, no undesired traffic is present during the test.
6	Applicability For networks and devices that support internal traffic generation or routing external IP traffic
7	Test Case Sequence 1. Start monitoring probes to collect RSRP, RSRQ and IP throughput, if applicable. 2. Using the traffic generator, begin pinging from the UE to the end-point using one of the ICMP traffic patterns defined. 3. Record RTT, number of packets “ICMP request” sent and number of packets “ICMP reply” received. 4. Stop monitoring probes 5. Calculate and record the average RTT, the average PL rate and the average ping success rate, the average RSRP, the average RSRQ and the average IP throughput per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5.1.6. Location accuracy tests

Location accuracy tests were not conducted during the 5GENESIS experimentation Phase 1. Corresponding test cases will be added in a later revision of this deliverable (D6.2 or D6.3).

5.1.7. Reliability tests

Reliability tests were not conducted during the 5GENESIS experimentation Phase 1. Corresponding tests will be added in a later revision of this deliverable (D6.2 or D6.3).

5.1.8. Service creation time tests[11]

5.1.8.1. Service creation time calibration tests

Average (expected mean) service creation time for deploying virtual instruments on a single compute host

<table>
<tr><th colspan="2">Test Case</th><th><i>TC-Ser-001</i></th><th><i>Service Creation Time</i></th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Average Service Creation Time
for deploying virtual instruments on a single compute host
This KPI refers to the time needed to deploy a single virtual machine (VM), which may act as virtual instruments, on a single compute machine. The main intend is to evaluate the capability of the system to deploy a service (VNFs) on an existing, i.e. previously deployed, 4G/5G Network. Since the test case mandates that both VMs are deployed on the same compute host, i.e. in a single availability zone, potential effects of deploying VNFs in different zones do not impact the results.
For this test, a simple VM running Debian-9 is deployed. To assure correct deployment of a service offered by the VM, a “ping” is triggered from that machine towards a remote host. The latter step not being part of the time measured. As such, the deployed VNF offers the service “reachability test”.
Source A → Virtual Machine / VNF deployed
Destination B → Remote server in the testbed or outside the testbed
Underlying system → OpenStack
Layer → Network layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring the service creation time, the deployment of a single VM (Debian-9-based) is triggered by the orchestrator of the testbed. The time between triggering the deployment process and the indication of a successful deployment is measured.
After the deployment, the VM is used to conduct a “ping” (offered service) originating at the deployed VM towards any reachable remote host. Since the duration of this service test is not part of the measurement, the specification on how to parameterize the “ping” is out of scope of this test case.
The test case shall include the consecutive execution of several iterations according to the following properties.
• Duration of a single iteration → deployment of the the single VNF / VM
• Number of replica → At least 25 iterations</td></tr>
</table>

---

[11] The set of measurements for the service creation time, provided in this deliverable, are related only to the service activation time measured in a well controlled subsystem and as such the provided values depict the contribution of the activation procedure to the overall time and not the total amount of time needed for the creation of a service

<table>
<tr><td>3</td><td>

Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) RTT:

Let avg_i be the calculated average service creation time for the i^{th} iteration, and x_(i,n) be the measured service creation time for each deployment within the iteration.

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average service creation time *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average service creation time *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The service creation time output should be provided as:

<table>
<tr><td colspan="3">Service Creation Time [s]</td></tr>
<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>
</table>

</td></tr>
<tr><td>4</td><td>

Complementary measurements

Deployment Success Rate in % (number of deployments, which resulted in a successful ping afterwards).

</td></tr>
<tr><td>5</td><td>

Pre-conditions

A deployed and working OpenStack facilitating the deployment of VNFs / VMs via an orchestrator (e.g. OpenBaton or OSM).

An existing remote server that can be used to conduct the service test, i.e. to ping.

No other VMs / VNFs are deployed on the compute hosted on which the single VNF / VM for this test case is deployed.

</td></tr>
<tr><td>6</td><td>

Applicability

This test case applies for all scenarios as it assumes an underlying network infrastructure to deploy the VNFs/VMs in.

</td></tr>
<tr><td>7</td><td>

Test Case Sequence

1. Trigger the deployment of the VNF/VM from the orchestrator, or from experiment coordinator (TAP) on top of the latter.
2. Wait for a response indicating successful deployment

</td></tr>
</table>

	3. Calculate the deployment time (time between steps 1 and 2) 4. Verify that the deployed service is working, i.e. ping the remote machine from the deployed VM; and record if this test is successful or fails. Note: a test case defined for the RTT calibration test (see section 5.1.5.1.) may be used for this. In that case, only 1 iteration and 1 replica is needed as the actual RTT value is not recorded. 5. Delete VNF (In other to have the same conditions in each iteration the VNF deployed should be removed) 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section "Calculation process and output"

95% percentile service creation time for deploying a virtual instrument on a single compute host

<table>
<tr><th colspan="2">Test Case</th><th>TC-Ser-002</th><th>Service Creation Time</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
95% percentile of the Service Creation Time for deploying virtual instruments on a single compute host
This KPI refers to the time needed to deploy a single VM, which may act as virtual instruments, on a single compute machine. The main intend is to evaluate the capability of the system to deploy a service (VNFs) on an existing, i.e. previously deployed, 4G/5G Network. Since the test case mandates that both virtual machines are deployed on the same compute host, i.e. in a single availability zone, potential effects of deploying VNFs in different zones do not impact the results.
For this test, a simple VM running Debian-9 is deployed. To assure correct deployment of a service offered by the VM, a "ping" is triggered from that machine towards a remote host. The latter step not being part of the time measured. As such, the deployed VNF offers the service "reachability test".
Source A → VM / VNF deployed
Destination B → Remote server in the testbed or outside the testbed
Underlying system → OpenStack
Layer → Network layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring the service creation time, the deployment of a single VM (Debian-9-based) is triggered by the orchestrator of the testbed. The time between triggering the deployment process until successful deployment is indicated is measured.
After the deployment, the VM is used to conduct a "ping" (offered service) originating at the deployed VM towards any reachable remote host. Since the duration of this service test is not part of the measurement, the specification on how to parameterize the "ping" is out of scope of this test case.
The test case shall include the consecutive execution of several iterations according to the following properties.
• Duration of a single iteration → deployment of the single VNF / VM
• Number of replica → At least 25 iterations</td></tr>
</table>

<table>
<tr>
<td>3</td>
<td>

Calculation process and output

The required output should be calculated according to the following methodology:

95% percentile:

Let p95_i be the 95% percentile measured in the ith iteration, and x_(i,n) be the measured service creation time for each VNF/VM deployment within the iteration.

$$p95_i = 95percent\ percentile\ for\ all\ x_{i,n}$$

Then, the (reported) 95% percentile shall be calculated as follows:

$$p95 = \frac{1}{i}\sum_{i} p95_i$$

For the reported 95% percentile, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The RTT output should be provided as:

<table>
<tr><td colspan="3">Service Creation Time [s]</td></tr>
<tr><td rowspan="2">95% percentile</td><td colspan="2">95% confidence interval for 95% percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>

</td>
</tr>
<tr>
<td>4</td>
<td>

Complementary measurements

Deployment Success Rate in % (number of deployments, which resulted in a successful ping afterwards).

</td>
</tr>
<tr>
<td>5</td>
<td>

Pre-conditions

A deployed and working OpenStack facilitating the deployment of VNFs / VMs via an orchestrator (e.g. OpenBaton or OSM).

An existing remote server that can be used to conduct the service test, i.e. to ping.

</td>
</tr>
<tr>
<td>6</td>
<td>

Applicability

This test case applies for all scenarios as it assumes an underlying network infrastructure to deploy the VNFs/VMs in.

</td>
</tr>
</table>

7	Test Case Sequence 1. Trigger the deployment of the VNF/VM from the orchestrator, or from experiment coordinator (TAP) on top of the latter. 2. Wait for a response indicating successful deployment 3. Calculate the deployment time (time between steps 1 and 2) 4. Verify that the deployed service is working, i.e. ping the remote machine from the deployed VM; and record if this test is successful or fails. Note: a test case defined for the RTT calibration test (see section 5.1.5.1.) may be used for this. In that case, only 1 iteration and 1 replica is needed as the actual RTT value is not recorded. 5. Delete VNF (In other to have the same conditions in each iteration the VNF deployed should be removed) 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5.1.9. Speed tests

Reliability tests were not conducted during the 5GENESIS experimentation Phase 1. Corresponding tests will be added in a later revision of this deliverable (D6.2 or D6.3).

5.1.10. Throughput tests

5.1.10.1. Throughput calibration tests

Average (expected mean) Throughput

Test Case		*TC-Thr-001*	*Throughput*
1	Target KPI ***Average Throughput Calibration*** The Throughput calibration test aims to assess the measurement capabilities of the measurement system employed in future Throughput tests. The calibration test employs traffic generation probes and a traffic reception probe, which are communicating with each other in an “empty” SUT. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Source of packets → measurement probe acting as traffic generator Destination of packets → measurement probe acting as recipient Underlying SUT → Network components (if applicable) between the source and the destination Measurement conducted at layer → Application layer		
2	Methodology For measuring Throughput, a packet stream is emitted from a source and received by a data sink (destination). The amount of data (Byte) successfully transmitted per unit of time (seconds) as measured by the traffic generator and the probes shall be recorded.		

<table>
<tr><td></td><td colspan="3">For consistency among calibration tests, a TCP-based traffic stream is created between the source and the destination using the iPerf3 tool. [iPerf.fr]; to reduce impacts of TCP slow-start algorithm, the first 20 s of a measurement are discarded.
The test case shall include the consecutive execution of several iterations according to the following properties.
• Duration of a single iteration → at least three (3) minutes, where the first 20 seconds of measurements are discarded.
• Records throughput over 5-second intervals within an iteration.
• Number of replica → At least 25 iterations.</td></tr>
<tr><td rowspan="7">3</td><td colspan="3"><i>Parameters</i>
The test case shall include the consecutive execution of several replica (iterations) according to the following parameters.</td></tr>
<tr><td>Parameter</td><td>iPerf3 Option</td><td>Suggested Value</td></tr>
<tr><td>Throughput measurement interval</td><td>--interval</td><td>5</td></tr>
<tr><td>Number of simultaneously transmitting probes/ processes/ threads</td><td>--parallel</td><td>1</td></tr>
<tr><td>Bandwidth limitation set to unlimited</td><td>n/a</td><td>Unlimited is the default for iPerf for TCP</td></tr>
<tr><td>Omit first n seconds of the test to skip TCP slowstart</td><td>--omit</td><td>20</td></tr>
<tr><td>Iteration duration</td><td>--time</td><td>180</td></tr>
<tr><td></td><td>Number of iterations</td><td>n/a</td><td>At least 25</td></tr>
<tr><td></td><td>Format to report iPferf results in (report in Mbits/sec)</td><td>--format</td><td>m</td></tr>
<tr><td>3</td><td colspan="3"><i>Calculation process and output</i>
The required output should be calculated according to the following methodology:
Mean (average) Throughput:
Let avg_i be the calculated average throughput for the i^{th} iteration, and x_(i,n) be the measured average throughput over the nth time interval within the ith iteration.
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Then, the overall (reported) average throughput <i>avg</i> shall be calculated as the average of all avg_i
$$avg = \frac{1}{i}\sum_{i} avg_i$$
For the overall average throughput <i>avg</i>, the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.
Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].</td></tr>
</table>

<table>
<tr><td></td><td>When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The Output shall be provided as:
<table>
<tr><td colspan="3">Throughput [Mbit/s]</td></tr>
<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>
</table>
</td></tr>
<tr><td>4</td><td>Complementary measurements
PL Rate</td></tr>
<tr><td>5</td><td>Pre-conditions
The scenario has been configured.
The traffic generator should support the generation of the traffic pattern defined in the Throughput traffic patterns section.
Ensure that no undesired traffic is present during the test.</td></tr>
<tr><td>6</td><td>Applicability

The measurement probes need to be capable of injecting traffic into the system and assessing successful or unsuccessful transmission of the data, as well as determining the throughput of the transmission.</td></tr>
<tr><td>7</td><td>Test Case Sequence
Start monitoring probes (deployment of propes running iPerf client and server).
Using the traffic generator, begin transmitting from the client probe to the server probe using one of the Throughput traffic patterns defined.
Record the Throughput for each time interval within a trial.
Stop the traffic generator.
Stop monitoring probes
Calculate and record the KPIs as needed per iteration as defined in “Calculation process and output”.
Repeat steps 1 to 6 for each one of the 25 iterations
Compute the KPIs as defined in section “Calculation process and output”</td></tr>
</table>

DL/UL Peak Throughput

<table>
<tr><th colspan="2">Test Case</th><th>TC-Thr-002</th><th>Throughput</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
DL/UL Peak Throughput (Speedtest)
The DL/UL Peak Throughput KPI is the maximum Downlink and Uplink Throughput between a client and an external Ookla Speedtest server over a MN[12].
Throughput measurement using Speedtest reflects the network experience when conducting a consumer-initiated test.[13]
Source → External Ookla Server (DL) / UE (UL)
Sink → UE (DL) / External Ookla Server (UL)
Underlying system → UE – RAN (LTE only or 5G-NR NSA) – EPC – External Packet Data Network
Layer → Application</td></tr>
<tr><td>2</td><td colspan="3">Methodology
The experimenter shall install Ookla's Speedtest Application on the UE.
The methodology for sending data streams to the sink shall follow Ookla's default provision.
The data traffic type shall be TCP.
The experiment includes the execution of N≥2 iterations, according to Speedtest's default parameters.</td></tr>
<tr><td>3</td><td colspan="3">Parameters
The experimenter shall configure the following parameters in the Speedtest Application:
Mode: Multiple Connections.
This parameter shows the maximum potential throughput by using multiple streams in parallel.
External Ookla Server: Select the nearest available Server in the experimenter's location.
This is mostly due to the fact that parameters (such as the TCP window size) controlling the transfer are not optimized for the increased latency that comes from an increase in distance.</td></tr>
<tr><td>4</td><td colspan="3">Calculation process and output
Speedtest calculates Throughput according to a predefined process devised by Ookla.
The overall (reported) average Throughput shall be calculated as the average of the throughput values R_n(T) measured over N iterations:
$$Average\ Throughput = \frac{1}{N}\sum_{n=1}^{N} R\_n(T)$$
The 95% confidence interval shall be reported using the Student-T-distribution for $v = N - 1$ degrees of freedom.
Throughput results shall be reported as follows:<table><tr><th>Throughput [Mbit/s]</th></tr></table></td></tr>
</table>

---

[12] https://www.speedtest.net/

[13] https://www.speedtest.net/about/knowledge/test-methods

<table>
<tr><td></td><td>
<table>
<tr><th></th><th>Lower Bound of 95% confidence interval</th><th>Mean</th><th>Upper Bound of 95% confidence interval</th></tr>
<tr><td>DL</td><td></td><td></td><td></td></tr>
<tr><td>UL</td><td></td><td></td><td></td></tr>
</table>
</td></tr>
<tr><td>5</td><td>Complementary measurements
The experimenter shall record the following Complementary Measurements, as reported by Speedtest:
Jitter (ms)
Ping (ms)
The overall (reported) average value of each metric shall be calculated as the average of the values measured over N iterations.
The 95% confidence interval for each metric shall also be reported using the Student-T-distribution for $v = N - 1$ degrees of freedom.</td></tr>
<tr><td>5</td><td>Pre-conditions
The scenario has been configured.
Ensure that no undesired traffic is present during the test.
There is only one connected UE to the network.</td></tr>
<tr><td>6</td><td>Applicability
The measurement probes need to be capable of injecting traffic into the system and assessing successful or unsuccessful transmission of the data, as well as determining the throughput of the transmission.</td></tr>
<tr><td>7</td><td>Test Case Sequence
Launch the “Speedtest” application at the UE.
Define the appropriate parameters as described in Section “*Parameters*” of this Test Case.
Initiate the measurement through Speedtest.
Record the Downlink / Uplink Throughput and the Complementary Measurements results.
Repeat Steps 1-4 according to the number of iterations.
Compute the average values of the Primary and Secondary KPIs, as defined in Section “Calculation Process and Output”.</td></tr>
</table>

Adaptive HTTP Streaming Throughput

<table>
<tr><th colspan="2">Test Case</th><th>TC-Thr-003</th><th>Throughput</th></tr>
<tr><td>1</td><td colspan="3">Target KPI
Adaptive HTTP Streaming Throughput
Throughput is calculated as the number of received bits per second. Only the TCP payload is considered, without the Layer 2 or IP headers (this is referred to as Goodput in RFC 2647).[14]
In the Adaptive Video Streaming over HTTP, the video is encoded on a server at several bitrates and divided into video segments of 2 to 10 seconds.[14]
The client is informed of the available bitrates by downloading a video manifest and then requests video segments from the server at a specific bitrate.
Depending on the network bandwidth and the local speed of video decoding and rendering, the client can decide to increase (upshift) or decrease (downshift) the requested bitrate for future video segments.
Reference Points:
Endpoint A → UE
Endpoint B → Server connected to SGi Interface (LTE) or N6 interface (5G)
Underlay system → UE – RAN (LTE only or 5G-NR) – Core Network (EPC or 5GC)</td></tr>
<tr><td>2</td><td colspan="3">Methodology
The experimenter shall configure Adaptive Video over HTTP Traffic to be sent over the endpoints (e.g. YouTube 1080p). The type of traffic will be reported in the test report.
The performance test duration shall be 3 minutes, where throughput shall be recorded every 2 seconds.
The experiment includes the execution of N≥2 iterations.</td></tr>
<tr><td>3</td><td colspan="3">Parameters
The experimenter shall configure the following parameters[14], according to the following table. In case any parameter value is changed, the value shall be reported in the final test report.
<table>
<tr><th>Parameter</th><th>Default Values</th><th>Comments</th></tr>
<tr><td>Segment duration (s)</td><td>2</td><td>N/A</td></tr>
<tr><td>Bitrates list (Mbps)</td><td>0.080 0.350 0.520 0.830 1.600 3.000 10.000 25.000</td><td>Series of bitrate levels in ascending order. Their value shall not exceed 1Gbps.</td></tr>
<tr><td>Shift Bias</td><td>Neutral</td><td>N/A</td></tr>
<tr><td>Shift Sensitivity</td><td>Medium</td><td>N/A</td></tr>
<tr><td>Starting Bitrate</td><td>Middle</td><td>N/A</td></tr>
</table></td></tr>
</table>

---

[14] Ixria IxChariot User Guide, October 2018, Software Version 9.6 SP1

<table>
<tr><td>4</td><td>

Calculation process and output

The overall (reported) average Throughput shall be calculated as the average of the throughput values R_n(T) measured over N iterations:

$$Average\ Throughput = \frac{1}{N}\sum_{n=1}^{N} R\_n(T)$$

The 95% confidence interval shall be reported using the Student-T-distribution for *v = N - 1* degrees of freedom.

Throughput results shall be reported as follows:

<table>
<tr><th colspan="3">Throughput [Mbps]</th></tr>
<tr><th>Lower Bound of 95% confidence interval</th><th>Mean</th><th>Upper Bound of
95% confidence interval</th></tr>
<tr><td></td><td></td><td></td></tr>
</table>

</td></tr>
<tr><td>5</td><td>

Complementary measurements

The experimenter may also record the following Complementary Measurements:

- RSSI (dBm).
- Video stopped counter (The number of times the video playback stopped).
- Average HTTP Video Rate (Mbps).
- Number of video segment sent with specific quality.

</td></tr>
<tr><td>6</td><td>

Pre-conditions

- The scenario has been configured.
- Ensure that no undesired traffic is present during the test.

</td></tr>
<tr><td>7</td><td>

Applicability

The measurement probes need to be capable of injecting traffic into the system and assessing successful or unsuccessful transmission of the data, as well as determining the throughput of the transmission.

</td></tr>
<tr><td>8</td><td>

Test Case Sequence

1. Launch the monitoring probes at the source and destination.
2. Define the appropriate parameters as described in Section “*Parameters*” of this Test Case.
3. Initiate the measurement through the Traffic Generator, as described in the “*Methodology Section*”.
4. Record the reported Throughput and Complementary Measurements results.
5. Repeat Steps 1-4 according to the number of iterations.
6. Compute the average values of the Primary and Secondary KPIs, as defined in Section “Calculation Process and Output”.

</td></tr>
</table>

Average Maximum User Data Rate

<table>
<tr><th colspan="2">Test Case</th><th>TC-Thr-004</th><th>Throughput</th></tr>
<tr><td>1</td><td colspan="3">
<p>Target KPI</p>
<p>Average Maximum User Data Rate Calibration</p>
<p>The maximum (max) user data rate calibration test aims to assess the measurement capabilities of the measurement system employed in future maximum user data rate tests.</p>
<p>The calibration test employs traffic generation probes and a traffic reception probe, which are communicating with each other in an “empty” SUT.</p>
<p>Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected.</p>
<p>Source of packets → measurement probe acting as traffic generator.</p>
<p>Destination of packets → measurement probe acting as recipient.</p>
<p>Underlying SUT → Network components between the source and destination.</p>
<p>Measurement conducted at layer → Application layer.</p>
</td></tr>
<tr><td>2</td><td colspan="3">
<p>Methodology</p>
<p>For measuring Throughput, a packet stream is emitted from a source and received by a data sink (destination). The amount of data (Byte) successfully transmitted per unit of time (seconds) as measured by the traffic generator and the probes shall be recorded.</p>
<p>For consistency among calibration tests, a UDP-based traffic stream is created between the source and the destination using the iPerf2 tool. [iPerf.fr]</p>
<p>The test case shall include the consecutive execution of several iterations according to the following properties.</p>

Duration of a single iteration → at least three (3) minutes.
Records throughput over 5-second intervals within an iteration.
Number of replicas → At least 25 iterations.

</td></tr>
<tr><td>3</td><td colspan="3">
<p>Parameters</p>
<p>The test case shall include the consecutive execution of several replica (iterations) according to the following parameters.</p>
<table>
<tr><th>Parameter</th><th>iPerf Option</th><th>Suggested Value</th></tr>
<tr><td>Throughput measurement interval</td><td>--interval</td><td>5</td></tr>
<tr><td>Number of simultaneously transmitting probes/ processes/ threads</td><td>--parallel</td><td>4 (in order to reach higher data rate)</td></tr>
<tr><td>Bandwidth limitation set to above the maximum bandwidth available</td><td>--b</td><td>Depends on the maximum theoretical throughput available the in network scenario</td></tr>
<tr><td>Iteration duration</td><td>--time</td><td>180</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>Number of iterations</td><td>n/a</td><td>At least 25</td></tr>
<tr><td></td><td>Format to report iPferf results in (report in Mbits/sec)</td><td>--format</td><td>m</td></tr>
<tr><td>4</td><td colspan="3">

Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) maximum user data rate:

Let avg_i be the calculated average max user data rate for the i^{th} iteration, and x_(i,n) be the measured average max user data rate over the nth time interval within the i^{th} iteration.

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average max user data rate *avg* shall be calculated as the average of all avg_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average max user data rate *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

Maximum user data rate output should be provided as:

<table>
<tr><td colspan="3">Max user data rate [Mbit/s]</td></tr>
<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>
</table>

</td></tr>
<tr><td>5</td><td colspan="3">

Complementary measurements

Note: PL rate is not recorded because constant traffic in excess of the available capacity will be used and the excess will be discarded, as expected.

</td></tr>
<tr><td>6</td><td colspan="3">

Pre-conditions

- The scenario has been configured.
- The traffic generator should support the generation of the traffic pattern defined in the max user data rate traffic patterns section.
- Ensure that no undesired traffic is present during the test.

</td></tr>
<tr><td>7</td><td colspan="3">

Applicability

The measurement probes need to be capable of injecting traffic into the system as well as determining the throughput of the transmission.

</td></tr>
<tr><td>8</td><td colspan="3">

Test Case Sequence

1. Start monitoring probes (deployment of propes running iPerf client and server).

</td></tr>
</table>

	2. Using the traffic generator, begin transmitting from the client probe to the server probe using one of the Max User Data Rate traffic patterns defined. 3. Record the Throughput for each time interval within a trial. 4. Stop the traffic generator. 5. Stop monitoring probes 6. Calculate and record the KPIs as needed per iteration as defined in “Calculation process and output”. 7. Repeat steps 1 to 6 for each one of the 25 iterations 8. Compute the KPIs as defined in section “Calculation process and output”

5.1.11. Ubiquity/Coverage tests

Note: in this deliverable Ubiquity and Coverage are terms that are used interchangeably.

5.1.11.1. Ubiquity/Coverage calibration tests

RAN coverage calibration test

Test Case		*TC-Ubi-001*	*RAN coverage calibration*
1	*Target KPI* ***RAN Coverage calibration test*** The aim of this test is to verify the proper operation of the SUT (E2E 5G network) and of the measurement methodology. More specifically, it aims at verifying that the network is operating properly (at least enabling basic data exchange) under ideal transmission conditions. It also aims to observe how QoS is degrading while the coupling loss between the RAN and the UE increases.T		
2	*Methodology* These tests use a simple approach involving ICMP ECHO messages to verify basic data connectivity. ICMP ECHO messages are exchanged between a couple of measurement probes (virtual or physical) installed at i) the UE and ii) a physical server or a VM behind the core network (EPC/5GC) respectively. In any case, the probe at the UE will issue an ICMP ECHO request and the probe at the core will reply with an ICMP ECHO reply. The TD-001 traffic profile will be used. ICMP PL will be measured at the request issuer (UE). The test includes two phases: Phase (i) includes measurements under ideal transmission conditions, while Phase (ii) includes measurements during which the SINR is gradually artificially degraded (see later Sec. “Test case sequence” for more details). Both the gNB/eNB remain stationary during both phases at the original distance (1m).		
3	*Calculation process and output* The calibration test is considered completed and successful if both the following conditions apply: For Step (ii): The loss rate increases with the SINR reduction. This is verified by: a) Plotting the decimal logarithm of PL against SINR, using the measured values b) Obtaining a simple linear regression estimator for Log(PL) as a function of SINR, i.e. $Log(PL) = a \cdot SINR + b$ c) Verifying that a is negative.		

4	*Complementary measurements* Along with PL, ICMP RTT (average, min, max and standard deviation) will be recorded. There should be no significant deviations in the RTT, especially for Phase (i) Also, RSSI and RSRQ will be measured, both at the UE and the eNB/gNB.
5	*Pre-conditions* The following pre-conditions should apply prior to the beginning of the tests: • The network will serve no other user traffic (verified in eNB/gNB monitoring) • A single UE will be connected • The UE will be stationary and at a distance of 1m from the eNB/gNB antenna, at line-of-sight conditions. Exact RAN parameters and configuration (antenna gain, band, bandwidth, other PHY configuration parameters) etc. will depend on the actual setup/scenario, yet they should be consistent across all tests for the specific scenario.
6	*Applicability* N/A
7	*Test Case Sequence* The test is conducted in two phases: i) Phase 1 - Verification of proper operation: 1. Perform one iteration (100 ICMP requests) under ideal transmission conditions (10 m distance), using maximum available RF power at eNB/gNB. 2. Verify that PL is no more than 0% ii) Phase 2: Analysis of service degradation: 1. Reduce the SINR (Signal to Interference and Noise Ratio) by 2 dB, by either of the following means: i. RF attenuator between the eNB/gNB and its antenna ii. Mixing the signal with the output of an external noise generator iii. Increasing the eNB/gNB – UE distance 2. Conduct one iteration (100 ICMP requests) and record the PL 3. If PL is less than 100%, repeat (1) 4. Assess service degradation using linear interpolation as described under “Calculation process and output”

Backhaul coverage calibration test

Test Case	*TC-Ubi-002*	*Backhaul coverage calibration*
1	*Target KPI* ***RAN Coverage calibration test*** Backhaul coverage mostly applies to cases where a wide-area wireless backhaul network is used to support a mobile 5G access network (“hotspot”). The aim of the backhaul coverage calibration text is to verify the proper operation of the system under test (E2E 5G network), as well as of the measurement methodology. More specifically, the calibration	

	test aims at verifying that the network is operating properly (at least enabling basic data exchange) under ideal transmission conditions in the backhaul. It also aims to observe how QoS is degrading while the coupling loss in the backhaul link increases.
2	*Methodology* These tests use a simple approach involving ICMP ECHO messages to verify basic data connectivity. ICMP ECHO messages are exchanged between a couple of measurement probes (virtual or physical) installed at i) the UE and ii) behind the core network (EPC/5GC) respectively. In any case, the probe at the UE will issue an ICMP ECHO request and the probe at the core will reply with an ICMP ECHO reply. The TD-001 traffic profile will be used. ICMP PL will be measured at the request issuer (UE). The test includes two phases: Phase (i) includes measurements under ideal transmission conditions in the backhaul link, while Phase (ii) includes measurements during which the SINR in the backhaul link is gradually artificially degraded (see later Sec. "Test case sequence" for more details) Both the backhaul antennas, as well as gNB/eNB remain stationary during both phases.
3	*Calculation process and output* The calibration test is considered completed and successful if both the following conditions apply: For Step (i): Loss rate is no more than 0% for the ideal conditions. For Step (ii): The loss rate increases with the backhaul SINR reduction. This is verified by: a) Plotting PL against SINR, using the measured values b) Obtaining a simple linear regression estimator for PL as a function of SINR, i.e. $$PL = a \cdot SINR + b$$ c) Verifying that a is negative.
4	*Complementary measurements* Along with PL, ICMP RTT (average, min, max and standard deviation) will be recorded. Also, RSSI and RSRQ (or equivalent values, depending on the wireless backhaul technology) will be measured, at both radio units of the wireless backhaul link.
5	*Pre-conditions* The following pre-conditions should apply prior to the beginning of the tests: • The network will serve no other user traffic (verified in eNB/gNB and also core network monitoring) • A single UE will be connected • The UE will be stationary and at a distance of 1m from the eNB/gNB antenna, at line-of-sight conditions (i.e. the RAN is considered stable) • The wireless backhaul connection will be reliable and stable, with SINR/RSSP well above (at least 10 dB) operating thresholds. Exact backhaul parameters and configuration (antenna gain, band, bandwidth, other PHY configuration parameters) etc. will depend on the actual setup/scenario, yet they should be consistent across all tests for the specific scenario.

6	*Applicability* N/A
7	*Test Case Sequence* The test is conducted in two phases: iii) Phase 1 - Verification of proper operation: 1. Perform one iteration (100 ICMP requests) under stable backhaul link conditions, with SINR/RSSP well above (at least 10 dB) above operating thresholds. 2. Verify that E2E PL is no more than 0% iv) Phase 2: Analysis of service degradation: 1. Reduce the backhaul SINR (Signal to Interference&Noise Ratio) by 2 dB, by either of the following means: i. RF attenuator between the backhaul unit and the antenna (both Rx and Tx). ii. Mixing the signal with the output of an external noise generator. iii. Increasing the distance between the backhaul units. 2. Conduct one iteration (100 ICMP requests) and record the PL. 3. If PL is less than 100%, repeat (1). 4. Assess service degradation using linear interpolation as described under “Calculation process and output”.

5.1.11.2. Ubiquity/Coverage tests

RAN coverage test

Test Case		*TC-Ubi-003*	*RAN Coverage*
1	*Target KPI* ***Ubiquity/Coverage***		
2	*Methodology* While coverage measurements, in a degree, depend on the specific service/application to be considered, for the tests in 5GENESIS we will use a simple approach involving ICMP ECHO messages to verify basic data connectivity. ICMP ECHO messages are exchanged on E2E basis between a couple of measurement probes (virtual or physical) installed at i) the UE and ii) behind the core network (EPC/5GC) respectively. In any case, the probe at the UE will issue an ICMP ECHO request and the probe at the core will reply with an ICMP ECHO reply. 1400-byte packets will be used, with the specifications defined in Sec. 6.1.3. One iteration consists of 100 consecutive requests, sent at a rate of 2 requests/sec. ICMP PL will be measured at the request issuer (UE). Measurements are taken into various UE locations (see later “Test case sequence”).		
3	*Calculation process and output* We consider n measurements at n different locations (see below “Test case sequence”). For each location i, we perform a measurement iteration and we assume: $x_i = 0$ if the location is considered out of coverage, i.e. ICMP PL is above a specified threshold (we assume 5%) $x_i = 1$ if the location is considered in coverage, i.e. ICMP PL is below a specified threshold (we assume 5%).		

	Then the coverage KPI (as percentage) is calculated as $$C = \frac{1}{n}\sum_{n} x_i \cdot 100\%$$
4	*Complementary measurements* Along with PL, ICMP RTT (average, min, max and standard deviation) will be recorded. Also, RSSI and RSRQ will be measured, both at the UE and the eNB/gNB.
5	*Pre-conditions* The following pre-conditions should apply prior to the beginning of the tests: • The calibration tests will have been completed and successful (see Section 5.1.11.1.) • The network will serve no other user traffic (verified in eNB/gNB monitoring) • A single UE will be connected Exact RAN parameters and configuration (antenna gain, band, bandwidth, other PHY configuration parameters) etc. will depend on the actual setup/scenario, yet they should be consistent across all tests for the specific scenario.
6	*Applicability* N/A
7	*Test Case Sequence* 1. Divide the area under consideration into a grid of equally spaced locations 2. With the gNB/eNB fixed, move the UE in location i 3. Perform a measurement iteration. 4. The location is considered out of coverage, i.e. ICMP PL is above a specified threshold (we assume 5%), it is considered in coverage otherwise. 5. Repeat steps (2-4) for all locations in the area 6. Calculate coverage as defined above ("Calculation process")

Backhaul coverage test

Test Case		*TC-Ubi-004*	*Backhaul coverage*
1	*Target KPI* ***Ubiquity/Coverage*** The backhaul coverage test addresses cases where a wide-area wireless backhaul network is used to support a mobile 5G access network ("hotspot").		
2	*Methodology* While coverage measurements, in a degree, depend on the specific service/application to be considered, for the tests in 5GENESIS we will use a simple approach involving ICMP ECHO messages to verify basic data connectivity. ICMP ECHO messages are exchanged on E2E basis between a couple of measurement probes (virtual or physical) installed at i) the UE and ii) behind the core network (EPC/5GC) respectively.		

	In any case, the probe at the UE will issue an ICMP ECHO request and the probe at the core will reply with an ICMP ECHO reply. 1400-byte packets will be used, with the specifications defined in Sec. 6.1.3. One iteration consists of 100 consecutive requests, sent at a rate of 2 requests/sec. ICMP PL will be measured at the request issuer (UE). In the backhaul coverage test, we consider the UE to be co-located with the (mobile) eNB/gNB in order to secure RAN coverage and to exclude outages which may be due to poor RAN signal. We assume the UE to be at a distance of max. 1m from the eNB/gNB, under LoS conditions. That is, the UE moves around together with the mobile 5G remote network (hotspot). Measurements are taken into various hotspot locations (see later "Test case sequence").
3	*Calculation process and output* We consider n measurements at n different locations (see below "Test case sequence") of the remote 5G network (hotspot). For each location i, we perform a measurement iteration and we assume: $x_i = 0$ if the location is considered out of coverage, i.e. ICMP PL is above a specified threshold (we assume 5%). $x_i = 1$ if the location is considered in coverage, i.e. ICMP PL is below a specified threshold (we assume 5%). Then the backhaul coverage KPI (as percentage) is calculated as $C = \frac{1}{n}\sum_{n} x_i \cdot 100\%$
4	*Complementary measurements* Along with packet loss, ICMP round trip time (average, min, max and standard deviation) will be recorded. Also, RSSI and RSRQ (or equivalent values, depending on the wireless backhaul technology) will be measured, at both radio units of the wireless backhaul link.
5	*Pre-conditions* The following pre-conditions should apply prior to the beginning of the tests: • The calibration tests will have been completed and successful (see Sec. 5.1.11.1.). • The network will serve no other user traffic (verified in eNB/gNB monitoring). • A single UE will be connected. • The UE will be stationary and at a distance of 1 m from the eNB/gNB antenna, at line-of-sight conditions (i.e. the RAN is considered stable). Exact backhaul parameters and configuration (antenna gain, band, bandwidth, other PHY configuration parameters) etc. will depend on the actual setup/scenario, yet they should be consistent across all tests for the specific scenario.
6	*Applicability* N/A
7	*Test Case Sequence* 1. Divide the area of the wireless backhaul network coverage under consideration into a grid of equally spaced locations. The backhaul radio node which is connected to the core network side is considered fixed.

	2. Move the second backhaul radio node, along with the entire 5G hotspot and the UE, in location i. If necessary, re-align the backhaul network antennas. 3. Perform a measurement iteration. 4. The location is considered out of backhaul coverage, i.e. ICMP PL is above a specified threshold (we assume 5%), it is considered in coverage otherwise. 5. Repeat steps (2-4) for all locations in the area 6. Calculate backhaul coverage as defined above ("Calculation process")

5.1.12. MCPTT tests

5.1.12.1. Average (expected mean) MCPTT access time test

Test Case		*TC-MAL-UC3-001*	*MCPTT*
1	Target KPI ***Average (expected mean) MCPTT access time Calibration*** The MCPTT access time calibration tests aims at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests. The MCPTT access time is defined as the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking. This time does not include confirmations from receiving users. The calibration test employs floor control messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an "empty" SUT. The calibration test is conducted at the application layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, they shall, preferably, be directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of floor control packets →measurement probe acting as client Destination of floor control packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Application layer		
2	Methodology The measurement procedure only requires to have an ongoing call, and requesting and releasing the token. The request is logged when the MCPTT request event is created with the next message: KPI1_PERFORMANCE,TOKEN REQUEST,currentTime On the other hand, the granted state (in case the token is granted to the requested user after a token idle situation) is logged with the next message: KPI1_PERFORMANCE,TOKEN GRANTED,currentTime Using the logged timestamps, we can calculate the accurate time that MCPTT access takes.		

<table>
<tr><td></td><td>The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
- Duration of a single replica (iteration) → the duration has to ensure that at least 20 KPI1_PERFORMANCE,TOKEN REQUEST,currentTime are sent during a single replica (iteration), and that the same amount of KPI1_PERFORMANCE,TOKEN GRANTED,currentTime are responded.
- Number of replica (iterations) → At least 25</td></tr>
<tr><td>3</td><td><i>Calculation process and output</i>
The required output should be calculated according to the following methodology:
Mean (average) MCPTT access time:
Let avg_i be the calculated average MCPTT access time for the ith replica (iteration), and x_(i,n) be the measured MCPTT access time for each repetition within the replica (iteration).
$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$
Then, the overall (reported) average MCPTT access time <i>avg</i> shall be calculated as the average of all x_i
$$avg = \frac{1}{i}\sum_{i} avg_i$$
For the overall average MCPTT access time <i>avg</i>, the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.
Note: This methodology accounts for non-Gaussian distributions of MCPTT access times within a single trial [5].
When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.
The MCPTT access time delay output shall be provided as:
<table>
<tr><td colspan="2">MCPTT access time [ms]</td></tr>
<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>
</table></td></tr>
<tr><td>4</td><td>Pre-conditions
In order to start the execution of this test, there must be an ongoing call.</td></tr>
<tr><td>5</td><td>Applicability
Measurement of 3GPP standardized MCPTT access time delay KPI.
This test case must be executed using MCPTT 3GPP compliant UEs.</td></tr>
</table>

6	Test Case Sequence 1. Start the MCPTT system. 2. Establish an MCPTT call. 3. Perform the token request, in order to generate the desired messages, which will be recorded. 4. Repeat the token or floor request and release procedure or finish the call. 5. Calculate and record the average MCPTT access time per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations 7. Compute the KPIs as defined in section “Calculation process and output”

5.1.12.2. 95%-percentile MCPTT access time test

Test Case		*TC-MAL-UC3-002*	*MCPTT*
1	Target KPI ***95%-percentile MCPTT access time Calibration*** The MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests. The MCPTT access time is defined as the time between when an MCPTT User requests to speak and when this user gets a signal to start speaking. This time does not include confirmations from receiving users. The calibration test employs floor control messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the application layer. Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected. Experimenters should include as part of their test case description an illustration of the measurement system, including if applicable and potential virtualization aspects. Source of floor control packets →measurement probe acting as client Destination of floor control packets → measurement probe acting as server Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Application layer		
2	Methodology The measurement procedure only requires having an ongoing call, and requesting and releasing the token. The request is logged when the MCPTT request event is created with for instance, the next message: KPI1_PERFORMANCE,TOKEN REQUEST,currentTime On the other hand, the granted state (in case the token is granted to the requested user after a token idle situation) is logged with for example, the next message: KPI1_PERFORMANCE,TOKEN GRANTED,currentTime Using the logged timestamps, we can calculate the accurate time that MCPTT Access takes.		

<table>
<tr><td></td><td>The test case shall include the consecutive execution of several replica (iterations) according to the following properties.
- Duration of a single replica (iteration) → the duration has to ensure that at least 20 KPI1_PERFORMANCE,TOKEN REQUEST,currentTime are sent during a single replica (iteration), and that the same amount of KPI1_PERFORMANCE,TOKEN GRANTED,currentTime are responded.
- Number of replica (iterations) → At least 25</td></tr>
<tr><td>3</td><td>*Calculation process and output*
The required output should be calculated according to the following methodology:
95%-percentile MCPTT access time:
Let MCPTTAT 95p_i be the 95%-percentile MCPTT access time measured in the ith replica (iteration), and x_(i,n) be the measured MCPTT access time for each repetition within the replica (iteration).
$$MCPTTAT_{95p_i} = 95\%\ \text{percentile}(x_{i,n})$$
Then, the (reported) 95%-percentile MCPTT access time (MCPTTAT_95p) shall be calculated as follows:
$$MCPTTAT\_95p = \frac{1}{i}\sum_{i} MCPTTAT\_95p_i$$
For the reported 95%-percentile MCPTT access time ($MCPTTAT\_95p$), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.
The MCPTT Access time delay output should be provided as:
<table>
<tr><td colspan="3">MCPTT access time [ms]</td></tr>
<tr><td rowspan="2">95% Percentile</td><td colspan="2">95% confidence interval for 95% percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>Pre-conditions
In order to start the execution of this test, there must be an ongoing call.</td></tr>
<tr><td>5</td><td>Applicability
Procedure to measure 3GPP standardized Access time delay.
This test case must be executed using MCPTT 3GPP compliant UEs.</td></tr>
<tr><td>6</td><td>Test Case Sequence
1. Start the MCPTT system.
2. Establish an MCPTT call.
3. Perform the token request, in order to generate the desired messages, which will be recorded.
4. Either request and release the token or floor again or finish the call.
5. Calculate and record the 95%-percentile MCPTT access time per iteration as defined in “Calculation process and output”.
6. Repeat steps 1 to 5 for each one of the 25 iterations.</td></tr>
</table>

	7. Compute the KPIs as defined in section “Calculation process and output”.

5.1.12.3. Average (expected mean) MCPTT E2E access time test

Test Case		TC-MAL-UC3-003	MCPTT
1	Target KPI *Average (expected mean) E2E MCPTT access time Calibration* The E2E MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further E2E MCPTT access time tests. According to the standard the E2E MCPTT Access time is defined as the time between when an MCPTT User requests to speak and when this user gets a signal to star speaking, including MCPTT call establishment and possibly acknowledgment from first receiving user before voice can be transmitted. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the application layer. Source of INVITE packet and destination of confirmation packet →measurement probe acting as client 1. Destination of INVITE packet and source of confirmation packet → measurement probe acting as client 2. Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT. Measurement conducted at layer → Application layer.		
2	Methodology In order to get the accurate time of the E2E MCPTT Access time delay we log the function that sends the INVITE and the one which receives the 200 OK relative to that INVITE from the MCPTT/MCS system, after the confirmation by the callee. To ensure that the 200 OK is the response to the sent INVITE by the caller we have logged the Command sequence (Cseq) field of the INVITE and 200 OK messages. For instance, in the following way: KPI2_PERFORMANCE,INVITE,currentTime,cseq KPI2_PERFORMANCE,200 OK,currentTime,cseq The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → the duration has to ensure that at least 20 KPI2_PERFORMANCE,INVITE,currentTime,cseq are sent during a single replica (iteration), and that the same amount of KPI2_PERFORMANCE,200 OK,currentTime,cseq are responded. • Number of replica (iterations) → At least 25		
3	*Calculation process and output* The required output shall be calculated according to the following methodology: **Mean (average) E2E MCPTT access time**:		

<table>
<tr><td></td><td>
Let avg_i be the calculated average E2E MCPTT access time for the ith replica (iteration), and x_(i,n) be the measured E2E MCPTT access time for each repetition within the replica (iteration).

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average E2E MCPTT access time *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average E2E MCPTT access time *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of E2E MCPTT access times within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

E2E MCPTT access time delay output should be provided as:

<table>
<tr><td colspan="3">E2E MCPTT access time [ms]</td></tr>
<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>
</table>
</td></tr>
<tr><td>4</td><td>
Pre-conditions

The service MCPTT/MCS VNFs should be instantiated, the whole set of services up and running, the UE connected to an actual RAT that is able to reach the deployed VNFs going through a core network.
</td></tr>
<tr><td>5</td><td>
Applicability

Measurement of 3GPP standardized E2E MCPTT access time delay

This test case must be executed using MCPTT 3GPP compliant UEs.
</td></tr>
<tr><td>6</td><td>
Test Case Sequence

1. Start the MCPTT system.
2. Establish an MCPTT call.
3. Finish the call.
4. Calculate and record the average E2E MCPTT access time per iteration as defined in “Calculation process and output”.
5. Repeat steps 1 to 3 for each one of the 25 iterations
6. Compute the KPIs as defined in section “Calculation process and output”
</td></tr>
</table>

5.1.12.4. 95%-percentile MCPTT E2E access time test

Test Case	*TC-MAL-UC3-004*	*MCPTT*
1	Target KPI ***95%-percentile E2E MCPTT access time Calibration*** The E2E MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further E2E MCPTT access time tests. According to the standard the E2E MCPTT Access time is defined as the time between when an MCPTT User requests to speak and when this user gets a signal to star speaking, including MCPTT call establishment and possibly acknowledgment from first receiving user before voice can be transmitted. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the application layer. Source of INVITE packet and destination of confirmation packet →measurement probe acting as client 1 Destination of INVITE packet and source of confirmation packet → measurement probe acting as client 2 Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Application layer	
2	Methodology In order to get the accurate time of the E2E MCPTT Access time delay we log the function that sends the INVITE and the one which receives the 200 OK relative to that INVITE from the MCPTT/MCS system, after the confirmation by the callee. To ensure that the 200 OK is the response to the sent INVITE by the caller we have logged the Command sequence (Cseq) field of the INVITE and 200 OK messages. For instance, in the following way: KPI2_PERFORMANCE,INVITE,currentTime,cseq KPI2_PERFORMANCE,200 OK,currentTime,cseq The test case shall include the consecutive execution of several replica (iterations) according to the following properties. • Duration of a single replica (iteration) → the duration has to ensure that at least 20 KPI2_PERFORMANCE,INVITE,currentTime,cseq are sent during a single replica (iteration), and that the same amount of KPI2_PERFORMANCE,200 OK,currentTime,cseq are responded. • Number of replica (iterations) → At least 25	
3	*Calculation process and output* The required output shall be calculated according to the following methodology: **95%-percentile E2E MCPTT access time**: Let E2EMCPTTAT 95p_i be the 95%-percentile E2E MCPTT access time measured in the ith replica (iteration), and x_(i,n) be the measured E2E MCPTT access time for each repetition within the replica (iteration). $E2EMCPTTAT_{95p_i} = 95\%\ \mathrm{percentile}(x_{i,n})$	

<table>
<tr><td></td><td>
<p>Then, the (reported) 95%-percentile E2E MCPTT access time (E2EMCPTTAT_95p) shall be calculated as follows:</p>

$$E2EMCPTTAT\_95p = \frac{1}{i}\sum_{i} E2EMCPTTAT\_95p_i$$

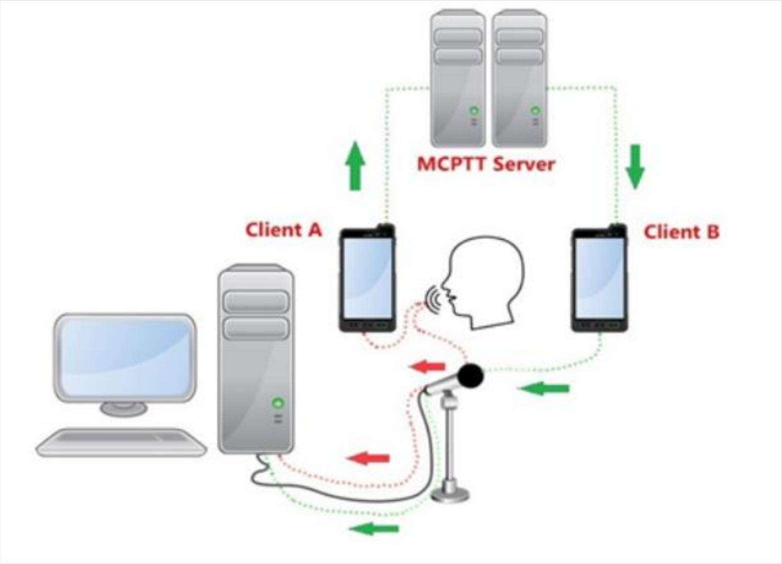
<p>For the reported 95%-percentile E2E MCPTT access time ($E2EMCPTTAT\_95p$), the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.</p>
<p>E2E MCPTT access time delay output shall be provided as:</p>
<table>
<tr><td colspan="3">E2E MCPTT access time [ms]</td></tr>
<tr><td rowspan="2">95% Percentile</td><td colspan="2">95% confidence interval for 95% Percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td></tr>
<tr><td>4</td><td>
<p>Pre-conditions</p>
<p>The service MCPTT/MCS VNFs should be instantiated, the whole set of services up and running, the UE connected to an actual RAT that is able to reach the deployed VNFs going through a core network.</p>
</td></tr>
<tr><td>5</td><td>
<p>Applicability</p>
<p>Measurement of 3GPP standardized E2E MCPTT access time delay.</p>
<p>This test case must be executed using MCPTT 3GPP compliant UEs.</p>
</td></tr>
<tr><td>6</td><td>
<p>Test Case Sequence</p>

Start the MCPTT system.
Establish an MCPTT call.
Finish the call.
Calculate and record the 95%-percentile E2E MCPTT access time per iteration as defined in “Calculation process and output”.
Repeat steps 1 to 3 for each one of the 25 iterations
Compute the KPIs as defined in section “Calculation process and output”

</td></tr>
</table>

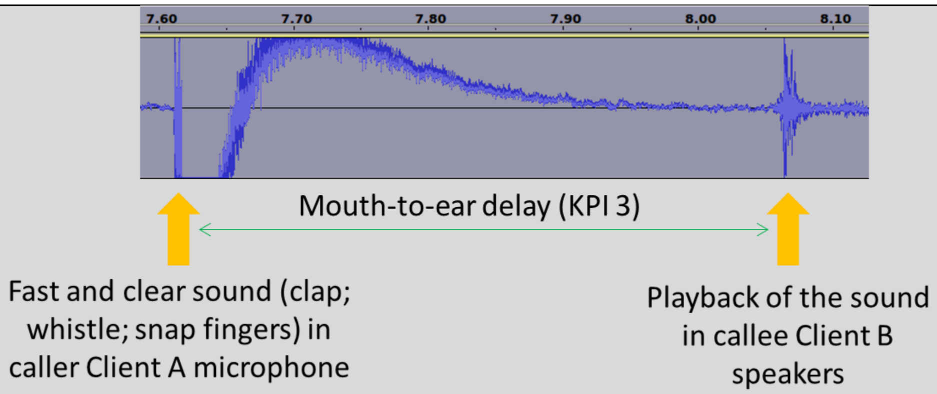
5.1.12.5. Average (expected mean) MCPTT mouth-to-ear delay test

<table>
<tr><td colspan="2">Test Case</td><td><i>TC-MAL-UC3-005</i></td><td><i>MCPTT</i></td></tr>
<tr><td>1</td><td colspan="3">
<p>Target KPI</p>
<p><i>Average (expected mean) MCPTT mouth-to-ear delay Calibration</i></p>
<p>The MCPTT mouth-to-ear delay calibration tests aim at assessing the measurement capabilities of the measurement system employed for further MCPTT mouth-to-ear tests.</p>
<p>The MCPTT mouth-to-ear delay is the time between an utterance by the transmitting user, and the playback of the utterance at the receiving user’s speaker.</p>
<p>The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the application layer.</p>
<p>Source of RTP packets →measurement probe acting as client A</p>
</td></tr>
</table>

<table>
<tr><td></td><td>Destination of RTP packets → measurement probe acting as client B
Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT
Measurement conducted at layer → Application layer</td></tr>
<tr><td>2</td><td>Methodology


Client A and B are placed next to each other providing that there is sufficient space so as to ensure no audio coupling. Next to the devices, a microphone is placed in order to capture the complete sound activity. Attached to the microphone, there is a need for a computer running a sound recording tool (e.g. Audacity).
Once defined the measurement setup, it is important to define the procedure in order to be capable of gathering significant samples.
1. Client A plays the role of caller and client B does the same as a callee.
2. Client A calls client B with implicit token request.
3. The call is established and client A has the token. Client B enables the speaker (up to the measurement configuration but helps capture the incoming sound).
4. Start recording in the audio recording tool.
5. Through the client A terminal, we make a clear and fast sound (e.g. clap, whistle, snap fingers). This sound is captures by the microphone and recorded by the audio recording tool in the computer.
6. The sound travels all the way until the client B. When the client B receives and plays the sound, the microphone captures it and the audio recording tool stores it, being able to measure the time-gap between the sound creation event and the reception event (MCPTT mouth-to-ear delay).
The next figure shows an example obtained with audacity while making a clear sound with a clap. Even though the audio recording tool also records echoes in the lab, it is clear that the delay is approximately 450ms (from slightly more than 7.6 s to slightly more than 8.05 s).</td></tr>
</table>

<table>
<tr>
<td></td>
<td>



The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

- Duration of a single replica (iteration) → the duration has to ensure that at least 20 messages of each required type are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25.

</td>
</tr>
<tr>
<td>3</td>
<td>

Calculation process and output

The required output should be calculated according to the following methodology:

Mean (average) MCPTT mouth-to-ear delay:

Let avg_i be the calculated average MCPTT mouth-to-ear delay for the i^{th} replica (iteration), and x_(i,n) be the measured MCPTT mouth-to-ear delay for each repetition within the replica (iteration).

$$avg_i = \frac{1}{n}\sum_{n} x_{i,n}$$

Then, the overall (reported) average MCPTT mouth-to-ear delay *avg* shall be calculated as the average of all x_i

$$avg = \frac{1}{i}\sum_{i} avg_i$$

For the overall average MCPTT mouth-to-ear delay *avg*, the 95% confidence interval shall be reported using the Student-T-distribution for $v = i - 1$ degrees of freedom to denote the precision of the experiment.

Note: This methodology accounts for non-Gaussian distributions of MCPTT mouth-to-ear delays within a single trial [5].

When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.

The MCPTT mouth-to-ear delay output shall be provided as:

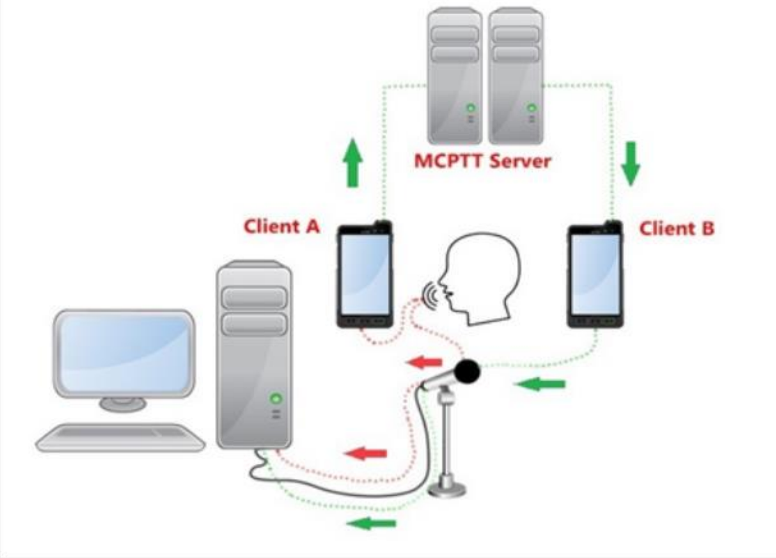
<table>
<tr><th colspan="3">MCPTT mouth-to-ear delay [ms]</th></tr>
<tr><td rowspan="2">Mean</td><td colspan="2">95% confidence interval for Mean</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>

</td>
</tr>
</table>

4	Pre-conditions The service MCPTT/MCS VNFs should be instantiated, the whole set of services up and running, the UE connected to an actual RAT that is able to reach the deployed VNFs going through a core network.
5	Applicability Measurement of 3GPP standardized MCPTT mouth-to-ear delay This test case must be executed using MCPTT 3GPP compliant UEs.
6	Test Case Sequence 1. Start the MCPTT system. 2. Establish an MCPTT call. 3. Talk to the microphone, in order to generate all the audio transmission and messages to be recorded for the measurement (as explained before) 4. Finish the call. 5. Calculate and record the average MCPTT mouth-to-ear delay per iteration as defined in “Calculation process and output”. 6. Repeat steps 1 to 5 for each one of the 25 iterations. 7. Compute the KPIs as defined in section “Calculation process and output”.

5.1.12.6. 95%-percentile MCPTT mouth-to-ear delay test

Test Case		*TC-MAL-UC3-006*	*MCPTT*
1	Target KPI ***95%-percentile MCPTT mouth-to-ear delay Calibration*** The MCPTT mouth-to-ear delay calibration tests aim at assessing the measurement capabilities of the measurement system employed for further MCPTT mouth-to-ear tests. The MCPTT mouth-to-ear delay is the time between an utterance by the transmitting user, and the playback of the utterance at the receiving user’s speaker. The calibration test employs ping messages exchanges between measurement probes (acting as client and server) which are directly communicating with each other of an “empty” SUT. The calibration test is conducted at the application layer. Source of RTP packets →measurement probe acting as client A Destination of RTP packets → measurement probe acting as client B Underlying SUT → none. Source and destination directly communicate with each other without an intermediate SUT Measurement conducted at layer → Application layer		
2	Methodology		

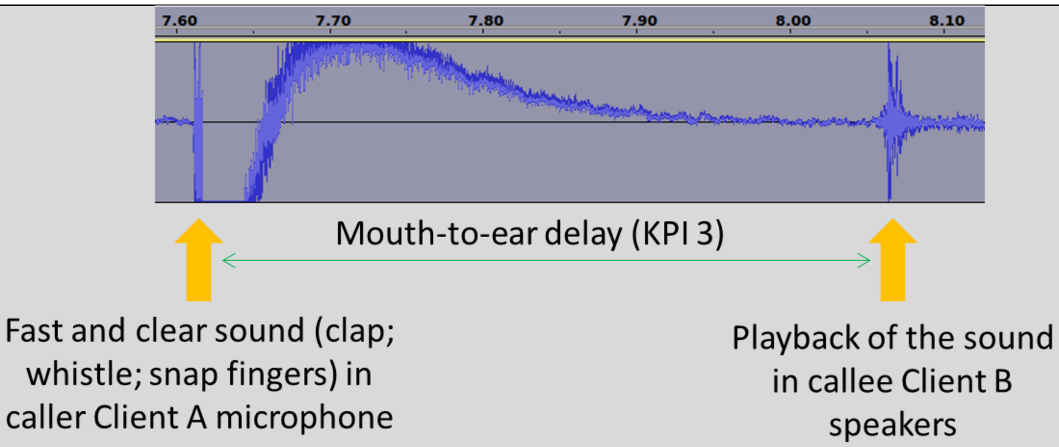


Client A and B are placed next to each other providing that there is sufficient space so as to ensure no audio coupling. Next to the devices, a microphone is placed in order to capture the complete sound activity. Attached to the microphone, there is a need for a computer running a sound recording tool (e.g. Audacity).

Once defined the measurement setup, it is important to define the procedure in order to be capable of gathering significant samples.

1. Client A plays the role of caller and client B does the same as a callee.
2. Client A calls client B with implicit token request.
3. The call is established and client A has the token. Client B enables the speaker (up to the measurement configuration but helps capture the incoming sound).
4. Start recording in the audio recording tool.
5. Through the client A terminal, we make a clear and fast sound (e.g. clap, whistle, snap fingers). This sound is captures by the microphone and recorded by the audio recording tool in the computer.
6. The sound travels all the way until the client B. When the client B receives and plays the sound, the microphone captures it and the audio recording tool stores it, being able to measure the time-gap between the sound creation event and the reception event (MCPTT mouth-to-ear delay).

The next figure shows an example obtained with audacity while making a clear sound with a clap. Even though the audio recording tool also records echoes in the lab, it is clear that the MCPTT mouth-to-ear delay is approximately 450ms (from slightly more than 7.6 s to slightly more than 8.05 s).

<table>
<tr>
<td></td>
<td>


The test case shall include the consecutive execution of several replica (iterations) according to the following properties.

- Duration of a single replica (iteration) → the duration has to ensure that at least 20 messages of each required type are sent during a single replica (iteration).
- Number of replica (iterations) → At least 25
</td>
</tr>
<tr>
<td>3</td>
<td>
Calculation process and output

The required output should be calculated according to the following methodology:

95%-percentile MCPTT mouth-to-ear delay:

Let MCPTTM2E 95p_i be the 95%-percentile MCPTT mouth-to-ear delay measured in the ith replica (iteration), and x_(i,n) be the measured MCPTT mouth-to-ear delay for each repetition within the replica (iteration).

$$MCPTTM2E_{95p_i} = 95\%\ \text{percentile}(x_{i,n})$$

Then, the (reported) 95%-percentile MCPTT mouth-to-ear delay (MCPTTM2E_95p) shall be calculated as follows:

$$MCPTTM2E\_95p = \frac{1}{i}\sum_{i} MCPTTM2E\_95p_i$$

For the reported 95%-percentile MCPTT mouth-to-ear delay ($MCPTTM2E\_95p$), the 95% confidence interval shall be reported using the Student-T-distribution for *v = i - 1* degrees of freedom to denote the precision of the experiment.

The MCPTT mouth-to-ear delay output shall be provided as:

<table>
<tr><td colspan="3">MCPTT mouth-to-ear delay [ms]</td></tr>
<tr><td rowspan="2">95% Per-centile</td><td colspan="2">95% confidence interval for 95% Percentile</td></tr>
<tr><td>Lower bound</td><td>Upper bound</td></tr>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
Pre-conditions

The service MCPTT/MCS VNFs should be instantiated, the whole set of services up and running, the UE connected to an actual RAT that is able to reach the deployed VNFs going through a core network.
</td>
</tr>
</table>

<table>
<tr><td>5</td><td>Applicability
Measurement of 3GPP standardized mouth-to-ear delay</td></tr>
<tr><td>6</td><td>Test Case Sequence
1. Start the MCPTT system.
2. Establish an MCPTT call.
3. Talk to the microphone, in order to generate all the audio transmission and messages to be recorded for the measurement. (as explained before)
4. Finish the call.
5. Calculate and record the 95%-percentile MCPTT mouth-to-ear delay per iteration as defined in “Calculation process and output”.
6. Repeat steps 1 to 5 for each one of the 25 iterations.
7. Compute the KPIs as defined in section “Calculation process and output”.</td></tr>
</table>

5.2. Application Level Tests

5.2.1. Video streaming jitter tests

5.2.1.1. Average (expected mean) jitter

<table>
<tr><td colspan="2">Test Case</td><td><i>TC-VideoStream-Jitter-001</i></td><td><i>VideoStreamJitter</i></td></tr>
<tr><td>1</td><td colspan="3">Target KPI
<i>Average Jitter</i>
The jitter test aims to assess the measurement capabilities of the measurement system employed in future jitter tests.
The calibration test employs traffic generators and a traffic receptor, which are communicating with each other in an “empty” system under test.
Measurement probes may be virtualized or physical. In case of virtualized probes, all probes shall be instantiated on the same (physical) host. In case of physical probes, the latter shall be preferably directly connected.
Source of packets → video generator
Destination of packets → video receptor
Underlying SUT → Network components between the source and destination
Measurement conducted at layer → Application layer</td></tr>
<tr><td>2</td><td colspan="3">Methodology
For measuring jitter, a RTP (Real Time Protocol) stream is emitted from a source and received by a data sink (destination). The amount of data (Byte) successfully transmitted per unit of time (seconds) and received shall be recorded.
The test case shall include the consecutive execution of several iterations according to the following properties.
• Duration of a single iteration → at least two (2) minutes.
• Records RTP stream received.
• Number of replica → At least 25 iterations.</td></tr>
</table>

<table>
<tr>
<td>3</td>
<td>
<p><i>Calculation process and output</i></p>
<p>The required output should be calculated according to the following methodology:</p>
<p>Jitter:</p>
<p>In each iteration the jitter is calculated according to RFC 3550 (RTP)</p>
<p>Then, the overall (reported) jitter shall be calculated as the average of all the iteration</p>
<p>$$avg = \frac{1}{i} \sum_{i} jitter_i$$</p>
<p>For the overall average jitter, the 95% confidence interval shall be reported using the Student-T-distribution for <i>v = i - 1</i> degrees of freedom to denote the precision of the experiment.</p>
<p>Note: This methodology accounts for non-Gaussian distributions of latencies within a single trial [5].</p>
<p>When reporting the output, information on the specific combination of values specified in the Experiment Descriptor shall be given.</p>
<p>The Jitter output shall be provided as:</p>
<table>
<tr><th colspan="3">Jitter [ms]</th></tr>
<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>
</table>
</td>
</tr>
<tr>
<td>4</td>
<td>
<p>Complementary measurements</p>

PL rate is recorded in order to detect jitter peaks due to packet lost.
Inter-packet delay

</td>
</tr>
<tr>
<td>5</td>
<td>
<p>Pre-conditions</p>

The scenario has been configured.
The traffic generator should support the generation RTP video traffic.
Ensure that no undesired traffic is present during the test.

</td>
</tr>
<tr>
<td>6</td>
<td>
<p>Applicability</p>
<p>The measurement probes need to be capable of recording the traffic received.</p>
</td>
</tr>
<tr>
<td>7</td>
<td>
<p>Test Case Sequence</p>

Start monitoring probes (deployment of probes running iPerf client and server).
Using the traffic generator, begin transmitting from the server to the client.
Record the traffic for each time interval within a trial.
Stop the traffic generator.
Stop monitoring probes.
Calculate and record the KPIs as needed per iteration as defined in “Calculation process and output”.
Repeat steps 1 to 6 for each one of the 25 iterations.
Compute the KPIs as defined in section “Calculation process and output”.

</td>
</tr>
</table>

6. TRAFFIC PROFILES

6.1. ICMP ECHO_REQUEST – ECHO_RESPONSE traffic

6.1.1. 56-byte-payload ECHO_REQUESTS

	Traffic Description	*TD-002*
1	Traffic sources ICMP ECHO_REQUEST – ICMP ECHO_RESPONSE traffic stream between source and destination and backwards. The stream of generated ICMP ECHO_REQUESTs shall comply to the following specification: • Sending rate of ICMP ECHO_REQUESTs → 2 Hz • Length of Data Field in the ICMP ECHO_REQUEST → 56 bytes	
2	Service Type (optional) *n/a*	

6.1.2. 32-byte-payload ECHO_REQUESTS

	Traffic Description	*TD-003*
1	Traffic sources ICMP ECHO_REQUEST – ICMP ECHO_RESPONSE traffic stream between source and destination and backwards. The stream of generated ICMP ECHO_REQUESTs shall comply to the following specification: • Sending rate of ICMP ECHO_REQUESTs → 2 Hz • Length of Data Field in the ICMP ECHO_REQUEST → 32 bytes	
2	Service Type (optional) *n/a*	

6.1.3. 1400-byte-payload ECHO_REQUESTS

	Traffic Description	*TD-001*
1	Traffic sources ICMP ECHO_REQUEST – ICMP ECHO_RESPONSE traffic stream between source and destination and backwards. The stream of generated ICMP ECHO_REQUESTs shall comply to the following specification: • Sending rate of ICMP ECHO_REQUESTs → 2 Hz • Length of Data Field in the ICMP ECHO_REQUEST → 1400 bytes	
2	Service Type (optional) *n/a*	

6.2. TCP/UDP traffic

Traffic Description		*TD-004*
1	Traffic sources UDP traffic stream between source and destination and backwards. The stream of generated UDP packets shall comply to the following specification: • Length of UDP datagaram (UDP header + UDP payload) → 1400 bytes • The bandwidth depends on test purpose. For maximum user data rate tests the bandwidth should be set above the maximum bandwidth available in the network scenario.	
2	Service Type (optional) *n/a*	

7. MALAGA PLATFORM EXPERIMENTS

7.1. Overview

The Malaga platform has eight different setups, which are enumerated in Table 10. Only the first two setups (setup 1 and setup 2) have been worked on in the first integration cycle (Phase 1) for the Malaga Platform. These two setups aim at quantifying the 4G baseline for the KPIs targeted in the Malaga platform and assessing the integration of the core network. The other setups will be worked on in the next quarters and will be described in future releases of WP6 deliverables, i.e. D6.2 and D6.3.

The TRIANGLE testbed, shown in Figure 6, is a lab experimentation environment that provides an E2E 4G environment where radio and network conditions are configurable. In addition, the testbed offers a radio emulator that enables reproducing mobility conditions, adding in the radio interface the multipath impairments and the AWGN (Additive white Gaussian noise). In setup 1, as shown in Figure 7, the UEs used for the tests have been connected through RF cables to the radio emulator, which is a Keysight E7515A UXM Wireless Test Set.

Table 10 Setups available at the Malaga platform.

#	Description
1	TRIANGLE testbed
2	Indoor E2E 4G: Compute node+Athonet EPC+Nokia small cell+Commercial UEs
3	Lab 5G Eurecom no core: ECM UE+ECM gNB, no S1, IP traffic
4	Lab 5G RunEl no core: RUN UE+ RUN gNB, no S1, IP traffic
5	Lab 5G NSA: ECM UE+ECM gNB, NSA with eNB S1-AP and gNB-S1-U to ATHONET EPC
6	Full E2E- Lab 5G: ECM UE+RUN gNB+ATH 5G CORE
7	Full E2E 4G with VIM: Setup 2 + VIM + Slice manager + ELCM+ Portal
8	Full E2E 5G with all components (including outdoor)

Figure 6 TRIANGLE testbed

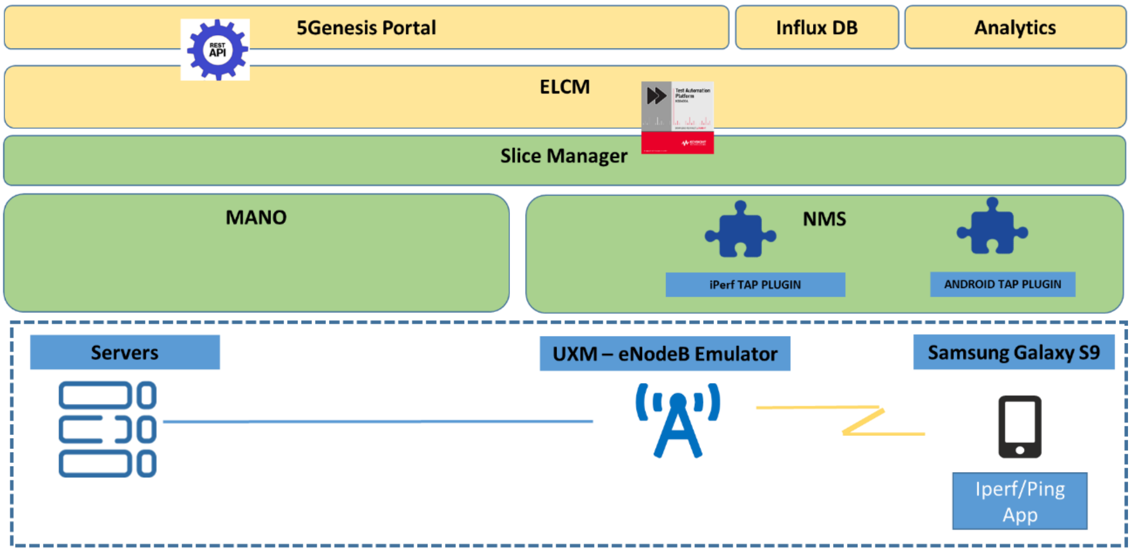


Figure 7 Malaga platform setup 1 based on the TRIANGLE testbed

For setup 1 four different scenarios have been identified.

Scenario 1 and scenario 2 have been designed to have a baseline of the KPI under test for 4G technology in ideal conditions. Scenario 2 has been used to measure the maximum user data rate available when using 4 carriers aggregated and a maximum modulation of 256 QAM.

Scenario 3 and scenario 4 have been defined to quantify the performance degradation introduced in the KPI under test due to radio impairments such as muti-path propagation and AWGN noise. Emulation channel conditions and SNR values are based on the scenarios provided in 3GPP TR 37.901 for testing throughput at the application level in User Equipments. Scenario 3 emulates the radio conditions perceived by the UE when the user is walking down an urban street at 1 to 3 km/h. It is also expected that the channel conditions vary due to the moving obstacles such as vehicles. Scenario 4 emulates the conditions perceived by the UE when the user is traveling by car. The speed of the car car is expected to be on the high end of the urban limits, possibly 40 to 60 km/h. The signal strength is lower than in the pedestrian scenario since the attenuation introduced by the chassis is taken into account. The maximum Block Level Error Rate (BLER) in these two scenarios is between 6% and 10%.

The configuration details for these scenarios are described in the tables below.

Table 11 Ideal scenario 1 Component Carrier (Scenario 1)

Band	3
Downlink bandwidth	20 MHz
Transmission mode	TM3
MIMO	2x2
CFI	1
PMI/RI Mode	Adaptative
Number of carriers	1
Modulation	Adaptative

Max modulation	27_256QAM
Max downlink OTA throughput	195 Mbits
Cell Power	-60 dBm
Noise/Interference	N/A
Channel model	Static MIMO
Channel model Doppler	N/A
Channel model correlation	N/A

Table 12 Ideal scenario carrier aggregation 4 component carriers 256 QAM (Scenario 2)

Band	3,3,7,7
Downlink bandwidth	20 MHz
Transmission mode	TM3
MIMO	2x2
CFI	1
PMI/RI Mode	Adaptative
Number of carriers	4
Modulation	Fixed
Max modulation	27_256QAM
Max downlink OTA throughput	780 Mbits
Cell Power	-60 dBm
Noise/Interference	N/A
Channel model	Static MIMO
Channel model Doppler	N/A
Channel model correlation	N/A

	Current	Peak	Average	Transferred
Downlink IP	0.00 bps	0.00 bps	0.00 bps	0.00 bits
Uplink IP	0.00 bps	0.00 bps	0.00 bps	0.00 bits
Downlink OTA	781.19 Mbps	781.20 Mbps	781.11 Mbps	88.56 Gbits
Uplink OTA	0.00 bps	0.00 bps	0.00 bps	0.00 bits

Figure 8 eNodeB emulator configuration for scenario 2 (4CC)

Table 13 Urban pedestrian scenario (Scenario 3)

Band	3
Downlink bandwidth	20 MHz
Transmission mode	TM3
MIMO	2x2
CFI	1
PMI/RI Mode	Adaptative
Number of carriers	1
Modulation	Adaptative
Max modulation	27_256QAM
Max downlink OTA throughput	195 Mbps
Cell Power	-70 dBm
Noise/Interference	20 dB
Channel model	Extended Pedestrian A model (EPA)
Channel model Doppler	5 Hz
Channel model correlation	Medium

Table 14 Urban driving scenario (Scenario 4)

Band	3
Downlink bandwidth	20 MHz
Transmission mode	TM3
MIMO	2x2
CFI	2
PMI/RI Mode	Adaptative
Number of carriers	1
Modulation	Adaptative
Max modulation	27_256QAM
Max downlink OTA throughput	195 Mbps
Cell Power	-80 dBm
Noise/Interference	15 dB
Channel model	Extended Vehicular A model (EVA)
Channel model Doppler	70 Hz
Channel model correlation	High

Compared to setup 1, setup 2 is much closer to a field deployment, where channel emulation is not available. Experiments have benen conducted to validate the integration of the Athonet core, the Video surveillance camera provided by the local police department, and the service and terminals provided by Nemergent for testing MCPTT services, as shown in Figure 10.

In Setup 2 the radio connection is over the air, using a Nokia small cell, which has been isolated to avoid interferences with a homemade solution that uses HNG100 metal fabric that assures an attenuation of 100 dB. Apart from the components outlined in the previous paragraph and the small cell, a compute node is also part of this setup. This compute node is where the VNF executing the MCS server part has been deployed.

Figure 9 Malaga platform setup 2 equipment

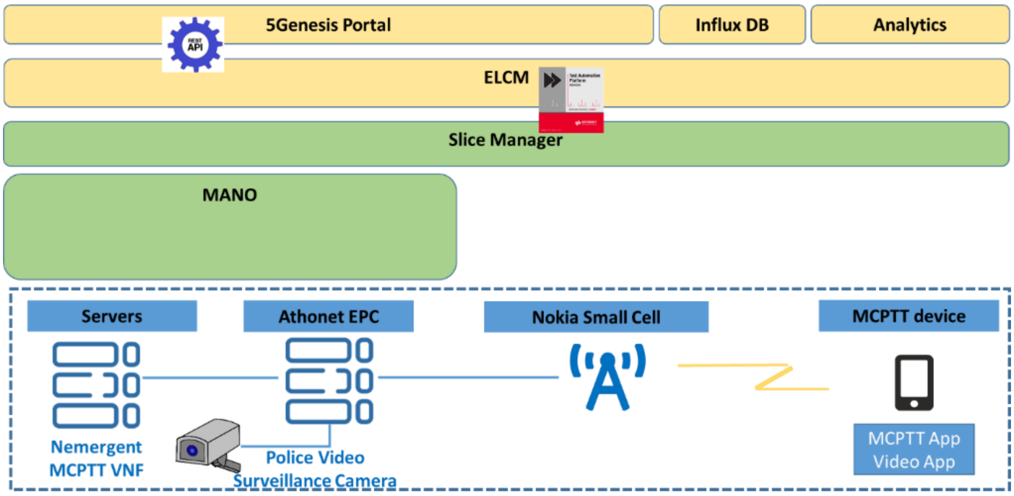


Figure 10 Malaga platform setup 2

The scenario for the setup 2 has the following configuration:

Band	7
Downlink bandwidth	10 MHz
Number of carriers	1
Modulation	Adaptative
Max modulation	64QAM
Cell Power	17 dBm

Table 15 and Table 16 report more details on the components used in the Malaga Platform and on the primary 5G KPIs targeted in the first integration cycle (Phase 1), respectively.

Table 15 Experimentation methodology components in the first integration cycle for the Malaga Platform, according to D2.3 [12]

Experimentation methodology component	Plan for integration and trial Phase 1	Status / Trial Phase I achievements
Open API's + Dispatcher	No	N/A
Experiment Life Cycle	PVC	N/A
Portal	POC	N/A
Custom experiments	POC	N/A
Standard experiments	No	N/A
E2E slices	No	N/A
VNF's	Yes	The Malaga Platform supports placement of VNFs in the testbed via the slice manager. Following the experiment lifecycle the VNFs are triggered by the ELCM which, ultimately, uses Keysight's Test Automation Platform (TAP) to communicate with the slice manager.
Scenarios	POC	Ideal/Urban Pedestrian/Urban driving
Un-attended experiments	POC:	RTT and Maximum user data rate test cases has been executed automatically in setup 1.
Attended experiments	POC	Live video streaming and MCPTT applications have been operated manually during the execution of the test cases.

Security Manager	NA	N/A

Table 16 Primary 5G KPIs evaluated at the Malaga Platform in the first trial

KPI to be evaluated at the Malaga Platform according to DoW	Evaluated in Phase 1 / First Trial	Comment
Capacity	no	Not scheduled for Phase 1
Speed	no	Not scheduled for Phase 1
Latency	no	Not scheduled for Phase 1
Reliability	no	Not scheduled for Phase 1
Density of Users	no	Not scheduled for Phase 1
Location accuracy	no	Not scheduled for Phase 1
Service Creation Time	yes	Calibration test to deploy a predefined VM as a VNF providing "MCPTT server" for test
Network management CAPEX/OPEX	no	Not scheduled for Phase 1
Additional 5G KPIs evaluated at the Malaga Platform	Evaluated in Phase 1 / First Trial	Comment
Delay (RTT)	yes	-
Maximum user data rate	yes	-
MCPTT	yes	-
Jitter	yes	-

7.2. Experiments and results

7.2.1. Round trip time calibration test

7.2.1.1. Summary and discussion of results for the E2E latency (RTT) calibration tests

The RTT calibration test has been used to characterize the delay mainly due to the radio interace. The test case has been executed in tree different scenarios defined in section 7.1: Ideal conditions with 1 carrier aggregation (scenario 1), urban pedestrian (scenario 3) and urban driving (scenario 4).

The mean RTT obtained in the three scenarios is very similar due to the fact that RTT is not in competition with other traffic, the modulation is adaptive and the UE has assigned all the radio resources, that is, the UE is able to transmit in all the subframes and in all the radio resource blocks. However, the maximum RTT value reached in scenario 3 and 4 is higher due to the lost packet at the radio link at the HARQ retransmissions (4 HARQ retransmissions have been configured in all the scenarios). Figure 11 depicts the obtained results.

The mean RTT reachable currently in the Malaga platform is 25 ms.

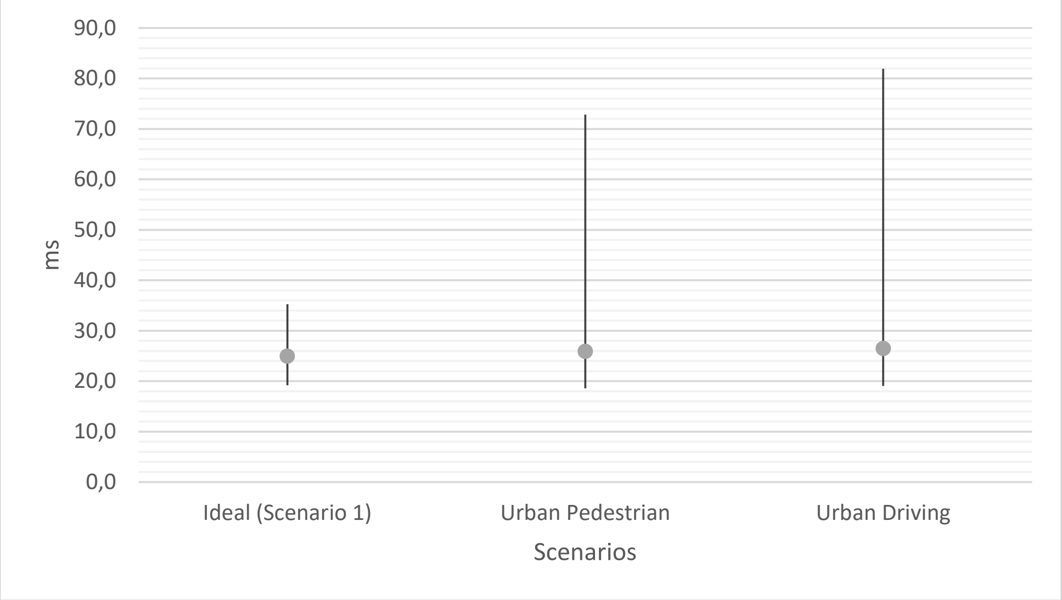


Figure 11 RTT results

7.2.1.2. Round trip time (RTT) between an UE and a VNF running on the compute node of the infrastructure

Test Case ID	TC-RTT-001, TC-RTT-002, TC-RTT-003, TC-RTT-04, TC-RTT-05			
General description of the test	The tests assess the average, minimum, maximum, 5% percentile and 95% percentile RTT between an UE and a VNF deployed on a single compute node in the network.			
Purpose	Characterize the impact on RTT of different radio scenarios.			
Executed by	Partner:	UMA	Date:	20.06.2019
Involved Partner(s)	UMA, NEMERGENT, ATOS			
Scenario	Ideal 1 CA			
Slicing configuration	VNF deployed at the compute node			
Components involved *(e.g. HW components, SW components)*	Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)			
Metric(s) under study (*Refer to those in Section 4*)	*Round Trip Time defined in Section 4.4.2*			
Additional tools involved	TAP for automated testing, VNF, Ping, Ping TAP plugin,			
Primary measurement results *(those included in the test case definition)*	Round Trip time			

	RTT [ms]		
	Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval

	Mean	24.844278	24.964893	25.085508
	Minimum	18.533175	19.192	19.850825
	Maximum	29.726087	35.268	40.809913
	5% percentile	20,611548	20.68	20,748452
	95% percentile	29,322434	29.444	29,565566
Complementary measurement results	n/a			

Test Case ID	TC-RTT-001, TC-RTT-002, TC-RTT-003, TC-RTT-04, TC-RTT-05			
General description of the test	The tests assess the average, minimum, maximum, 5% percentile and 95% percentile RTT between an UE and a VNF deployed on a single compute node in the network.			
Purpose	Characterize the impact on RTT of different radio scenarios.			
Executed by	Partner:	UMA	Date:	13.07.2019
Involved Partner(s)	UMA, NEMERGENT, ATOS			
Scenario	Urban pedestrian 1 CA			
Slicing configuration	VNF deployed at the compute node			
Components involved *(e.g. HW components, SW components)*	Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)			
Metric(s) under study (*Refer to those in Section 4*)	*Round Trip Time defined in Section 4.4.2*			
Additional tools involved	TAP for automated testing, VNF, Ping, Ping TAP plugin,			
Primary measurement results *(those included in the test case definition)*	Round Trip time			
	RTT [ms]			
		Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval
	Mean	25,873505	26.128263	26.383021
	Minimum	18.998598	19.716	20.433402
	Maximum	66.208212	74.388	82.567788
	5% percentile	21.04035	21.1288	21.21725
	95% percentile	29.922812	30.00459	30.086368
Complementary measurement results	n/a			

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-RTT-001, TC-RTT-002, TC-RTT-003, TC-RTT-04, TC-RTT-05</td></tr>
<tr><td>General description of the test</td><td colspan="4">The tests assess the average, minimum, maximum, 5% percentile and 95% percentile RTT between an UE and a VNF deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">Characterize the impact on RTT of different radio scenarios.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>13.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Urban driving 1 CA</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Round Trip Time defined in Section 4.4.2</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, VNF, Ping, Ping TAP plugin,</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip time
<table>
<tr><td colspan="4">RTT [ms]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>26.104108</td><td>26.456743</td><td>26.809378</td></tr>
<tr><td>Minimum</td><td>18.421226</td><td>19.084000</td><td>19.746774</td></tr>
<tr><td>Maximum</td><td>77,939744</td><td>81.924000</td><td>85,908256</td></tr>
<tr><td>5% percentile</td><td>20,80349</td><td>20.924</td><td>21,04451</td></tr>
<tr><td>95% percentile</td><td>28,842912</td><td>33.086</td><td>37,329088</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

7.2.2. Maximum user data rate calibration test

7.2.2.1. Summary and discussion of results for maximum user data rate calibration tests

In this section the maximum user data rate is tested in scenarios 1, 3 and 4. The goal is to measure how the user data rate decreases due to radio impairments. As shown in Figure 12, in the ideal scenario (scenario 1) the user data rate is close to the maximum throughput available in OTA for scenario 1 (195 Mbps). In scenario 3 and 4, the user date rate reduced to a quarter of the bandwidth available in OTA.

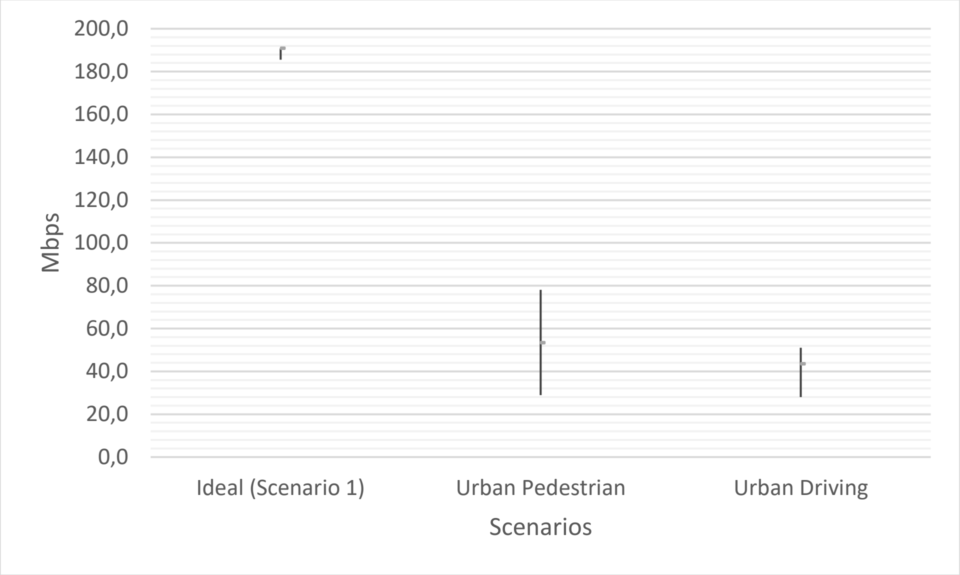


Figure 12 User data rate results

Finally, the test TC-Thr-003 has been executed in scenario 2, which is an ideal scenario featuring CA using 4 component carriers in two different bands. The theoretical maximum user data rate available is close to 800 Mbps. However, the mean data rate measured at the application is 626 Mbps, which is the maximum throughput currently provided by the Malaga platform.

7.2.2.2. Maximum user data rate test calibration

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the calculation of the average maximum user data rate available.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure the maximum user data rate available in different scenarios.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>13.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Ideal 1 CA (See scenario 1 defined in section 7.1)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Throughtput*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, VNF, Ping, Ping TAP plugin,</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Maximum user data rate<table><tr><td colspan="4">Maximum user data rate [Mbps]</td></tr><tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr></table></td></tr>
</table>

<table>
<tr><td></td><td><table><tr><td>Mean</td><td>190.614122</td><td>190.743123</td><td>190.872124</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the calculation of the average maximum user data rate available.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure the maximum user data rate available in different scenarios.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>12.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Urban pedestrian 1 CA (See scenario 3 defined in section 7.1)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, VNF, Ping, Ping TAP plugin.</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Maximum user data rate<table><tr><td colspan="4">Maximum user data rate [Mbps]</td></tr><tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr><tr><td>Mean</td><td>50,307196</td><td>51.649192</td><td>52.991188</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the calculation of the average maximum user data rate available.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure the maximum user data rate available in different scenarios.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>15.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Urban driving 1 CA (See scenario 4 defined in section 7.1)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
</table>

<table>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td>*Throughput*</td></tr>
<tr><td>Additional tools involved</td><td>TAP for automated testing, VNF, Ping, Ping TAP plugin.</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Maximum user data rate
<table>
<tr><th colspan="4">Maximum user data rate [Mbps]</th></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>43.298815</td><td>43.452077</td><td>43.605339</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the calculation of the average maximum user data rate available.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure the maximum user data rate available in different scenarios.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>12.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Ideal CA 4 Carrier Components (See scenario 2 defined in section 7.1)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">Samsung Galaxy S9, UXM (eNodeB emulator), Openstack server (compute node)</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Throughput*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, VNF, Ping, Ping TAP plugin,</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Maximum user data rate
<table>
<tr><th colspan="4">Maximum user data rate [Mbps]</th></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>598,6001068</td><td>626.8947368</td><td>655,18936809</td></tr>
</table></td></tr>
</table>

Complementary measurement results	n/a

7.2.3. MCPTT

7.2.3.1. Summary and discussion of results for the MCPTT tests

Considering the standardized thresholds for each KPI and good values obtained well below the max set threshold, the results clearly show two important facts:

1) The testing environment is innocuous enough not to induce additional delays, therefore it provides the perfect ground to perform reliable tests on technology and services, and
2) The tested MCS service is efficient in a way that the service itself only consumes a third part of the total threshold to achieve the measured task.

The very good results obtained with 4G under lab circumstances give an optimistic expectation on the capability of the Malaga Platform to deliver similarly good results also outside of the lab and deploying the still under development 5G features and services.

7.2.3.2. Average MCPTT access time test

Test Case ID	TC-MAL-UC3-001, TC-MAL-UC3-002			
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 aim at assessing the measurement capabilities of the measurement system employed for further MCPTT access time tests.			
Executed by	Partner:	UMA	Date:	09.07.2019
Involved Partner(s)	UMA, NEM, ATOS			
Scenario	Athonet 4G Core with Nokia small cell with -17 dBm power and LTE band 7. The measurements are taken at the application level, in the Nemergent MCS application.			
Slicing configuration	VNF deployed at the compute node			
Components involved *(e.g. HW components, SW components)*	NEM MCS applications and MCS server VNF, Nokia small cell eNB, Athonet 4G EPC, NEM (SONIM) UEs			
Metric(s) under study (*Refer to those in Section 4*)	*MCPTT*			
Additional tools involved	Logcat Android log command-line tool			

<table>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td>MCPTT Access time
<table>
<tr><th colspan="3">MCPTT access time [ms]</th></tr>
<tr><th rowspan="2">Mean</th><th colspan="2">95% confidence interval for Mean</th></tr>
<tr><th>Lower bound</th><th>Upper bound</th></tr>
<tr><td>55,192</td><td>50,114</td><td>60,271</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

7.2.3.3. 95%-percentile MCPTT access time test

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-MAL-UC3-001, TC-MAL-UC3-002</td></tr>
<tr><td>General description of the test</td><td colspan="4">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>09.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEM, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Athonet 4G Core with Nokia small cell with -17 dBm power and LTE band 7. The measurements are taken at the application level, in the Nemergent MCS application.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">NEM MCS applications and MCS server VNF, Nokia small cell eNB, Athonet 4G EPC, NEM (SONIM) UEs</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>MCPTT</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Logcat Android log command-line tool</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">MCPTT Access time
<table>
<tr><th>MCPTT access time [ms]</th></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td><table><tr><td rowspan="2">95% Per-centile</td><td colspan="2">95% confidence interval for Min</td></tr><tr><td>Lower bound</td><td>Upper bound</td></tr><tr><td>92,308</td><td>83,227</td><td>101,388</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

7.2.3.4. Average MCPTT E2E access time test

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-MAL-UC3-003, TC-MAL-UC3-004</td></tr>
<tr><td>General description of the test</td><td colspan="4">MCPTT E2E 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.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure time from request to speak to permission granted in a MCPTT call, including call establishment. The E2E MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further E2E MCPTT access time tests.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>10.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEM, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Athonet 4G Core with Nokia small cell with -17 dBm power and LTE band 7. The measurements are taken at the application level, in the Nemergent MCS application.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">NEM MCS applications and MCS server VNF, Nokia small cell eNB, Athonet 4G EPC, NEM (SONIM) 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 E2E access time<table><tr><td colspan="3">E2E MCPTT access time [ms]</td></tr><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>247,459</td><td>225,921</td><td>268,997</td></tr></table></td></tr>
</table>

Complementary measurement results	n/a

7.2.3.5. 95%-percentile MCPTT E2E access time test

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-MAL-UC3-003, TC-MAL-UC3-004</td></tr>
<tr><td>General description of the test</td><td colspan="4">MCPTT E2E 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.</td></tr>
<tr><td>Purpose</td><td colspan="4">Measure time from request to speak to permission granted in a MCPTT call, including call establishment. The E2E MCPTT access time calibration tests aim at assessing the measurement capabilities of the measurement system employed for further E2E MCPTT access time tests.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>10.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEM, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Athonet 4G Core with Nokia small cell with -17 dBm power and LTE band 7. The measurements are taken at the application level, in the Nemergent MCS application.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">VNF deployed at the compute node</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">NEM MCS applications and MCS server VNF, Nokia small cell eNB, Athonet 4G EPC, NEM (SONIM) 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 E2E access time
<table>
<tr><td colspan="3">E2E MCPTT access time [ms]</td></tr>
<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>295,983</td><td>270,817</td><td>321,149</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

7.2.4. Video streaming jitter

7.2.4.1. Summary and discussion of results for the video streaming jitter tests

The tests for jitter using video streaming offer very good results and allow a comparison among setup 1 and 2, both as baseline measurements in 4G. The results show that the TRIANGLE setup throws better results for jitter, mainly because the RF connection is via cables, unlike in setup 2. Nevertheless, the objective is to reproduce ideal conditions in both scenarios.

The jitter in both setups is of around 1 ms. This is due to the ideal conditions and the granularity of LTE, where transmission opportunities happen at 1 ms boundaries.

The complementary results do not differ significantly. In the case of interpacket delay, this value may be imposed by the video streaming traffic nature, since it is a burst kind of traffic with very low delay between consecutive packets. Finally, packet lost is non-existent, which is perfectly reasonable, given the ideal conditions used during the test and because no other traffic was introduced.

7.2.4.2. Video streaming average jitter test in Setup 1

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-VideoStreamJitter-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">Video streaming requested from VLC in UE to the police video camera</td></tr>
<tr><td>Purpose</td><td colspan="4">Video test for baseline measurements in Setup 1</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>28.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA</td></tr>
<tr><td>Scenario</td><td colspan="4">TRIANGLE Ideal (-85 dBm) with config 5 (MIMO 2x2 256QAM up to 800 Mbps)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">-</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">TRIANGLE testbed with Keysight E7515A UXM radio, Samsung S9 UE</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Video streaming jitter</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">VLC video player, TestelDroid</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Jitter
<table>
<tr><td colspan="3">Jitter [ms]</td></tr>
<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>0,366</td><td>0,352</td><td>0,379</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">PL is 0 packets for all 25 iterations
<table>
<tr><td>Interpacket delay [ms]</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td><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>1,111</td><td>1,067</td><td>1,156</td></tr></table></td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-VideoStreamJitter-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">Video streaming requested from VLC in UE to the police video camera</td></tr>
<tr><td>Purpose</td><td colspan="4">Video test for baseline measurements in Setup 2</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>15.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA</td></tr>
<tr><td>Scenario</td><td colspan="4">Athonet 4G Core with Nokia small cell with -17 dBm power and LTE band 7</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">Nokia small cell eNB, Athonet 4G EPC, Samsung S9 UE</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Video streaming jitter</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">VLC video player, TestelDroid, Wireshark</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Jitter<table><tr><td colspan="3">Jitter [ms]</td></tr><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>1,127</td><td>1,071</td><td>1,182</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">PL is 0 packets for all 25 iterations<table><tr><td colspan="3">Interpacket delay [ms]</td></tr><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>1,036</td><td>0,983</td><td>1,088</td></tr></table></td></tr>
</table>

7.2.5. Service creation time calibration tests

7.2.5.1. Summary and discussion of results for the Service creation time calibration tests

The service creation time test consists of the creation of a new service instance, measuring the time elapsed until the instance is running, and the deletion of this instance in order to leave the system in the original state before repeating this sequence. This process is repeated for 25 iterations, with 20 consecutive deployments per iteration.

In the Malaga platform this test has been performed by automating the deployment, time measurement and deletion by using a custom Python script that sends the required calls to the Katana Slice Manager. The script is able to wait until the slices are created and deleted, but the time measurement is taken from the values calculated by the Slice Manager, which are more accurate.

For this test all the required images and configurations are already available in OpenStack, therefore only the deployment time is considered, not including the onboarding process.

7.2.5.2. Service creation time

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, 5% percentile and 95% percentile deployment time of a single VM a compute node.</td></tr>
<tr><td>Purpose</td><td colspan="4">Characterize the capability of the system to deploy a VNF on an existing 4G/5G network.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UMA</td><td>Date:</td><td>15.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">UMA, NEMERGENT, ATOS</td></tr>
<tr><td>Scenario</td><td colspan="4">Ideal 1 CA</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">OpenStack server, Katana Slice Manager</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Service Creation Time defined in Section 4.1.9*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Python scripting</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Service Creation Time
<table>
<tr><th colspan="3">Service Creation Time [s]</th></tr>
<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>13.669952 s</td><td>13,585577 s</td><td>13,754327 s</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

8. ATHENS PLATFORM EXPERIMENTS

8.1. Overview

The Athens Platform consists of numerous deployed radio access and transport network testbeds, as described in detail in deliverable D4.1[15].

Figure 13 Athens platform compute nodes and ATH EPC deployed for phase 1

In the context of deliverable D6.1, the selected testbeds for conducting the developed test cases include a pre-commercial 5G-NR NSA deployment and an experimental LTE MN based on ECM Openair Interface and Athonet vEPC.

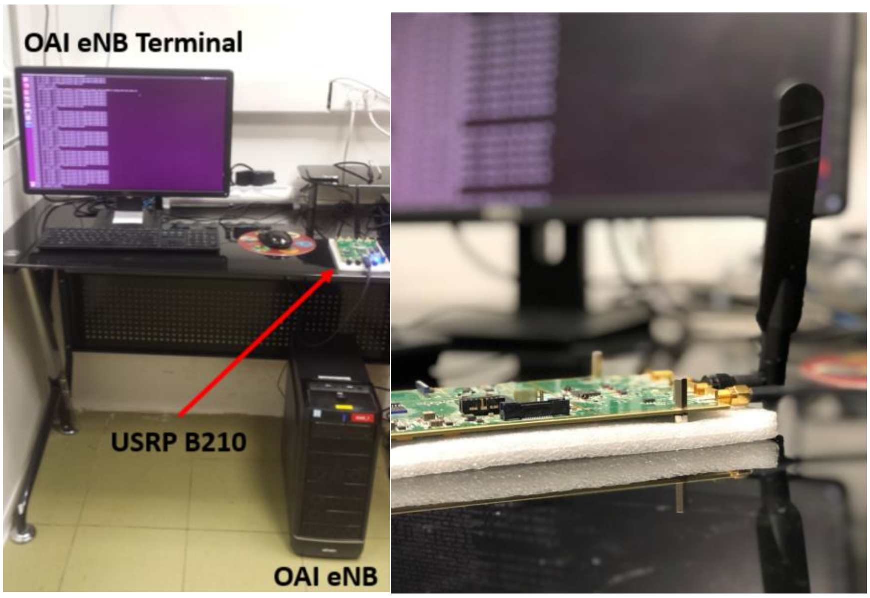


Figure 14 Openair Interface Radio Access Setup with USRP B210

---

[15] 5GENESIS Deliverable D4.1 The Athens Platform (Release A), v1.0, March 31st 2019

Table 17 presents the planned progress for the Athens platform for Phase 1.

Table 17 Summary of planned progress for Athens platform

Experimentation methodology component	Plan for integration and trial Phase 1	Status / Trial Phase I achievements
Open API's + Dispatcher	No	N/A
Experiment Life Cycle	No	N/A
Portal	No	N/A
Custom test cases	No	N/A
Standard test cases	Yes	The standard test cases as defined in section 5 of this document were executed
E2E slices	No	N/A
VNF's	No	N/A
Scenarios	LTE	Phase 1 includes only 4G core and radio, to be upgraded to 5G in Phase 2.
Un-attended test cases	No	N/A
Attended test cases	No	N/A
Security Manager	No	N/A

Athens' Platform validated KPIs are summarized in Table 18.

Table 18 Athens platform KPI validation

KPI to be evaluated at the Athens Platform according to DoA	Evaluated in Phase 1 / First Trial	Comment
Ubiquity	No	Not scheduled for Phase 1
Latency	Yes	Phase 1 focused on RTT measurements (see below)
Reliability	No	Not scheduled for Phase 1
Service creation time	No	Not scheduled for Phase 1
Additional 5G KPIs evaluated at the Athens Platform	**Evaluated in Phase 1 / First Trial**	**Comment**
Delay (RTT)	Yes	-
Throughput	Yes	-

The following section showcases the results of selected test cases on different deployment scenarios, according to the methodologies defined in deliverable D6.1.

8.2. Experiments and results

8.2.1. Throughput

8.2.1.1. Summary and discussion of results

The goal of the Throughput Measurement was the utilization of the testing methodologies developed as part of deliverable D6.1 for assessing the throughput capabilities of various deployed MN testbeds in Athens Platform.

The Throughput test cases conducted are the following:

1. DL/UL Peak Throughput -SpeedTest (5G-NR NSA precommercial setup)
2. Average Throughput (LTE SISO 20 MHz)
3. Average Throughput (LTE SISO 5 MHz)
4. Adaptive HTTP Streaming Throughput (LTE SISO 20 MHz)

The Test Case 1 is part of the Confidential Annex of this deliverable released within 5GENESIS.

The Throughput results in the OAI LTE Testbed of Experiments 2 and 3 align with the expected performance as reported by ECM for the specific radio configurations[16].

In addition to these measurements, Experiment 4 utilized both throughput and video quality based on the Adaptive HTTP Streaming scenario provided by Ixria IxChariot Traffic Generator[17]. This experiment served mainly as a baseline scenario for evaluating network performance based on user perception.

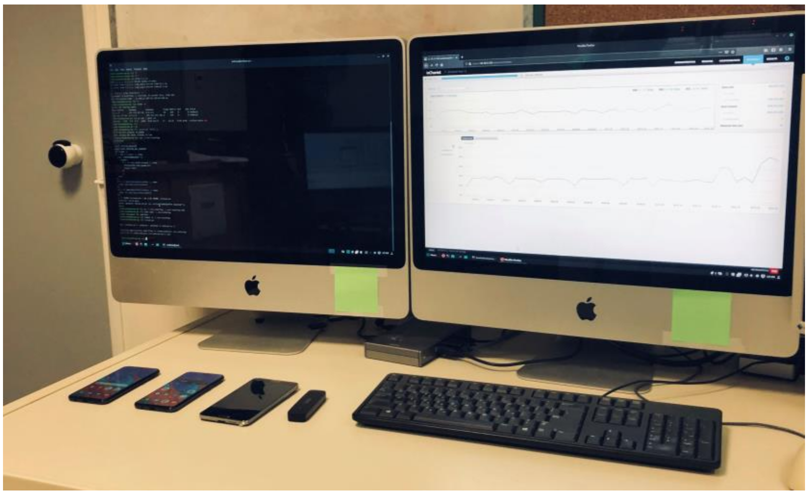
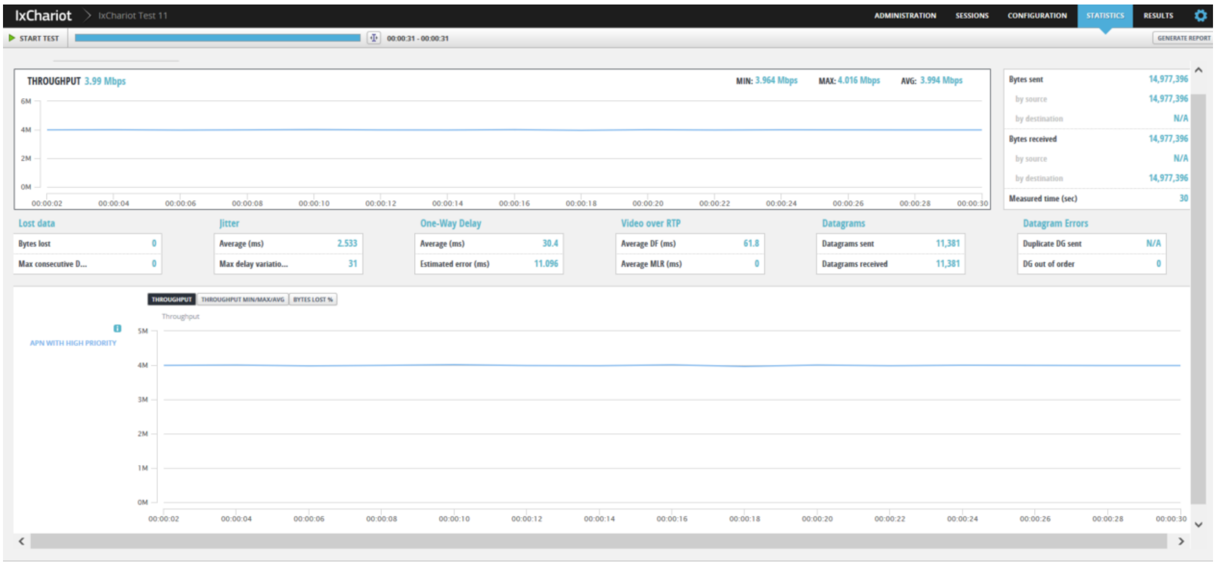


Figure 15 Measurement Probes for running the tests with IxChariot Traffic Generator

---

[16] https://gitlab.eurecom.fr/oai/openairinterface5g/blob/develop/doc/FEATURE_SET.md

[17] https://www.ixiacom.com/products/ixchariot



THROUGHPUT 3.99 Mbps

MIN: 3.964 Mbps MAX: 4.016 Mbps AVG: 3.994 Mbps

Bytes sent 14,977,396
by source 14,977,396
by destination N/A
Bytes received 14,977,396
by source N/A
by destination 14,977,396
Measured time (sec) 30

Lost data
Bytes lost 0
Max consecutive D... 0

Jitter
Average (ms) 2.533
Max delay variatio... 31

One-Way Delay
Average (ms) 30.4
Estimated error (ms) 11.096

Video over RTP
Average DF (ms) 61.8
Average MLR (ms) 0

Datagrams
Datagrams sent 11,381
Datagrams received 11,381

Datagram Errors
Duplicate DG sent N/A
DG out of order 0

Figure 16 GUI of IxChariot Traffic Generator

Considering that during the experiment the radio conditions were adequate, the majority of the video segments received by the end user were in the High-Quality scale (Bit Rate higher than 10 Mbps), as expected.

8.2.1.2. E2E DL/UL Peak Throughput in 5G pre-commercial equipment

This experiment has been performed in **5G pre-commercial equipment**. Therefore, this subsection is released as Confidential within the consortium and is provided as an autonomous annex to this public deliverable. **For further information please refer to Annex 4 of this document**.

8.2.1.3. Average Throughput (LTE SISO 20 MHz)

Test Case ID	TC-Thr-001			
General description of the test	Throughput Measurement			
Purpose	Assess Throughput Capabilities of NCSRD OAI LTE Laboratory Setup (20 MHz Bandwidth)			
Executed by	Partner:	NCSRD	Date:	20.06.2019
Involved Partner(s)	ATH, COS, ECM			
Scenario	• One connected COTS UE in experimental LTE laboratory setup • Radio configuration: SISO, 20 MHz (100 RBs), FDD, DL 2680 MHz, UL 2560 MHz)			
Slicing configuration	N/A			
Components involved *(e.g. HW components, SW components)*	OpenAirInterface and Athonet LTE Testbed			
Metric(s) under study (*Refer to those in Section 4*)	*Throughput*			

<table>
<tr><td>Additional tools involved</td><td>N/A</td></tr>
<tr><td><i>Primary measurement results</i>
<i>(those included in the test case definition)</i></td><td><table><tr><th colspan="3">Throughput [Mbps]</th></tr><tr><th rowspan="2">Average</th><th colspan="2">95% confidence interval for Average</th></tr><tr><th>Lower bound</th><th>Upper bound</th></tr><tr><td>69.42</td><td>69.31</td><td>69.54</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>N/A</td></tr>
</table>

8.2.1.4. Average Throughput (LTE SISO 5 MHz)

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">Throughput Measurement</td></tr>
<tr><td>Purpose</td><td colspan="4">Assess Throughput Capabilities of NCSRD OAI LTE Laboratory Setup (5 MHz Bandwidth)</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>NCSRD</td><td>Date:</td><td>20.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">ATH, COS, ECM</td></tr>
<tr><td>Scenario</td><td colspan="4">• One connected COTS UE in experimental LTE laboratory setup
• Radio configuration: SISO, 5 MHz (25 RBs), FDD, DL 2680 MHz, UL 2560 MHz)</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">N/A</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">OpenAirInterface and Athonet LTE Testbed</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Throughput</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">N/A</td></tr>
<tr><td><i>Primary measurement results</i>
<i>(those included in the test case definition)</i></td><td colspan="4"><table><tr><th colspan="3">Throughput [Mbps]</th></tr><tr><th rowspan="2">Average</th><th colspan="2">95% confidence interval for Average</th></tr><tr><th>Lower bound</th><th>Upper bound</th></tr><tr><td>16.55</td><td>16.48</td><td>16.63</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">N/A</td></tr>
</table>

8.2.1.5. Adaptive HTTP Streaming Throughput

Test Case ID	TC-Thr-003

<table>
<tr><td>General description of the test</td><td colspan="4">Throughput Measurement</td></tr>
<tr><td>Purpose</td><td colspan="4">Baseline scenario for evaluating network performance based on user perception.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>NCSRD</td><td>Date:</td><td>11.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">ATH, COS, ECM</td></tr>
<tr><td>Scenario</td><td colspan="4">One connected COTS UE in experimental LTE laboratory setup
Radio configuration: SISO, 20 MHz (100 RBs), FDD, DL 2680 MHz, UL 2560 MHz)
Number of iterations: 5
Duration of each iteration: 3 minutes
Traffic Type: YouTube 4K</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">N/A</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">OpenAirInterface and Athonet LTE Testbed</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Ixria IxChariot Traffic Generator
Xiaomi Note 4 with NCSRD Android Application for recording received signal strength values</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4"><table><tr><th colspan="3">Throughput [Mbps]</th></tr><tr><th>Lower Bound of 95% confidence interval</th><th>Mean</th><th>Upper Bound of 95% confidence interval</th></tr><tr><td>17.79</td><td>17.84</td><td>17.90</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">RSSI: -60.83 +/- 3.13 dBm
Video Stopped counter:
<table><tr><th>Iteration</th><th>Video Stopped Counter</th></tr><tr><td>1</td><td>13</td></tr><tr><td>2</td><td>13</td></tr><tr><td>3</td><td>13</td></tr><tr><td>4</td><td>12</td></tr><tr><td>5</td><td>9</td></tr></table>
Number of video segment sent with specific quality
<table><tr><th>Iteration</th><th>Very Low</th><th>Low</th><th>Medium</th><th>High</th><th>Very High</th></tr><tr><td>1</td><td>0</td><td>3</td><td>1</td><td>4</td><td>62</td></tr></table></td></tr>
</table>

2	0	3	1	4	62
3	0	3	1	4	62
4	0	3	0	12	59
5	0	3	0	17	57

- Average Video Rate: 17.51+/-0.13 Mbps

8.2.2. Round-Trip-Time

8.2.2.1. Summary and discussion of results

The aim of the Round-Trip-time experiment was the evaluation of the delay on a deployed setup in NCSRD, using the testing methodologies defined in this Deliverable. The Round-Trip-Time is measured between the UE and an external server connected to the Core Network and as such, the reported result of 37 ms is expected in this case.

8.2.2.2. E2E network layer Round-trip-time

Test Case ID	TC-RTT-006, TC-RTT-007, TC-RTT-008, TC-RTT-09, TC-RTT-10			
General description of the test	RTT Measurement			
Purpose	Assess RTT of NCSRD OAI LTE Laboratory Setup			
Executed by	Partner:	NCSRD	Date:	20.06.2019
Involved Partner(s)	ATH, COS, ECM			
Scenario	• One connected COTS UE in experimental LTE laboratory setup • Radio configuration: SISO, 20 MHz (100 RBs), FDD, DL 2680 MHz, UL 2560 MHz)			
Slicing configuration	N/A			
Components involved *(e.g. HW components, SW components)*	OpenAirInterface and Athonet LTE Testbed			
Metric(s) under study (*Refer to those in Section 4*)	*Round Trip Time*			
Additional tools involved	N/A			

<table>
<tr>
<td>Primary measurement results
(those included in the test case definition)</td>
<td>
<table>
<tr><th colspan="4">RTT [ms]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>36.48</td><td>36.67</td><td>36.86</td></tr>
<tr><td>Minimum</td><td>21.71</td><td>21.94</td><td>22.18</td></tr>
<tr><td>Maximum</td><td>49.93</td><td>53.70</td><td>57.48</td></tr>
<tr><td>5th Percentile</td><td>23.44</td><td>23.66</td><td>23.88</td></tr>
<tr><td>95th Percentile</td><td>49.03</td><td>49.24</td><td>49.45</td></tr>
</table>
</td>
</tr>
<tr>
<td>Complementary measurement results</td>
<td>N/A</td>
</tr>
</table>

9. LIMASSOL PLATFORM EXPERIMENTS

9.1. Overview

The goal of the first phase of experimentation in the Limassol platform has been to:

- Verify the proper functionality of the Phase 1 version of the network, i.e. an E2E 4G network with satellite backhaul, featuring virtualisation at core and edge and basic satellite edge processing;
- Assess the performance of the infrastructure in terms of throughput;
- Assess the value of satellite edge functionalities;

Table 19 recaps the components planned to be available for the first integration and experimentation cycle and relates them to the achievements of the first trial phase.

Table 19 Experimentation methodology components in the first integration cycle for the Limas-sol Platform, according to D2.3

Experimentation methodology component	Plan for integration and trial Phase 1	Status / Trial Phase I achievements
Open API's + Dispatcher	No	N/A
Experiment Life Cycle	No	N/A
Portal	No	N/A
Custom experiments	No	N/A
Standard experiments	No	N/A
E2E slices	No	N/A
VNF's	Yes (LBO, IoT interoperability, WAN acceleration)	Local break-out at the satellite edge was implemented and its effect on throughput and (especially) RTT is reflected in the measurements below. IoT interoperability was deployed as a virtual function both at the edge and the core. WAN acceleration was not deployed, since its functionality in a single-backhaul scenario would be limited. It will be used in Phase 2 in a dual-backhaul setup.
Scenarios	LTE	Phase 1 includes only 4G core and radio, to be upgraded to 5G in Phase 2.
Un-attended experiments	No	

Attended experiments	POC: IoT scenarios over satellite backhaul	Apart from tests with generic traffic, Phase 1 experimentation also included tests with IoT application traffic.
Security Manager	No	N/A

Table 20 lists the KPIs evaluated in the first trial and summarizes the kind of evaluation measurements conducted.

Table 20 Primary 5G KPIs evaluated at the Limassol Platform in the first trial

KPI to be evaluated at the Limassol Platform according to DoA	Evaluated in Phase 1 / First Trial	Comment
Ubiquity	No	Not scheduled for Phase 1
Latency	Yes	Phase 1 focused on RTT measurements (see below)
Reliability	No	Not scheduled for Phase 1
Service creation time	No	Not scheduled for Phase 1
Additional 5G KPIs evaluated at the Limassol Platform	**Evaluated in Phase 1 / First Trial**	**Comment**
Delay (RTT)	Yes	-
Throughput	Yes	-

Figure 17 depicts the physical topology of the Limassol platform, as it has been implemented for experimentation Phase 1.

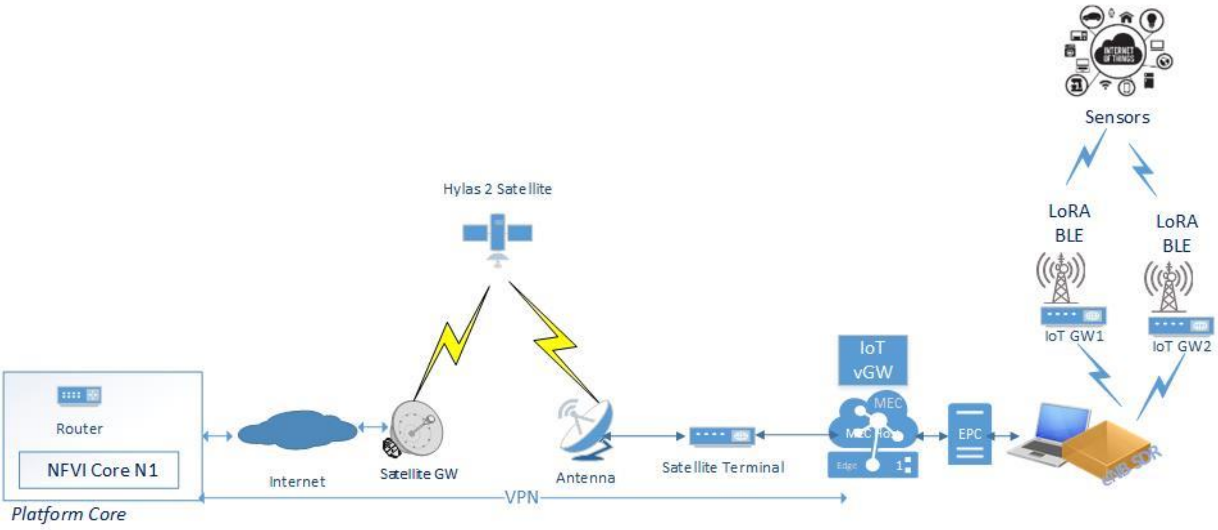


Figure 17. Actual topology of Limassol platform implemented for Phase 1 experimentation

Figure 18 Limassol platform setup equipment

The core of the platform is supported by an OpenStack cluster, acting as core NFV Infrastructure (NFVI), but also hosting virtualized management functions. Open Source MANO (OSM) Release Five has been deployed as a VM, acting as NFV Orchestrator.

In addition to the Coordination/MANO components, the OpenStack cluster also hosts IoT interoperability functions, installed as separate VMs to complete the portfolio of platform services.

The Limassol platform offers a satellite backhaul towards providing wide-area coverage for M2M communications and voice services. The satellite gateway is the entry point for the satellite backhaul, a key element of the Limassol platform. Connectivity of the core NFVI to the satellite gateway is implemented over an L3VPN. The satellite backhaul is established in Ka-band, using the Hylas 2 satellite.

At the satellite edge, edge computing equipment is deployed to allow the deployment of local traffic handling functions (such as e.g. LBO or data adaptation functions). A single edge node is currently used, allowing the deployment of functions either as VMs or containers. The EPC functions (provided by Athonet) are also implemented at the satellite edge.

The radio access component at LTE Band 7. It is SDR-based, based on the Eurecom OpenAirInterface platform, consisting of a software-driven baseband unit (BBU), an RF digital front-end (DFE), and the antenna modules. The UEs are COTS devices (4G dongles and smartphones).

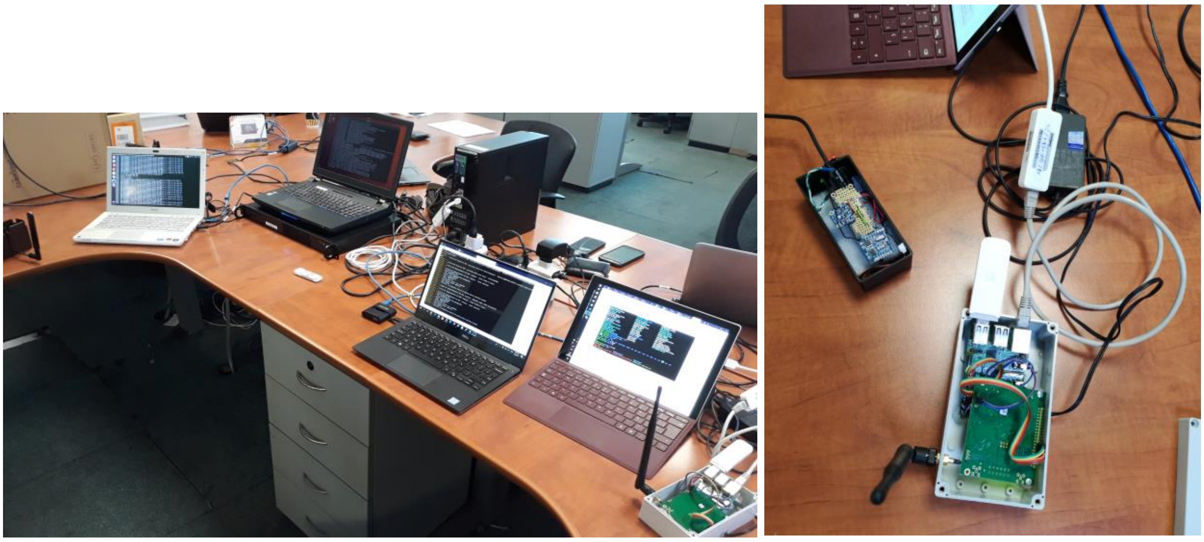


Figure 19 Limassol Platform experiments setup with Ues and IoT-oriented RANs

Finally, IoT-oriented RANs are included, such as BLE and LoRa, connecting the IoT physical gateways with the edge sensors and actuators.

9.2. Experiments and results

9.2.1. E2E latency (RTT)

9.2.1.1. Summary and discussion of results for the E2E latency (RTT) tests

The aim of the RTT tests has been to measure the E2E round-trip latency between several points in the Limassol platform network. Basically, the focus has been in two configurations: RTT in the edge 4G network segment (4G and EPC at the edge) and in the whole E2E path, including the satellite domain, the Internet and the core network (see topology in the previous section).

The derived RTT values, as presented in the next sections, were close to the ones expected, i.e. around 30 ms for the 4G network at the edge, and around 650 ms for the E2E path including the satellite backhaul. This difference implies a ~95% reduction of the perceived latency for the traffic which is locally handled at the satellite edge instead of traversing the backhaul, and emphasises the value of LBO functionalities at the satellite edge.

9.2.1.2. E2E latency (RTT) between UE and satellite edge

Test Case ID	TC_LIM_RTT_01
General description of the test	The test measures the RTT over the satellite edge segment, i.e. between i) the 4G UE and ii) a virtual endpoint (VM) deployed at the edge.
Purpose	This test assesses the performance (in terms of latency) at the 4G network at the edge, including the latency of the 4G RAN and EPC.

<table>
<tr><td>Executed by</td><td>Partner:</td><td>SHC</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in lab environment, with the 4G UE and the eNB at ~5 m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5 MHz (25 PRBs). ICMP packets of 32 bytes were used.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">COTS UE (4G USB dongle), SDR-based eNB with Eurecom OAI, Athonet EPC, edge node</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Latency (RTT)</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Ubuntu-based VMs acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><th colspan="3">RTT [ms]</th></tr>
<tr><th rowspan="2">Mean</th><th colspan="2">95% confidence interval for Mean</th></tr>
<tr><th>Lower bound</th><th>Upper bound</th></tr>
<tr><td>42.9</td><td>42.6</td><td>43.3</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

9.2.1.3. E2E latency (RTT) between UE and platform core

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_LIM_RTT_02</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test measures the RTT over the entire Limassol network chain, i.e. between i) the 4G UE and ii) a virtual endpoint (VM) deployed at the core network segment. The traffic traverses the 4G RAN, EPC and satellite backhaul, as well as an (uncontrolled) path through the Internet in order to reach the platform core.</td></tr>
<tr><td>Purpose</td><td colspan="4">This test assesses the performance (in terms of latency) of the entire network chain, including the latency of the 4G RAN, EPC and satellite backhaul.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>SHC</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in lab environment, with the 4G UE and the eNB at ~5 m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5 MHz (25 PRBs). ICMP packets of 32 bytes were used. The satellite backhaul is configured at 15 Mbps DL, 5 Mbps UL, guaranteed BW.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
</table>

<table>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>COTS UE (4G USB dongle), SDR-based eNB with Eurecom OAI, Athonet EPC, edge node, iDirect satellite terminal, core NFVI infrastructure</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td>Latency (RTT)</td></tr>
<tr><td>Additional tools involved</td><td>Ubuntu-based VMs acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Round Trip time
<table>
<tr><th colspan="3">RTT [ms]</th></tr>
<tr><th rowspan="2">Mean</th><th colspan="2">95% confidence interval for Mean</th></tr>
<tr><th>Lower bound</th><th>Upper bound</th></tr>
<tr><td>650.2</td><td>649.9</td><td>650.6</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

9.2.2. E2E Application Layer Latency

9.2.2.1. Summary and discussion of results for the end-to-end Application Layer latency tests

The objective of the latency tests at the application level has been to measure the E2E delay between the IoT Use Case components over the Limassol platform network. This will allow to analyze later on the behavior of the real IoT traffic over the infrastructure including the satellite backhaul.

Moreover, the objective is to compare the latency in an application environment with two configurations: with the virtual part of the gateway instantiated at the edge 4G network segment (4G and EPC at the edge), and with the same component instantiated at the core, measuring the whole E2E path, including the satellite domain. The purpose of this comparison is to enlight the improvements in the latency when the virtual functions are moved to the edge, justifying the implementation a real MEC with LBO for the future enhancements into the 5G infrastructure.

In this measurement campaign we must have in account not only the network latency of the infrastructure, but also the delay produced by the arrival to the concrete virtual resource within the edge/core. In addition, we must account on the capsulation and encapsulation of the data packet, as these measurements are taken at the application level.

The obtained latency values, presented in the following sections, were close to the ones expected, i.e. around 70 ms for the 4G network at the edge, and around 350 ms for the E2E path including the satellite backhaul. Once again, the difference implies a high reduction of the perceived latency for the traffic that is locally handled at the satellite edge instead of traveling until the backhaul, and emphasizes the value of the LBO functionalities at the satellite edge.

9.2.2.2. E2E Application Layer Latency between IoT Physical Gateway (UE) and Virtual Gateway (VNF) at the edge

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_LIM_LAT_IoT_01</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test measures the latency over the satellite edge segment, i.e. between i) the IoT Physical Gateway with 4G connection to ii) its counterpart the virtual gateway function (VNF) deployed at the edge (MEC) server by means of a containerization. Within this test the latency between devices and the gateway it is neglected.</td></tr>
<tr><td>Purpose</td><td colspan="4">This test assesses the performance (in terms of latency) of the IoT application running over the 4G network at the edge, including the latency of the 4G RAN and EPC.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UPV</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in a lab environment, with the 4G UE and the eNB at ~5 m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5 MHz (25PRBs). Packets exchanged were from different sizes as real application traffic was used by the experiment.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">Physical Gateway (Raspberry Pi-based), 4G USB dongle, SDR-based eNB with Eurecom OAI, Athonet EPC, the edge node, Virtual Gateway function instance.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Latency</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Node-RED dashboard to visualize the timestamps and values obtained.</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Uplink Latency
<table>
<tr><th colspan="3">Latency [ms]</th></tr>
<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>69.6</td><td>67.8</td><td>71.4</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

9.2.2.3. E2E Application Layer Latency between IoT Physical Gateway (UE) and Virtual Gateway (VNF) at the core.

Test Case ID	TC_LIM_LAT_IoT_02
General description of the test	The test measures the uplink latency over the entire Limassol network chain, i.e. between i) the IoT Physical Gateway with 4G connection to ii) its counterpart the virtual gateway function (VNF) deployed at the

<table>
<tr><td></td><td colspan="4">core network segment. The traffic traverses the 4G RAN, EPC and satellite backhaul until reaching the core, and there to the VM that contains the VNF. Within this test the latency between devices and the gateway it is neglected.</td></tr>
<tr><td>Purpose</td><td colspan="4">This test assesses the performance (in terms of latency) of the IoT application running over the whole 4G Limassol Infrastructure, including the latency of the 4G RAN, EPC and satellite backhaul.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>UPV</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in lab environment, with the 4G UE and the eNB at ~5m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5MHz (25PRBs). Packets exchanged were from different sizes as real application traffic was used by the experiment. The satellite backhaul is configured at 15 Mbps DL, 5 Mbps UL, guaranteed BW.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">Physical Gateway (Raspberry Pi-based), 4G USB dongle, SDR-based eNB with Eurecom OAI, Athonet EPC, iDirect satellite terminal, core NFVI infrastructure, Virtual Gateway function instance.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Latency</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Node-RED dashboard to visualize the timestamps and values obtained.</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Uplink Latency
<table>
<tr><th colspan="3">Latency [ms]</th></tr>
<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>347.7</td><td>342.1</td><td>353.3</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

9.2.3. Throughput

9.2.3.1. Summary and discussion of results for the throughput tests

The aim of the throughput tests has been to measure the E2E download (DL) and upload (UL) throughput between several points in the Limassol platform network. As in the RTT tests, the focus has been in two configurations: throughput measured across the edge 4G network segment (4G and EPC at the edge) and across the whole E2E path, including the satellite domain, the Internet and the core network (see topology in the previous section).

The derived throughput values, as presented in the next sections, were close to the ones expected and were actually corresponding to the nominal capacity of the backhaul and access links. The satellite link has been configured at 15 Mbps DL / 5 Mbps UL, and also the performance of the RAN (at 5 MHz/25 PRBs) is somewhat lower (~10 Mbps DL). That is, the experimental values verify the set expectations.

Unlike the RTT measurements, in the throughput case there is no differentiation between the measurements with the edge and the core. That is because, in the current configuration, the satellite backhaul does not constitute a bottleneck in the E2E chain. In other words, the local breakout function does not bring an added value. In future platform releases, when the capacity of the cellular RAN will increase, we expect to observe a significant difference.

9.2.3.2. Throughput between UE and satellite edge

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_LIM_THR_01</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test measures the download and upload throughput over the satellite edge segment, i.e. between i) the 4G UE and ii) a virtual endpoint (VM) deployed at the edge.</td></tr>
<tr><td>Purpose</td><td colspan="4">This test assesses the performance (in terms of throughput) at the 4G network at the edge, including the latency of the 4G RAN and EPC.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>SHC</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in lab environment, with the 4G UE and the eNB at ~5 m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5 MHz (25 PRBs). ICMP packets of 32 bytes were used.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">COTS UE (4G USB dongle), SDR-based eNB with Eurecom OAI, Athonet EPC, edge node</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Throughput*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Ubuntu-based VMs acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Download (VM @ edge -> UE)
<table>
<tr><td colspan="3">Throughput (Mbps)</td></tr>
<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>9.09</td><td>8.74</td><td>9.44</td></tr>
</table>
Upload (UE -> VM @ edge)
<table>
<tr><td colspan="2">Throughput (Mbps)</td></tr>
<tr><td>Mean</td><td>95% confidence interval for Mean</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td><table><tr><td></td><td>Lower bound</td><td>Upper bound</td></tr><tr><td>4.09</td><td>3.92</td><td>4.26</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

9.2.3.3. Throughput between UE and platform core

<table>
<tr><td>Test Case ID</td><td colspan="4">TC_LIM_THR_02</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test measures the download and upload throughput over the entire Limassol network chain, i.e. between i) the 4G UE and ii) a virtual endpoint (VM) deployed at the core network segment. The traffic traverses the 4G RAN, EPC and satellite backhaul, as well as an (uncontrolled) path through the Internet in order to reach the platform core.</td></tr>
<tr><td>Purpose</td><td colspan="4">This test assesses the performance (in terms of throughput) of the entire network chain, including the latency of the 4G RAN, EPC and satellite backhaul.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>SHC</td><td>Date:</td><td>05.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">SHC, PTL, AVA, UPV</td></tr>
<tr><td>Scenario</td><td colspan="4">The test is performed indoors, in lab environment, with the 4G UE and the eNB at ~5 m apart, line-of-sight and SISO configuration. Operation in LTE Band 7, 5 MHz (25 PRBs). ICMP packets of 32 bytes were used. The satellite backhaul is configured at 15 Mbps DL, 5 Mbps UL, guaranteed BW.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">No slicing used in Phase 1.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">COTS UE (4G USB dongle), SDR-based eNB with Eurecom OAI, Athonet EPC, edge node, iDirect satellite terminal, core NFVI infrastructure</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">Ubuntu-based VMs acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Download (VM @ platform core -> UE)<table><tr><td colspan="3">Throughput (Mbps)</td></tr><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>9.81</td><td>9.48</td><td>10.15</td></tr></table>Upload (UE -> VM @ platform core)<table><tr><td>Throughput (Mbps)</td></tr></table></td></tr>
</table>

<table>
<tr><td></td><td><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>4.41</td><td>4.33</td><td>4.48</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

10. SURREY PLATFORM EXPERIMENTS

10.1. Overview

The goal of the Surrey Platform for the Phase1 is to provide initial evaluations of the 5G KPIs.

Table 21 recaps the components planned to be available for the first integration and experimentation cycle and puts then in relation to the achievements of the first trial phase. The Surrey Platform in cycle 1, does not yet have capability to execute unattended and attended experiments, although it integrates commercial components (5G rel. 15 compliant RAN & core segments). The initial measurements (in cycle 1) address the following KPIs:

- Peak Throughput (single-user).

Table 21 Experimentation methodology components in the first integration cycle for the Berlin Platform, according to D2.3 [12]

Experimentation methodology component	Plan for integration and trial Phase 1	Status / Trial Phase I achievements
Open API's + Dispatcher	No	Not started. Expect availability by Q2-2020.
Experiment Life Cycle	No	ELCM implementation currently ongoing, expect availability by Q4-2019.
Portal	No	Portal implementation currently ongoing, expect availability by Q4-2019.
Custom experiments	No	The surrey Platform will support custom experiments over an E2E 5G network core, i.e., the network core (Rel.15) supporting the separation of control and data plane, while still allowing LTE base stations and UEs to be used as the anchor, in NSA configuration. Custome experimentation availability by Q1-2020.
Standard experiments	No	Expect availability by Q1-2020.
E2E slices	No	Expect availability by Q1-2020.
VNF's	Yes	The surrey Platform will support dynamic placement of VNFs in the testbed (cycle 2). Placement of VNFs may be triggered by the orchestration tool as well as by Keysight's TAP, which is chosen in 5GENESIS to execute and control experiments. Expect availability by Q4-2019.
Scenarios	POC	The instantiation of the Surrey Platform for the initial (limited) mMTC trial (established in the lab).

Un-attended experiments	POC: Delay and Throughput	Experiments conducted in cycle 1 on the surrey Platform are not automated and controlled by TAP. Expect availability by Q1-2020.
Attended experiments	POC: 4G Delay and Throughput	The Surrey Platform supports outdoor experiments using 5G RAN connected to the 5G packet core (separation of control and data plane existing) Expect availability by Q1-2020.
Security Manager	NA	-

Table 22 lists the planned KPIs and measurements.

Table 22 Primary 5G KPIs evaluated at the surrey Platform in the first trial

KPI to be evaluated at the surrey Platform according to DoW	Evaluated in Phase 1 / First Trial	Comment
Density of Users	No	scheduled for Phase 2
Service Creation Time	No	scheduled for Phase 2
Reliability	No	scheduled for Phase 3
Energy efficiency	No	Q2-2020 / Phase 2
Additional 5G KPIs evaluated at the Surrey Platform	**Evaluated in Phase 1 / First Trial**	**Comment**
Delay (RTT)	No	scheduled for Phase 2
Throughput (peak)	Yes	scheduled for Phase 1

Peak throughput experiments were executed on the instantiation of the Surrey Platform as illustrated in Figure 20. Note that the current 5G network deployment consist of 3 x BBU each having 2 or 3 sectors distributed to various location on the university site. All tests were performed using a “CPE1.0” unit (with no other UEs present), operating in 3.5 GHz band over 100 MHz bandwidth (the max. supported throughput by CPE 1.0 is limited due to hardware limitations). The Surrey 5GENESIS team is currently engaged in negotiations to procure one or more “CPE-pro” units that are cable of reaching greater than 1 Gbps througputs, for cycle 2 experimentations.

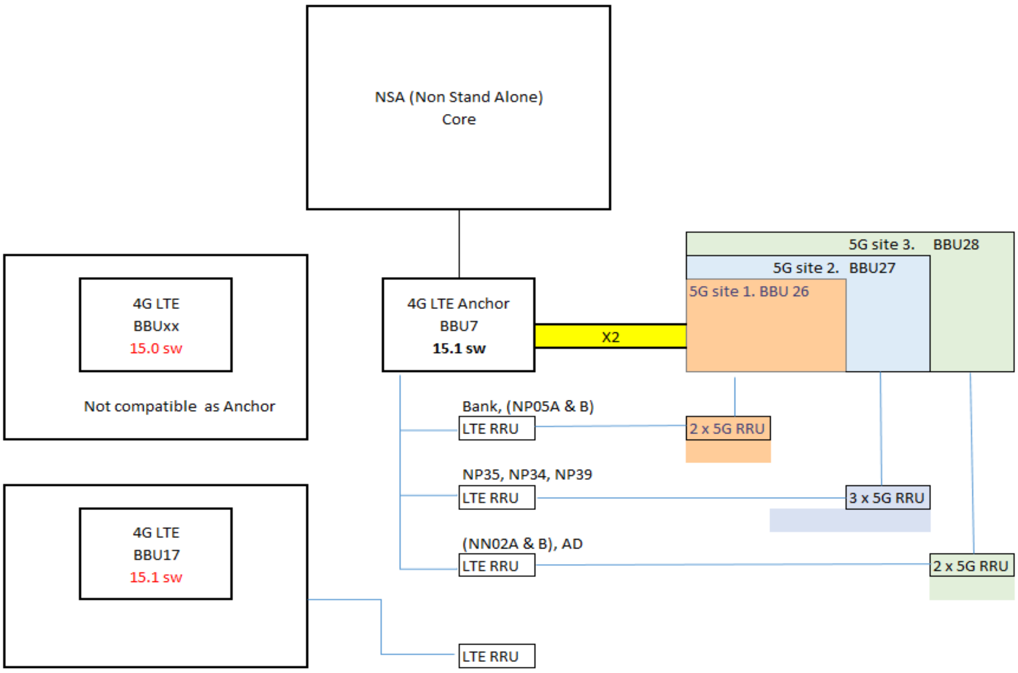


Figure 20 Instantiation of the 5GENESIS Architecture for the Surrey Platform Phase I Trials.

10.2. Experiments and results

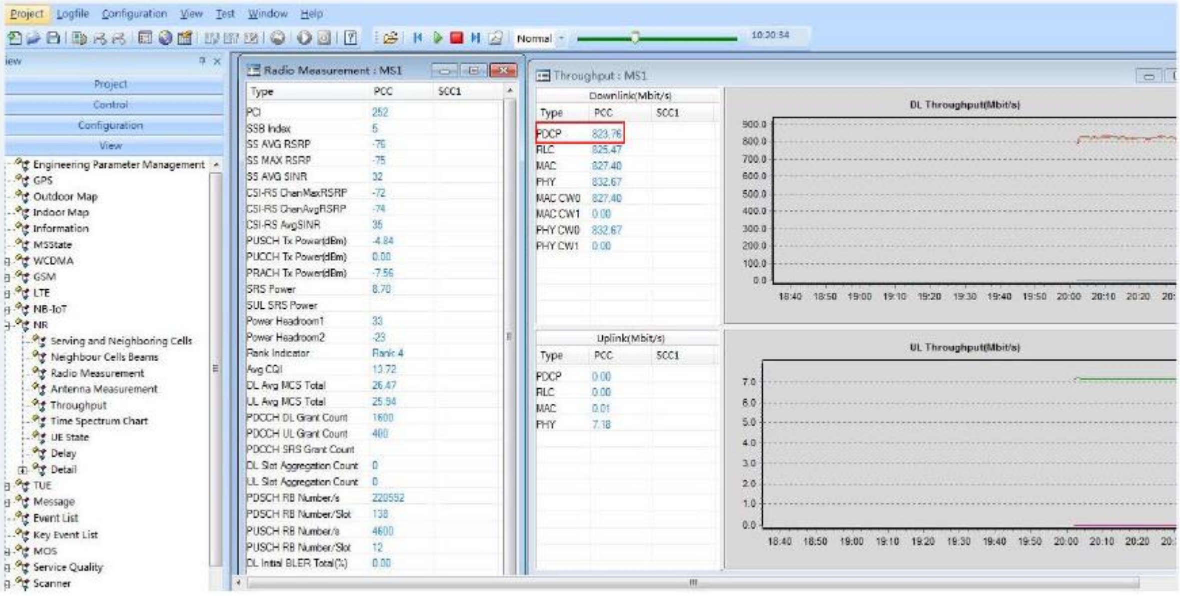
10.2.1. Peak Throughput

10.2.1.1. Summary and discussion of results for the peak throughput tests

These experiments are devoted to test and validate the infrastructure in a first step in order to assess if the envisioned 5G peak-throughput KPIs can be achieved with the platform's infrastructure regardless of deployed CN infrastructure. The raw measurements results are depicted in Table 23. The following figures provide more information on some relevant aspects of the Surrey Platform setup and results.

Table 23 Raw results from experiments (18 iterations) outdoors

5G BBU	NR cell ID	DL peak UDP rate	DL peak TCP rate	position
BBU-26	252	823	723	near window in test lab 01
BBU-26	253	826	726	BBU-26-253
BBU-26	253	822	721	BBU-26-253
BBU-26	254	820	723	Bbu-26-254
BBU-26	254	823	724	Bbu-26-254
BBU-26	252	820	720	near window in test lab 01
BBU-27	255	819	716	BBU-27-255
BBU-27	256	809	667	BBU-27-256
BBU-27	256	807	664	BBU-27-256
BBU-27	257	824	702	BBU-27-257
BBU-27	257	822	701	BBU-27-257
BBU-27	255	821	715	BBU-27-255
BBU-28	258	810	702	BBU-27-258
BBU-28	259	802	682	BBU-27-259
BBU-28	259	800	680	BBU-27-259
BBU-28	260	821	691	BBU-27-260
BBU-28	260	820	690	BBU-27-260
BBU-28	258	808	700	BBU-27-258



Radio Measurement : MS1

Type	PCC	SCC1
PCI	252	
SSB Index	5	
SS AVG RSRP	-76	
SS MAX RSRP	-75	
SS AVG SINR	32	
CSI-RS ChanMaxRSRP	-72	
CSI-RS ChanAvgRSRP	-74	
CSI-RS AvgSINR	35	
PUSCH Tx Power(dBm)	-4.84	
PUCCH Tx Power(dBm)	0.00	
PRACH Tx Power(dBm)	-7.56	
SRS Power	8.70	
SUL SRS Power		
Power Headroom1	33	
Power Headroom2	-23	
Rank Indicator	Rank 4	
Avg CQI	13.72	
DL Avg MCS Total	26.47	
UL Avg MCS Total	25.94	
PDCCH DL Grant Count	1600	
PDCCH UL Grant Count	400	
PDCCH SRS Grant Count		
DL Slot Aggregation Count	0	
UL Slot Aggregation Count	0	
PDSCH RB Number/s	220552	
PDSCH RB Number/Slot	138	
PUSCH RB Number/s	4600	
PUSCH RB Number/Slot	12	
DL Initial BLER Total(%)	0.00	

Throughput : MS1

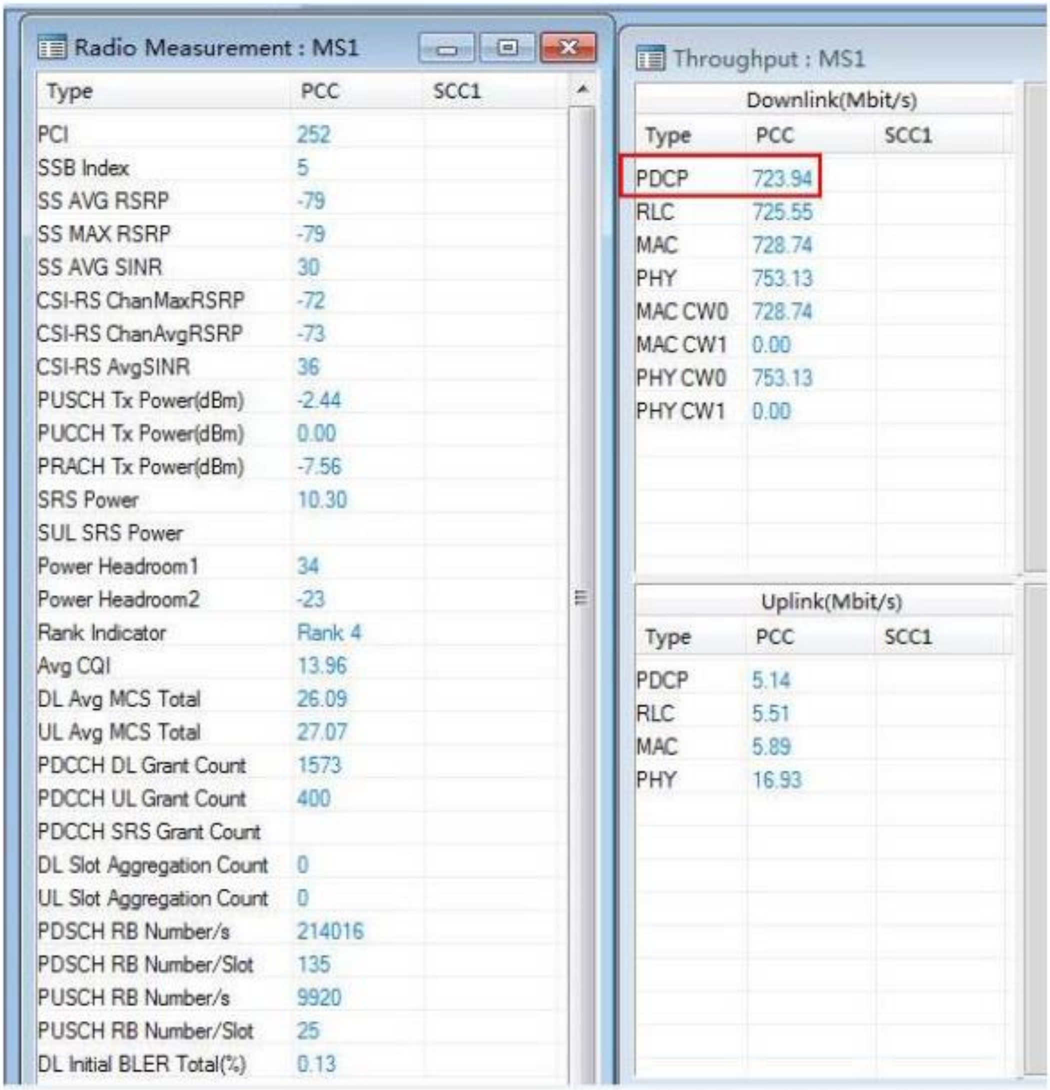
Downlink(Mbit/s)

Type	PCC	SCC1
PDCP	823.76	
RLC	825.47	
MAC	827.40	
PHY	832.67	
MAC CW0	827.40	
MAC CW1	0.00	
PHY CW0	832.67	
PHY CW1	0.00	

Uplink(Mbit/s)

Type	PCC	SCC1
PDCP	0.00	
RLC	0.00	
MAC	0.01	
PHY	7.18	

Figure 21 Results snapshot (BBU 26 NR cell 252) - Surrey Platform Phase I Trials, for UDP peak traffic.



Radio Measurement : MS1

Type	PCC	SCC1
PCI	252	
SSB Index	5	
SS AVG RSRP	-79	
SS MAX RSRP	-79	
SS AVG SINR	30	
CSI-RS ChanMaxRSRP	-72	
CSI-RS ChanAvgRSRP	-73	
CSI-RS AvgSINR	36	
PUSCH Tx Power(dBm)	-2.44	
PUCCH Tx Power(dBm)	0.00	
PRACH Tx Power(dBm)	-7.56	
SRS Power	10.30	
SUL SRS Power		
Power Headroom1	34	
Power Headroom2	-23	
Rank Indicator	Rank 4	
Avg CQI	13.96	
DL Avg MCS Total	26.09	
UL Avg MCS Total	27.07	
PDCCH DL Grant Count	1573	
PDCCH UL Grant Count	400	
PDCCH SRS Grant Count		
DL Slot Aggregation Count	0	
UL Slot Aggregation Count	0	
PDSCH RB Number/s	214016	
PDSCH RB Number/Slot	135	
PUSCH RB Number/s	9920	
PUSCH RB Number/Slot	25	
DL Initial BLER Total(%)	0.13	

Throughput : MS1

Downlink(Mbit/s)

Type	PCC	SCC1
PDCP	723.94	
RLC	725.55	
MAC	728.74	
PHY	753.13	
MAC CW0	728.74	
MAC CW1	0.00	
PHY CW0	753.13	
PHY CW1	0.00	

Uplink(Mbit/s)

Type	PCC	SCC1
PDCP	5.14	
RLC	5.51	
MAC	5.89	
PHY	16.93	

Figure 22 Results snapshot (BBU 26 NR cell 252) - Surrey Platform Phase I Trials, for TCP peak traffic.


BBU26-252-position

Near the window in test bed lab 01 (Bank)

BBU26-253-position

BBU26-254-position

BBU27-255-position

BBU27-256-position

BBU27-257-position

BBU28-258-position

BBU28-259-position

BBU 28-260-position

Figure 23 physical location of test calls using CPE 1.0

Results show that the UDP peak data rate reaches almost 800 Mbps (commensurate with maximum performance of CPE 1.0), and that the peak TCP rate reaches 700 Mbps.

Test Case ID	TC-Thr-001			
General description of the test	The test assesses the average, minimum, and maximum of single-user peak throughput between two communication end-points i.e. CPE & gNB.			
Purpose	The test is used to primarily assess the achievable peak throughput.			
Executed by	Partner:	UNIS	Date:	15.MAY.2019

<table>
<tr><td>Involved Partner(s)</td><td>UNIS</td></tr>
<tr><td>Scenario</td><td>The communication endpoints of the service are both within availability zone.</td></tr>
<tr><td>Slicing configuration</td><td>N/A</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>RRH, BBU, CPE 1.0.</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td>*Peak throughput*</td></tr>
<tr><td>Additional tools involved</td><td>Vendor-specific tool</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Peak throughput
<table>
<tr><th colspan="4">Peak-Tput [Mbps] - UDP</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Higher bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>812.5</td><td>816.5</td><td>820.5</td></tr>
<tr><td>Minimum</td><td>-</td><td>800</td><td>-</td></tr>
<tr><td>Maximum</td><td>-</td><td>826</td><td>-</td></tr>
</table>
<table>
<tr><th colspan="4">Peak-Tput [Mbps] - TCP</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Higher bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>692.7</td><td>702.6</td><td>712.6</td></tr>
<tr><td>Minimum</td><td>-</td><td>664</td><td>-</td></tr>
<tr><td>Maximum</td><td>-</td><td>726</td><td>-</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

10.2.2 Service Creation Time for IoT HTTP-UDP and MQTT-UDP Virtual Functions

10.2.2.1 Summary and discussion of results for service creation time

Figure 24 shows the workflow of the information within the Surrey platform. First, sensors carried by users when visiting a large-scale event on campus will collect sensing data including information about temperature, air quality, presence, movement, acceleration, and other parameters. This data is collected and transmitted using one or more of the available air interfaces and is then passed to the IoT virtual gateway that understands and translates the various incoming IoT protocols into UDP-over-IP packets, and forwards the data to the Surrey server.

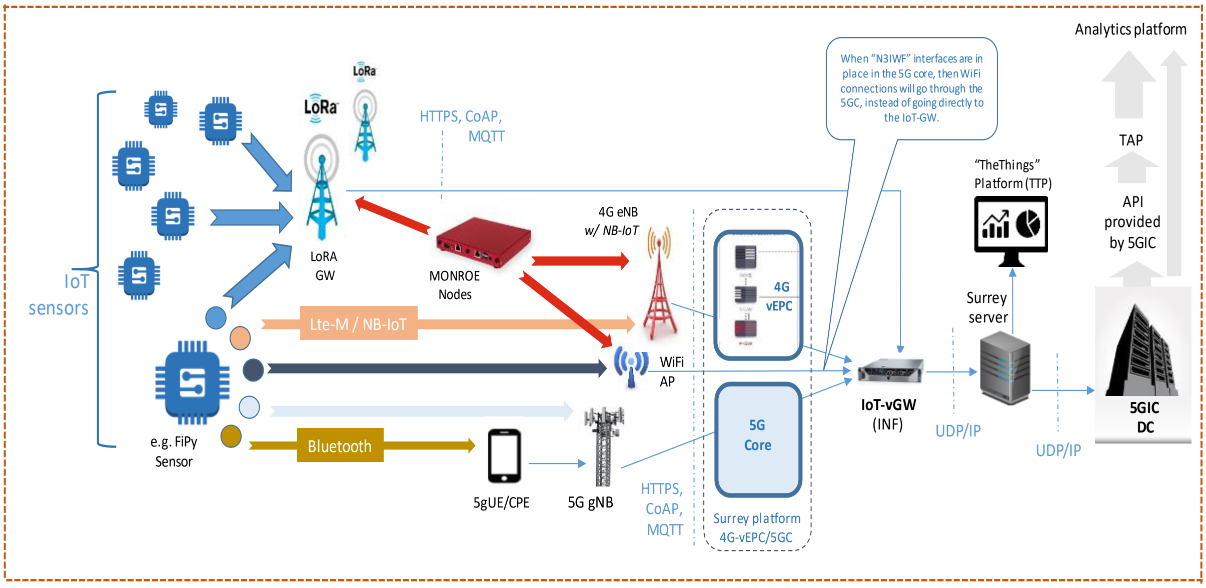


Figure 24: Network topology for the use case scenario at the 5GIC/ICS Site

The IoT vGW is mainly composed by mapping functions, which in the form of VNFs are deployed in the underlying NFVI infrastructure, in order to intervene between the GW and the IoT sensors to translate the various underlying data protocols to a common universal data protocol. This translation provides interoperability on top of plenty of IoT data sensors, allowing its use as part of the same experiment, even if they are using different data protocols.

During phase 1 of the project, two mapping functions (HTTP-UDP and MQTT-UDP) have been deployed as VNFs at the NFVI of the Surrey platform, which currently combines the two main technologies i.e. OSM and OpenStack as depicted in Figure 25.

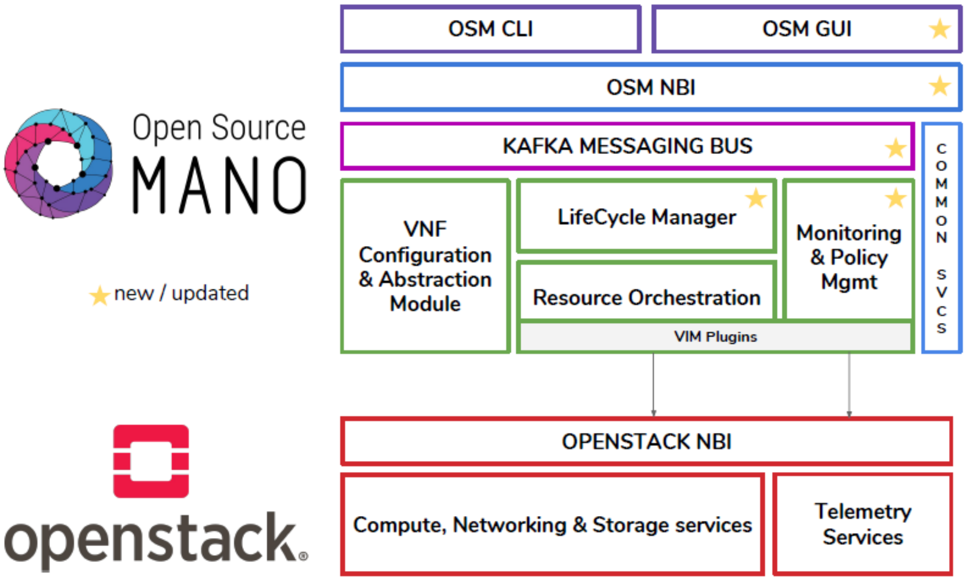
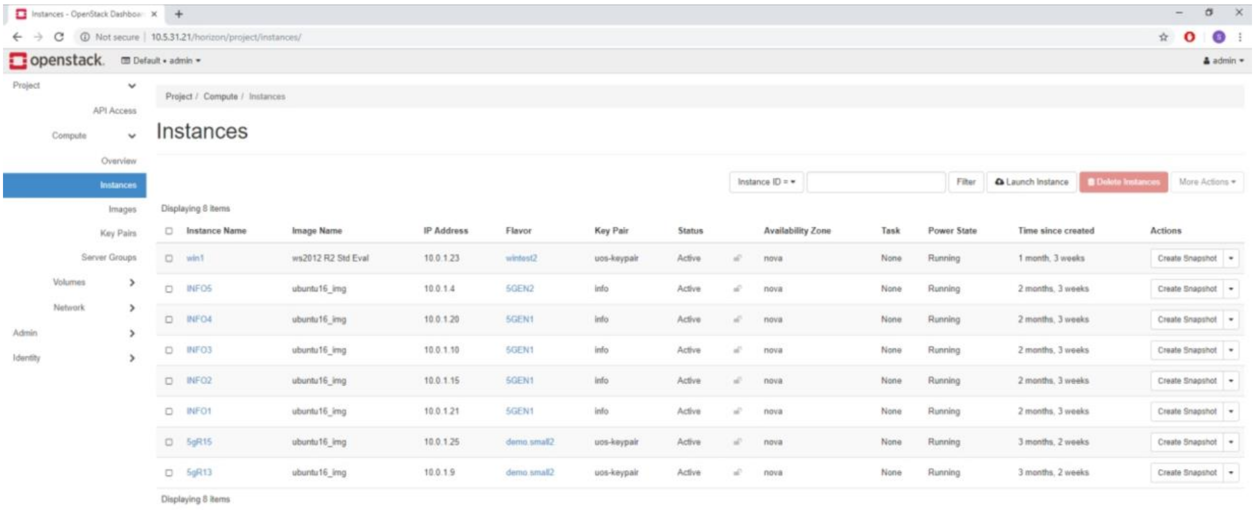


Figure 25: The architecture as adopted by the Surrey platform

Figure 26 depicts the instances of the VMs that were deployed at the Surrey NFVI, where VMs INF0x are used for deploying and testing the mapping functions as well as the IoT vGW component.



Instances

Displaying 8 items

☐	Instance Name	Image Name	IP Address	Flavor	Key Pair	Status		Availability Zone	Task	Power State	Time since created	Actions
☐	win1	ws2012 R2 Std Eval	10.0.1.23	wintest2	uos-keypair	Active		nova	None	Running	1 month, 3 weeks	Create Snapshot
☐	INF05	ubuntu16_img	10.0.1.4	5GEN2	info	Active		nova	None	Running	2 months, 3 weeks	Create Snapshot
☐	INF04	ubuntu16_img	10.0.1.20	5GEN1	info	Active		nova	None	Running	2 months, 3 weeks	Create Snapshot
☐	INF03	ubuntu16_img	10.0.1.10	5GEN1	info	Active		nova	None	Running	2 months, 3 weeks	Create Snapshot
☐	INF02	ubuntu16_img	10.0.1.15	5GEN1	info	Active		nova	None	Running	2 months, 3 weeks	Create Snapshot
☐	INF01	ubuntu16_img	10.0.1.21	5GEN1	info	Active		nova	None	Running	2 months, 3 weeks	Create Snapshot
☐	5gR15	ubuntu16_img	10.0.1.25	demo.small2	uos-keypair	Active		nova	None	Running	3 months, 2 weeks	Create Snapshot
☐	5gR13	ubuntu16_img	10.0.1.9	demo.small2	uos-keypair	Active		nova	None	Running	3 months, 2 weeks	Create Snapshot

Displaying 8 items

Figure 26: The deployed VMs at the NFVI of Surrey Platform

The VMs INF02 and INF03 were used for measuring the deployment time of HTTP-UDP and MQTT-UDP VNFs, following the test case TC-Ser-001. Each VM had been allocated the following resources: 2 core INTEL XEON 2.3 GHz and 3GB RAM.

10.2.1.2. Service Creation Time IoT HTTP-UDP Mapping Function

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">Service Creation Time of IoT mapping HTTP to UDP VNF for the IoT interoperability provision during use-case 1 of Surrey Platform</td></tr>
<tr><td>Purpose</td><td colspan="4">The measurement of the time needed for the mapping function HTTP-UDP to be deployed at the Surrey platform’s NFVI PoP in order to assess the time constraints that the provision of such a function is needed.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>INFOLYSiS</td><td>Date:</td><td>02.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">INFOLYSiS, UNIS</td></tr>
<tr><td>Scenario</td><td colspan="4">• A mapping VNF HTTP-to-UDP deployed at Openstack/NFVI of Surrey Platform
• Number of iterations: 25
• VM Configuration: 2 core INTEL XEON 2.3GHz, 3GB RAM</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">N/A</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">• NFVI Infrastructure at Surrey Platform
• OpenStack
• HTTP-to-UDP mapping VNF</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Service Creation Time</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">N/A</td></tr>
</table>

<table>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td><table><tr><th colspan="4">Service Creation Time [s]</th></tr><tr><th></th><th>Lower Bound of 95% confidence interval</th><th>Mean</th><th>Upper Bound of 95% confidence interval</th></tr><tr><td>Mapping VNF HTTP-UDP</td><td>12.2</td><td>15.4</td><td>18.4</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>N/A</td></tr>
</table>

10.2.1.3. Service Creation Time IoT MQTT-UDP Mapping Function

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">Service Creation Time of IoT mapping MQTT to UDP VNF for the IoT interoperability provision during use-case 1 of Surrey Platform</td></tr>
<tr><td>Purpose</td><td colspan="4">The measurement of the time needed for the mapping function MQTT -UDP to be deployed at the Surrey platform's NFVI PoP in order to assess the time constraints that the provision of such a function is needed.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>INFOLYSiS</td><td>Date:</td><td>02.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">INFOLYSiS, UNIS</td></tr>
<tr><td>Scenario</td><td colspan="4">A mapping VNF MQTT-to-UDP deployed at OpenStack/NFVI of Surrey Platform.
Number of iterations: 25.
VM Configuration : 2 core INTEL XEON 2.3 GHz, 3GB RAM.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">N/A</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">NFVI Infrastructure at Surrey Platform
OpenStack
MQTT-to-UDP mapping VNF</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Service Creation Time</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">N/A</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4"><table><tr><th colspan="4">Service Creation Time [s]</th></tr><tr><th></th><th>Lower Bound of 95% confidence interval</th><th>Mean</th><th>Upper Bound of 95% confidence interval</th></tr><tr><td>Mapping VNF MQTT-UDP</td><td>12.6</td><td>15.7</td><td>18.8</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">N/A</td></tr>
</table>

11. BERLIN PLATFORM EXPERIMENTS

11.1. Overview

The goal of the Berlin Platform for the first trial phase is to:

- validate the correct functionality of the experimentation methodology components provided by WP3 and WP4 for the first integration cycle,
- assess the performanc of the network infrastructure of the Berlin Platform towards its suitability for conducting further 5G E2E KPI evaluations, and
- conduct initial evaluations (calibration tests) for the 5G KPIs evaluated by the Berlin Platform.

Table 24 recaps the components planned to be available for the first integration and experimentation cycle and puts then in relation to the achievements of the first trial phase. All components planned for cycle 1 were available and successfully trialed. Besides, the maturity of components exceeds the plan for phase 1 in regard that all experiments are controlled and conducted using a commercial tool, i.e. Keysight's TAP. As such, the Berlin Platform does not use as planned a proof-of-concept for executing unattended and attended experiments, but integrates with a commercial test automation platform, as agreed within the 5GENESIS consortium.

Besides, measurements for the first of the four 5G KPIs, which are to be evaluated at the Berlin Platform, were conducted, namely covering:

- Service Creation time (c.f. Section 5.1.8).

In surplus, initial measurements to assess

- Delay (Round-Trip-Time, c.f. Section5.1.5), and
- Throughput (c.f. Section 5.1.10)

thus exceeding the contractual obligations per DoW.

Table 24 Experimentation methodology components in the first integration cycle for the Berlin Platform, according to D2.3 [12]

Experimentation methodology component	Plan for integration and trial Phase 1	Status / Trial Phase I achievements
Open API's + Dispatcher	No	-
Experiment Life Cycle	No	-
Portal	No	-
Custom experiments	POC	The Berlin Platform supports custom experiments over an E2E (pre) 5G network core, i.e., the network core (OpenEPC Rel.3) supports the separation of control and data plane, while

		still allowing LTE BSs and UEs to be used as the RAN.
Standard experiments	No	-
E2E slices	No	-
VNF's	Yes	The Berlin Platform supports dynamic placement of VNFs in the testbed. Placement of VNFs may be triggered by the orchestration tool as well as by Keysight's TAP, which is chosen in 5GENESIS to execute and control experiments.
Scenarios	POC (mmWave backhauling)	The instantiation of the Berlin Platform for the first trial provides mmWave backhaul links (established in the lab).
Un-attended experiments	POC: Delay and Throughput evaluations of the core testbed infrastructure	All experiments conducted in Phase I on the Berlin Platform are fully automated and controlled by TAP.
Attended experiments	POC: 4G Delay and Throughput	The Berlin Platform supports in the lab _unattended_ experiments using 4G RAN connected to a pre-5G packet core (separation of control and data plane existing)
Security Manager	NA	

Table 25 lists the KPIs evaluated in the first trial and summarizes the kind of evaluation measurements conducted.

Table 25 Primary 5G KPIs evaluated at the Berlin Platform in the first trial

KPI to be evaluated at the Berlin Platform according to DoW	Evaluated in Phase 1 / First Trial	Comment
Density of Users	no	Not scheduled for Phase 1
Service Creation Time	yes	Calibration test to deploy a predefined VM as a VNF providing "virtual instrument functionality" for test Dynamic provisioning of full Open5GCore Rel.3 providing E2E network connectivity.
Speed	no	Not scheduled for Phase 1
Reliability	no	Not scheduled for Phase 1

Additional 5G KPIs evaluated at the Berlin Platform	Evaluated in Phase 1 / First Trial	Comment
Delay (RTT)	yes	-
Throughput	yes	-

All experiments were executed on the instantiation of the Berlin Platform as illustrated in Figure 27. As such, the Berlin Platform provides for the first trial phase a fully orchestrated testbed spanning across two geographical distinct sites – namely Fraunhofer FOKUS and IHP – and allowing for deterministic, dynamic placement of VNFs in each of the seven availability zones. As the networks at the Fokus and IHP site operate under a single, trusted management domain from the testbed perspective, a single OpenStack instance has been deployed across the two sites interconnected via GEANT, which allows as well for life-migration of VNFs.

For the execution of test cases assessing throughput or delay, measurement endpoints – also denoted to as “virtual instruments” – could be dynamically and deterministically instantiated at either availability zone. Table 26 shows the combination of any two measurement points and the resulting (E2E) link assessed therefore.

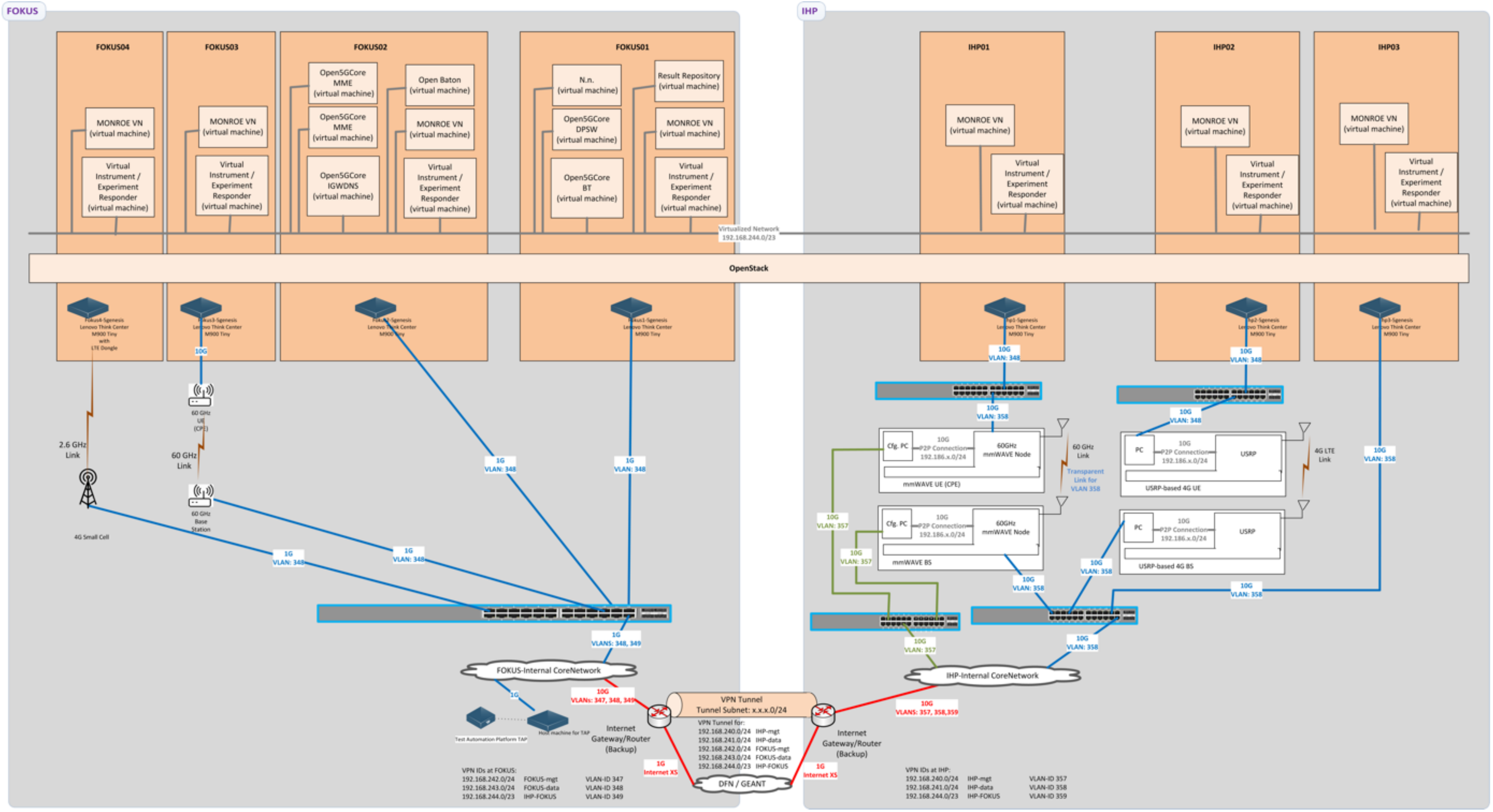


Figure 27 Instantiation of the 5GENESIS Architecture for the Berlin Platform Phase I Trials.

Availability Zone	FOKUS04	FOKUS03	FOKUS02	FOKUS01	IHP01	IHP02	IHP03	Comment
FOKUS04	x							Assessment of the performance of the compute node, i.e. the performance offered by OpenStack among instances deployed in the same machine
FOKUS03		x						
FOKUS02			x					
FOKUS01				x				
IHP01					x			
IHP02						x		
IHP03							x	
FOKUS04		x						E2E link including a LTE RAN, the packet core, and a 60 GHz backhaul link
FOKUS04			x					Placement of one service endpoint as a VNF within the core network. Link includes a LTE RAN, the packet core.
FOKUS04				x				Placement of one service endpoint as a VNF within the core network. Link includes a LTE RAN, the packet core.
FOKUS04					x			E2E link including a LTE RAN, packet core, wide area internet connection (inter-data-center connection) and a mmWAVE backhaul
FOKUS04						x		E2E link including two LTE RANs (one at each zone), a packet core, and wide area internet connection (inter-data-center connection)
FOKUS04							x	E2E link including a LTE RAN, packet core, wide area internet connection (inter-data-center connection)
FOKUS03			x					E2E link including a 60GHz mmWave backhaul
FOKUS03				x				E2E link including a 60GHz mmWave backhaul
FOKUS03					x			E2E link including two mmWave backhauls
FOKUS03						x		E2E link including a 60GHz backhaul, packet core, and LTE RAN
FOKUS03							x	E2E-link including a 60GHz backhaul and wide area internet connection (inter-data-center connection)

FOKUS02				x				E2E link between two VNFs placed in the same data center but at different compute nodes
FOKUS02					x			E2E link including a wide area internet connection (inter-data-center connection) and mmWave backhaul
FOKUS02						x		E2E link including a packet core, wide area internet connection (inter-data-center connection) and LTE RAN
FOKUS02							x	E2E link between two data centers (i.e. wide area internet connection (inter-data-center connection))
FOKUS01					x			E2E link including a wide area internet connection (inter-data-center connection) and a mmWAVE backhaul
FOKUS01						x		E2E link including a packet core, wide area internet connection (inter-data-center connection), and LTE RAN
FOKUS01							x	E2E link including a wide area internet connection (inter-data-center connection)
IHP01						x		E2E link including a mmWAVE backhaul, two wide area internet connection (inter-data-center connection)[18], a packet core, and a LTE RAN
IHP01							x	E2E link including a mmWAVE backhaul
IHP02							x	E2E link including a LTE RAN, two wide area internet connection (inter-data-center connection), and packet core

Table 26 Possible E2E links assessable via a measurement campaign in the first trial at the Berlin Platform

---

[18] Note that in this set-up, the wide-area internet connection (inter-data-center connection) has to be traversed twice as the packet core is instantiated at Fraunhofer FOKUS.

The following figures show selected testbed components of the Berlin platform, i.e. the 60 GHz prototype at IHP (Figure 28) and the FemtoCell, 60 GHz cell and compute node at Fokus.

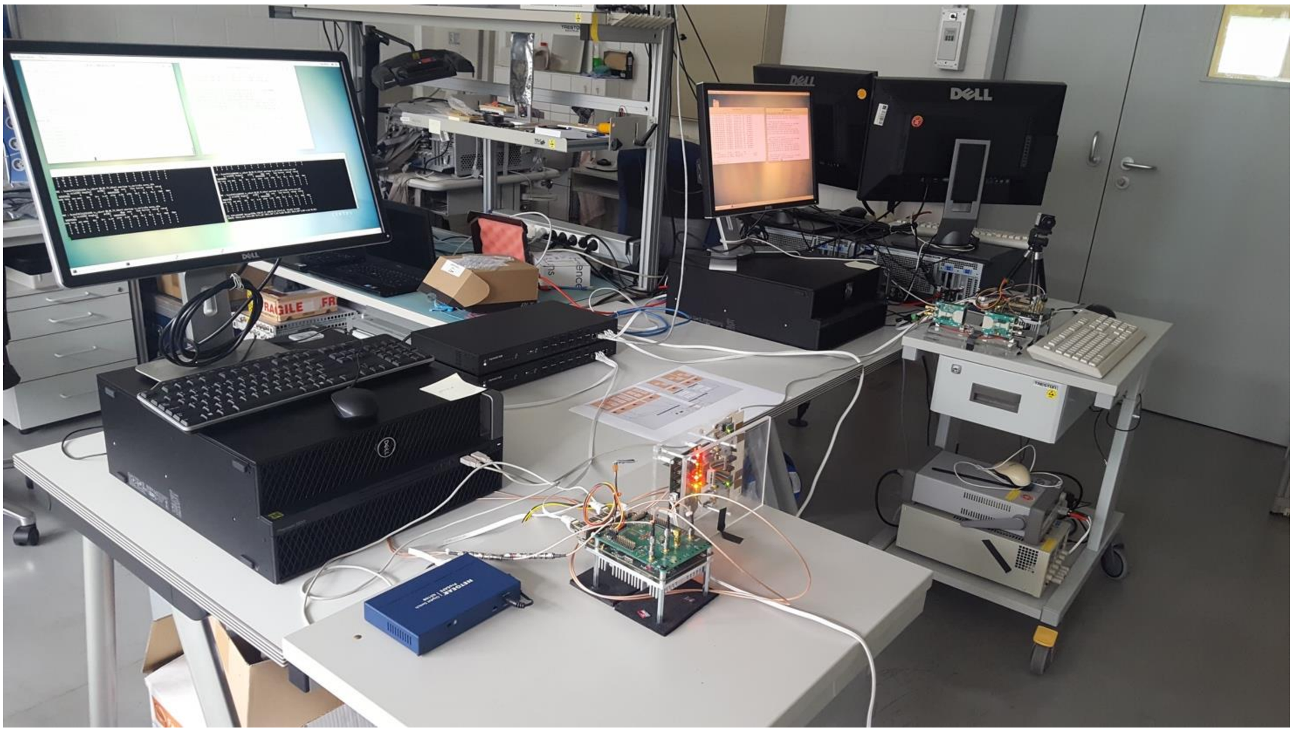


Figure 28 60 GHz link, one compute node

Figure 29 Compute nodes deployed at FOKUS for phase 1

Figure 30 60 GHz MetroLinq mmWave system deployed at FOKUS for Phase 1

Figure 31 AirSpan LTE Femto Cell deployed at FOKUS for Phase 1

Table 27 summarizes the technical specifications of key infrastructure components of the Berlin testbed:

Table 27 Different types of compute nodes deployed during Phase 1

Nodes	Model	CPU	Memory	Networking	OS	kernel
FOKUS-01, FOKUS-02	Lenovo ThinkCentre M910q	Intel Core i7-7700T CPU @ 2.90GHz	32 GB	1 Gbps Ethernet	centos-release-7-6.1810.2.el7.centos.x86_64	3.10.0-957.21.3.el7.x86_64 GNU/Linux
FOKUS-03	Lenovo ThinkCentre M900	Intel Core i7-6700T CPU @ 2.80GHz	16GB	1 Gbps Ethernet	centos-release-7-6.1810.2.el7.centos.x86_64	3.10.0-957.21.3.el7.x86_64 GNU/Linux
FOKUS-04	Lenovo ThinkCentre M900	Intel Core i7-6700T CPU @ 2.80GHz	16 GB	1 Gbps Ethernet	Ubuntu 16.04.6 LTS	4.4.0-142-generic GNU/Linux
FOKUS-05	Lenovo ThinkCentre M92	Intel Core i3-2120T CPU @ 2.60GHz	8 GB	1 Gbps Ethernet Huawei E3372	Ubuntu 16.04.6 LTS	4.4.0-142-generic GNU/Linux
IHP01 and IHP03	Dell Precision 5820	Intel Core i9-7980XE CPU @ 2.60GHz, CPUs 36 (threads 2, cores per socket 18, sockets 1)	32 GB	Intel 10G Ethernet controller X550T	CentOS7 64bit	3.10.0-957-21.3.el7.x86_64

From the inital set of trials executed in Phase 1 at the Berlin platform, the measurement results as detailed in the following sections show:

- For any Ultra-Reliable Low-Latency Communication (URLLC) use case
  - all involved components, including the 5G packet core, need to be placed within one premise. The delay imposed by the wide-area, GEANT-based connection between premises does not allow for ultra-low latency communication with delays in the order of 5 ms. This result supports the approach, that edge-base deployments are mandatory for URLLC use cases in order to meet related 5G E2E KPIs. For the Berlin Platform, placing involved components in within a single site / edge-compute location guarantees a possible delay of below 1ms which is more than suitable for URLLC use cases.
  - mmWave 60 GHz backhauling can be used in local deployments to connect nearby vincinities in case required throughput is up to approx. 1 Gbps.
  - The compute power of existing compute nodes is sufficient.

- For non-URLLC use cases,
    - the placement of VNFs / VMs is permissible in any availability zone for the Berlin Platform; the experienced delay is in all cases below 5 ms (10 ms RTT).
    - The compute power of existing compute nodes is sufficient.
    - Existing compute nodes need to be upgraded towards devices having 10GBase-T interfaces.[19]
    - The existing GANT wide-area connection is not suitable for eMBB evaluations in which VNFs/VMs are deployed within availability zones at different sites.

With completing the ongoing acquisition process to updgrade the Berlin Platform to a 100 Gbps-based infrastructure, all anticipated trials can we well executed. A continuous calibration-test-based evaluation of any platform is highly recommended to assure that (a) platform compoentns involved in a given trial are theoretically capable of fulfilling a 5G KPI of a 5G system placed on-top of the testbed infrastructure and (b) E2E measurement results can be thoroughly interpreted by highlighting the potential impact of the behavior of the underlying testbed on the 5G system running on top of it.Experiments and results

These experiments are devoted to test and validate the infrastructure in a first step in order to assess if the envisioned 5G E2E KPIs can be achieved with the platform's infrastructure regardless of deployed components, such as a 5G Core, on-top of the infrastructure.

11.1.1. E2E latency (RTT)

11.1.1.1. Summary and discussion of results for the E2E latency (RTT) calibration tests

The RTT *calibration tests aim at evaluating the* ***influence of the deployed physical components*** *for the compute hosts as well as the characteristics of the platform architecture* ***on the achievable delay***. As such, the calibration tests set the limit towards the achievable delay once a full E2E 5G system is deployed on the platform.

The first set of experiments evaluate the **performance of isolated, individual compute hosts**. The following three figures show the observed mean (average), minimum, and maximum RTT in case OpenStack deployes the source and destination of the round-trip-time measurement on a single compute host (i.e. availability zone). In addition to the average, minimum, and maximum RTT, the *graphs show the precision of the assessment* by sketching 95% confidence values respectively for each node.

For all experiments we observe mean RTTs well below 1 ms (see Figure 32). In general, the mean RTT observed at the compute hosts in zones FOKUS01, FOKUS02, and FOKUS03 is slightly higher as compared to that in zones IHP01 and IHP03. This can be explained as the latter two hosts have a significant higher computational power and larger memory as compared to the hosts in the FOKUS zones.

For all hosts, the observed minimum RTT is well below 0.2 ms (see Figure 33).

---

[19] Note: a corresponding procurement process for the Berlin Platform is ongoing as part of WP4.

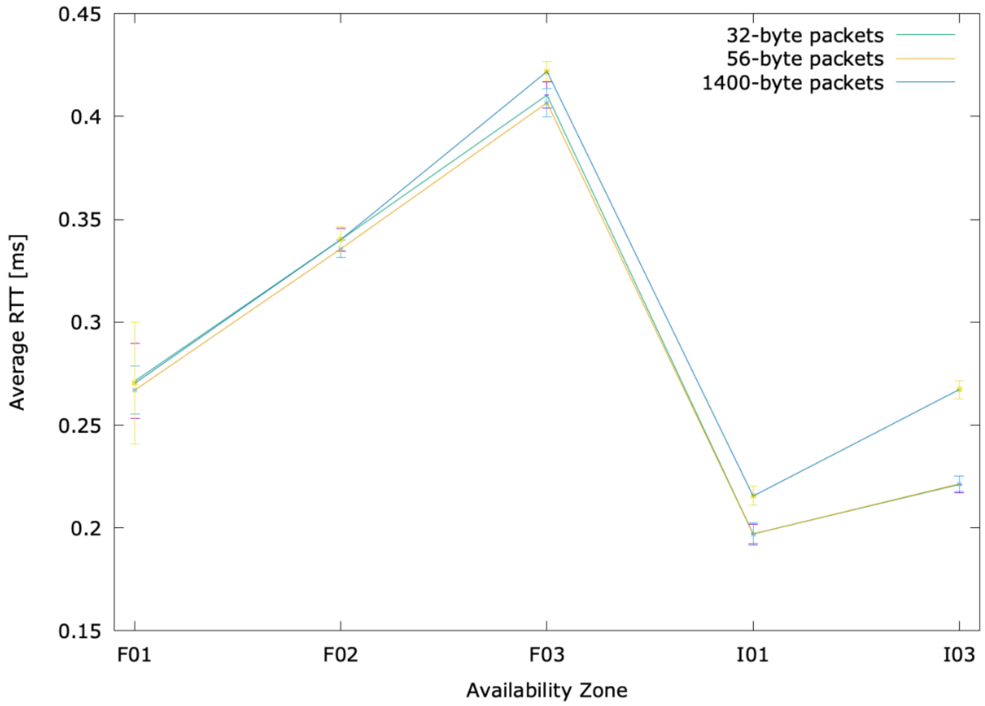


Figure 32 Average Round-Trip-Time between VNFs/VMs deployed in the same availability zone

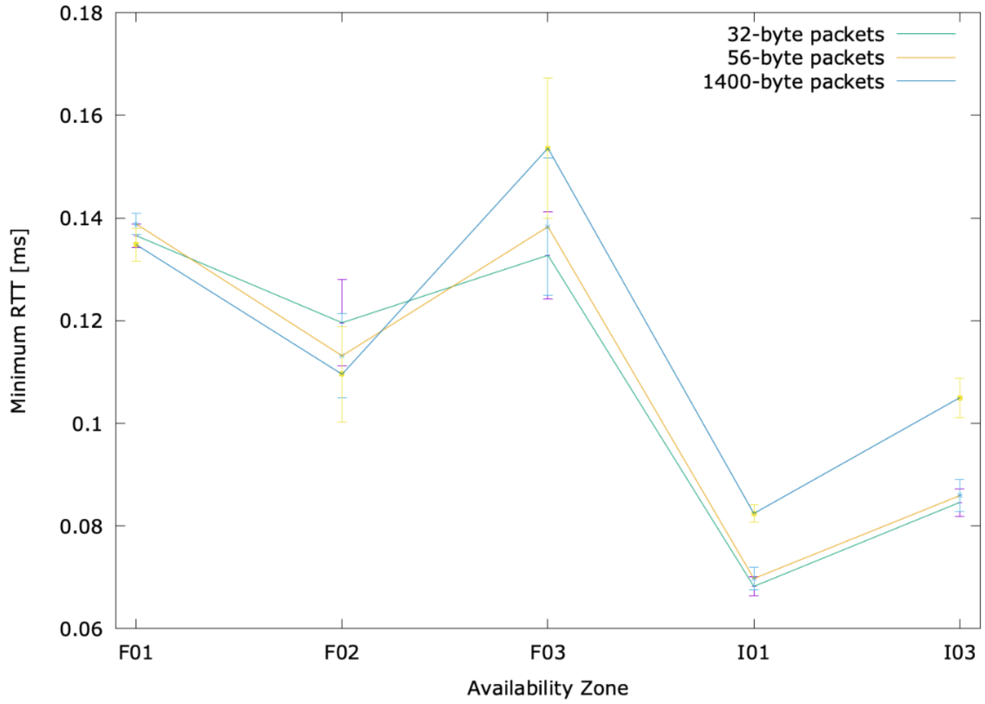


Figure 33 Minimum Round-Trip-Time between VNFs/VMs deployed in the same availability zone

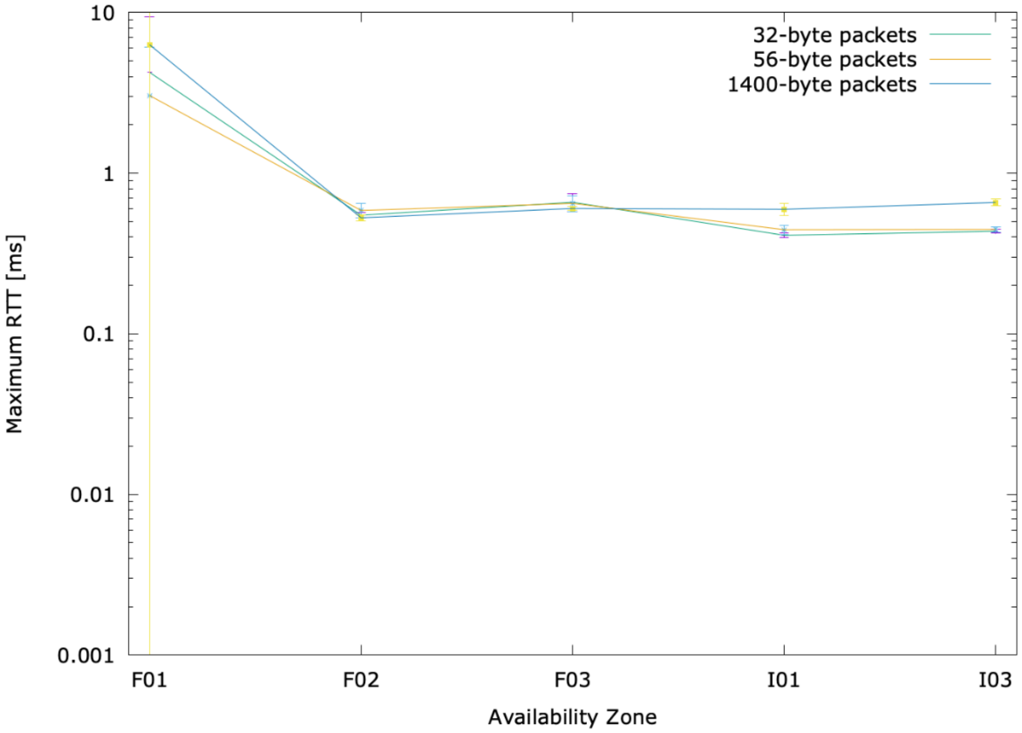


Figure 34 Maximum Round-Trip-Time between VNFs/VMs deployed in the same availability zone

The maximum observed RTT is below 1 ms with the only exception observed at the compute node in the FOKUS01 zone (Figure 34). This can be explained to the fact that in zone FOKUS01, the controlers for Open Baton (orchestration) and OpenStack are in parallel deployed, causing an increased demand towards computational power and memory. This correlation is backed-up by the high variations in the observered maximum RTT between each individual repetition of the experiment causing a limited precision value for the observation for the compute host in zone FOKUS01. Despite the low precision value for the compute host in zone FOKUS01, we observe with 95% confidence that the maximum RTT is guaranteed to be below 10 ms, which corresponds to an approximate 5ms latency as required for 5G E2E KPI evaluations for delay.

The second set of experimnts evaluates the **influence of intermediate switches connecting availability zones** within one premise.[20] Corresponding compute hoste are directly attached to the swith via Ethernet cables, i.e. the performance of the pure, cable-based testbd infrastructure is evaluated. Figure 35 depicts the minimum, maximum, and average experienced RTT. The graph shows as well the *the precision of the assessment* for each statistical value by sketching 95% confidence intervalls respectively.

---

[20] Note that only the performance of the switches at the FOKUS premise could be evaluated in an isolated manner as the IHP premesis did not provide as of now two compute nodes, which are immediately connected to local switches via Ethernet cables.

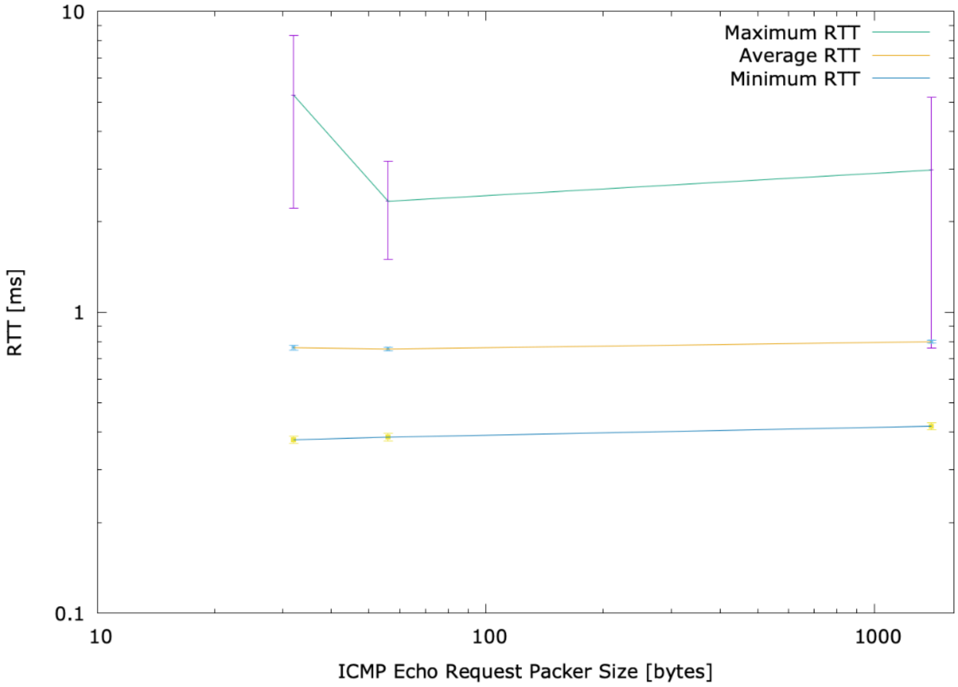


Figure 35 Round-Trip-Time between VNFs/VMs deployed in different availability zones located in the same premise

We observe an average RTT of 0.8 ms (minimum RTT of 0.4 ms) at high precision values (95% confidence intervals are below 1% of the measured values). Also, the influence of the packet size is negletably small being less than 0.04 ms when increasing the packet size from 32 bytes, over 56 bytes, to 1024 bytes. Due to higher variations in the obsereved maximum RTT between independent iterations of the experiments, corresponding precision values are lower (i.e. larger confidence intervals); still, we can state with 95% confidence that regardless of the packet size, the maximum experienced RTT is less than 8 ms.

The third set of experimnts evaluates the **influence of the wide-area, GEANT-based connction between the fascilities of the Berlin platform.** Corresponding compute hoste are directly attached to the swith via Ethernet cables and the switches are directly attachted to the backbone network of each premises which are then interconnected via GEANT. Figure 36 depicts the minimum, maximum, and average experienced RTT. The graph shows as well the *the precision of the assessment* for each statistical value by sketching 95% confidence intervalls respectively.

As anticipated, we observe a significant increase in delay due to the wide-are, GEANT-based interconnection between the sites. The average RTT is at about 14.5 ms (minimum RTT of approx. 14 ms) at high precision values (95% confidence intervals are below 0.5% of the measured values and are so small that they are not visible in Figure 36). Also, the influence of the packet size is negletably small being less than 0.5 ms when increasing the packet size from 32 bytes, over 56 bytes, to 1024 bytes. Due to higher variations in the obsereved maximum RTT between independent iterations of the experiments, corresponding precision values are lower (i.e. larger confidence intervals); still, we can state with 95% confidence that regardless of the packet size, the maximum experienced RTT is less than 17.5 ms.

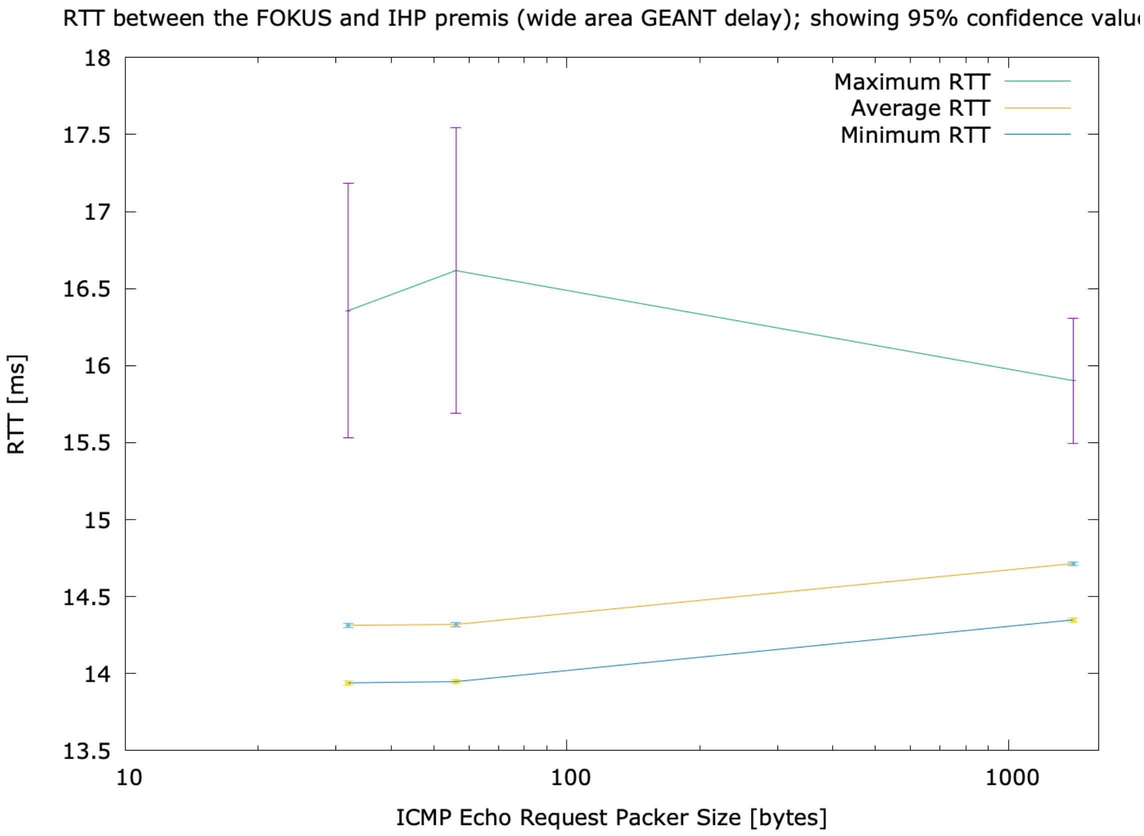


Figure 36 Round-Trip-Time between VNFs/VMs deployed in different availability interconnected via a wide-area, GEANT-based link

In summary, we observe that the placement of measurement endpoints in different availability zones as a significant impact on the suitability of the Berlin Platform on conducting 5G E2E evaluations (c.f. Figure 37). In particular, we conclude:

- For any URLLC use case, all involved components, including the 5G packet core, need to be placed within one premise. The delay imposed by the wide-area, GEANT-based connection between premises does not allow for ultra-low latency communication with delays in the order of 5ms. This result supports the approach, that edge-base deployments are mandatory for URLLC use cases in order to meet related 5G E2E KPIs. For the Berlin Platform, placing involved components in within a single site / edge-compute location guarantees a possible delay of below 1 ms which is more than suitable for URLLC use cases.
- For non-URLLC use cases, the placement of VNFs / VMs is permissible for the Berlin Platform; the experienced delay is in all cases below 5 ms (10 ms RTT).

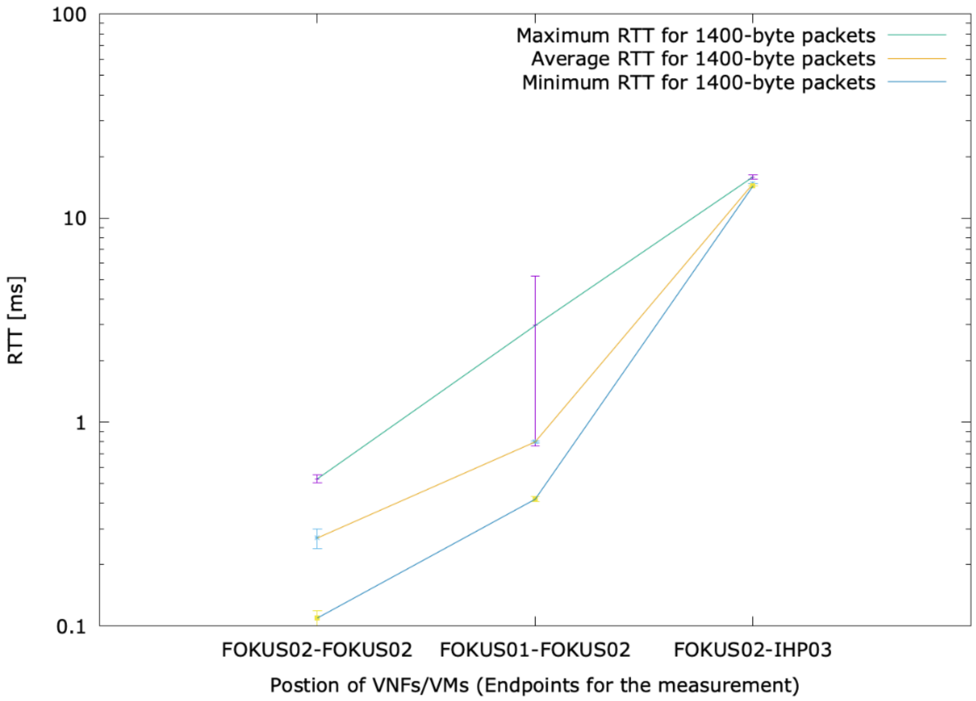


Figure 37 Influence of placement of VNFs/VMs in different availability zones towards the RTT

Further, we conclude from the calibration tests for subsequent experiments in Phases 2 and 3:

- Components controlling the platform infrastructure, such as the orchestration (Open Baton) and the OpenStack controller should run on a dedicated compute host which does not act as an endpoint for a measurement and which is not used to run virtual instruments involved in the experiments.
- Core 5G network functionalities, such as the 5G packet core should run on a dedicated compute host.
- Calibration tests should be repeated for those dedicated components under load once the final infrastructure is delivered to the Berlin platform. I.e., calibration tests on a single compute hosts should be run in parallel to the running management components as well as for a fully deployed 5G packet core to guarantee that deployed (physical) components to not impose any performance bottle necks for the final 5G E2E experiments.
- Especially the performance of any intermediate network component – including compute hosts, switches, and the inter-site wide-area connectivity via GEANT – has to be evaluated in an isolated manner before conducting experiments on top of the infrastructure to assess the 5G E2E KPIs.
- Experiments should include monitoring data of involved components highlighting, e.g., memory usage and CPU load, as recorded for all involved components.

Despite the limited performance of the compute hosts available in Phase 1, results for the delay show that existing hosts currently do not impose a limitation towards the 5G KPI assessment conducted in Phase 1.

11.1.1.2. E2E latency (RTT) between two VNFs running on the same compute node

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>24.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS01 availability zone.</td></tr>
<tr><td>4Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in FOKUS01 is a FOKUS-01</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Latency (RTT)</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>0.25307100</td><td>0.27144000</td><td>0.28980900</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.13424000</td><td>0.13656000</td><td>0.13888000</td><td>32</td></tr>
<tr><td>Maximum</td><td>0.0</td><td>4.23668000</td><td>9.42618000</td><td>32</td></tr>
<tr><td>Mean</td><td>0.25537800</td><td>0.26700000</td><td>0.27862200</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.13677400</td><td>0.13884000</td><td>0.14090600</td><td>56</td></tr>
<tr><td>Maximum</td><td>0.0</td><td>3.03860000</td><td>6.09245200</td><td>56</td></tr>
<tr><td>Mean</td><td>0.24076000</td><td>0.27036000</td><td>0.29996000</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.13160000</td><td>0.13480000</td><td>0.13800000</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.0</td><td>6.29204000</td><td>14.99054000</td><td>1400</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>26.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS02 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in FOKUS02 is a FOKUS-02</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Latency (RTT)</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>0.33461500</td><td>0.34020000</td><td>0.34578500</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.111158</td><td>0.1196</td><td>0.128042</td><td>32</td></tr>
<tr><td>Maximum</td><td>0.524011</td><td>0.54696</td><td>0.569909</td><td>32</td></tr>
<tr><td>Mean</td><td>0.331336</td><td>0.33572</td><td>0.340104</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.104914</td><td>0.11316</td><td>0.121406</td><td>56</td></tr>
<tr><td>Maximum</td><td>0.526087</td><td>0.5852</td><td>0.644313</td><td>56</td></tr>
<tr><td>Mean</td><td>0.334242</td><td>0.3402</td><td>0.346158</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.100291</td><td>0.10956</td><td>0.118829</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.503944</td><td>0.52688</td><td>0.549816</td><td>1400</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>27.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS03 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in FOKUS03 is a FOKUS-03</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Latency (RTT)*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>0.404089</td><td>0.41044</td><td>0.416791</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.124254</td><td>0.13276</td><td>0.141266</td><td>32</td></tr>
<tr><td>Maximum</td><td>0.571349</td><td>0.65812</td><td>0.744891</td><td>32</td></tr>
<tr><td>Mean</td><td>0.399671</td><td>0.40668</td><td>0.413689</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.124953</td><td>0.13832</td><td>0.151687</td><td>56</td></tr>
<tr><td>Maximum</td><td>0.573772</td><td>0.64776</td><td>0.721748</td><td>56</td></tr>
<tr><td>Mean</td><td>0.417054</td><td>0.42192</td><td>0.426786</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.139976</td><td>0.1536</td><td>0.167224</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.584339</td><td>0.60204</td><td>0.619741</td><td>1400</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>02.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the IHP01 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in IHP01 is a IHP-01</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Latency (RTT)*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>0.19218</td><td>0.19708</td><td>0.20198</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.066424</td><td>0.06824</td><td>0.070056</td><td>32</td></tr>
<tr><td>Maximum</td><td>0.398017</td><td>0.41012</td><td>0.422223</td><td>32</td></tr>
<tr><td>Mean</td><td>0.191552</td><td>0.19704</td><td>0.202528</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.067498</td><td>0.06976</td><td>0.072022</td><td>56</td></tr>
<tr><td>Maximum</td><td>0.415997</td><td>0.44428</td><td>0.472563</td><td>56</td></tr>
<tr><td>Mean</td><td>0.210859</td><td>0.21548</td><td>0.220101</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.08075</td><td>0.08244</td><td>0.08413</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.543076</td><td>0.59496</td><td>0.646844</td><td>1400</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>02.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the IHP03 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in IHP03 is a IHP-03</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Latency (RTT)*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>0.217166</td><td>0.22112</td><td>0.225074</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.081885</td><td>0.0846</td><td>0.087315</td><td>32</td></tr>
<tr><td>Maximum</td><td>0.423225</td><td>0.4348</td><td>0.446375</td><td>32</td></tr>
<tr><td>Mean</td><td>0.217634</td><td>0.22144</td><td>0.225246</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.082811</td><td>0.08592</td><td>0.089029</td><td>56</td></tr>
<tr><td>Maximum</td><td>0.428728</td><td>0.44604</td><td>0.463352</td><td>56</td></tr>
<tr><td>Mean</td><td>0.26288</td><td>0.26728</td><td>0.27168</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.101185</td><td>0.105</td><td>0.108815</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.624165</td><td>0.65632</td><td>0.688475</td><td>1400</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.1.3. E2E latency (RTT) between two VNFs located in different availability zones interconnected via a single physical switch

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Rtt-001, TC-Rtt-002, TC-Rtt-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on different compute nodes within the same network segment.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the local network and the performance of the virtualization layer (SDN) as the entire communication resides within a single network segment.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>20.06.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are within the FOKUS01 and FOKUS02 availability zones</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute nodes used in FOKUS01 and FOKUS02 are both of type FOKUS- 01</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Latency (RTT)</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip time
<table>
<tr><td colspan="4">RTT [ms]</td><td rowspan="2">ICMP Echo Request Packet Size [byte]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>0.75004100</td><td>0.76352000</td><td>0.77699900</td><td>32</td></tr>
<tr><td>Minimum</td><td>0.36664800</td><td>0.37748000</td><td>0.38831200</td><td>32</td></tr>
<tr><td>Maximum</td><td>2.21574100</td><td>5.27128000</td><td>8.32681900</td><td>32</td></tr>
<tr><td>Mean</td><td>0.74589000</td><td>0.75560000</td><td>0.76531000</td><td>56</td></tr>
<tr><td>Minimum</td><td>0.37438000</td><td>0.38492000</td><td>0.39546000</td><td>56</td></tr>
<tr><td>Maximum</td><td>1.49798100</td><td>2.34112000</td><td>3.18425900</td><td>56</td></tr>
<tr><td>Mean</td><td>0.78905700</td><td>0.79944000</td><td>0.80982300</td><td>1400</td></tr>
<tr><td>Minimum</td><td>0.40667300</td><td>0.41872000</td><td>0.43076700</td><td>1400</td></tr>
<tr><td>Maximum</td><td>0.76115800</td><td>2.97176000</td><td>5.18236200</td><td>1400</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.1.4. E2E latency (RTT) of the wide area inter-data-center

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Rtt-001, TC-Rtt-002, TC-Rtt-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on compute nodes in different data centers.</td></tr>
<tr><td>Purpose</td><td colspan="4">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).</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>05.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are within the FOKUS02 and IHP03 availability zones</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in FOKUS02 is of type FOKUS- 01 and the compute node used in IHP03 is of type IHP-I03</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Latency (RTT)</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip time
<table>
<tr><th colspan="4">RTT [ms]</th><th rowspan="2">ICMP Echo Request Packet Size [byte]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>14.30094800</td><td>14.31344000</td><td>14.32593200</td><td>32</td></tr>
<tr><td>Minimum</td><td>13.92721100</td><td>13.93952000</td><td>13.95182900</td><td>32</td></tr>
<tr><td>Maximum</td><td>15.52849000</td><td>16.35680000</td><td>17.18511000</td><td>32</td></tr>
<tr><td>Mean</td><td>14.30596700</td><td>14.31904000</td><td>14.33211300</td><td>56</td></tr>
<tr><td>Minimum</td><td>13.93495600</td><td>13.94788000</td><td>13.96080400</td><td>56</td></tr>
<tr><td>Maximum</td><td>15.69014100</td><td>16.61644000</td><td>17.54273900</td><td>56</td></tr>
<tr><td>Mean</td><td>14.70318900</td><td>14.71456000</td><td>14.72593100</td><td>1400</td></tr>
<tr><td>Minimum</td><td>14.33809700</td><td>14.34888000</td><td>14.35966300</td><td>1400</td></tr>
<tr><td>Maximum</td><td>15.49571400</td><td>15.90212000</td><td>16.30852600</td><td>1400</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.2. Throughput KPIs

11.1.2.1. Summary and discussion of results for the Throughput calibration tests

The throughput *calibration tests aim at evaluating the* ***influence of the deployed physical components*** *for the compute hosts as well as the characteristics of the platform architecture* ***on the achievable throughput***. As such, the calibration tests set the limit towards the achievable throughput once a full E2E 5G system is deployed on the platform.

The first set of experiments evaluate the **performance of isolated, individual compute hosts**. . shows the observed mean (average) throughput in case OpenStack deploys the source and destination of the throughput measurement on a single compute host (i.e. availability zone). In addition to the average throughput, the *graphs show the precision of the assessment* by sketching 95% confidence values respectively for each node.

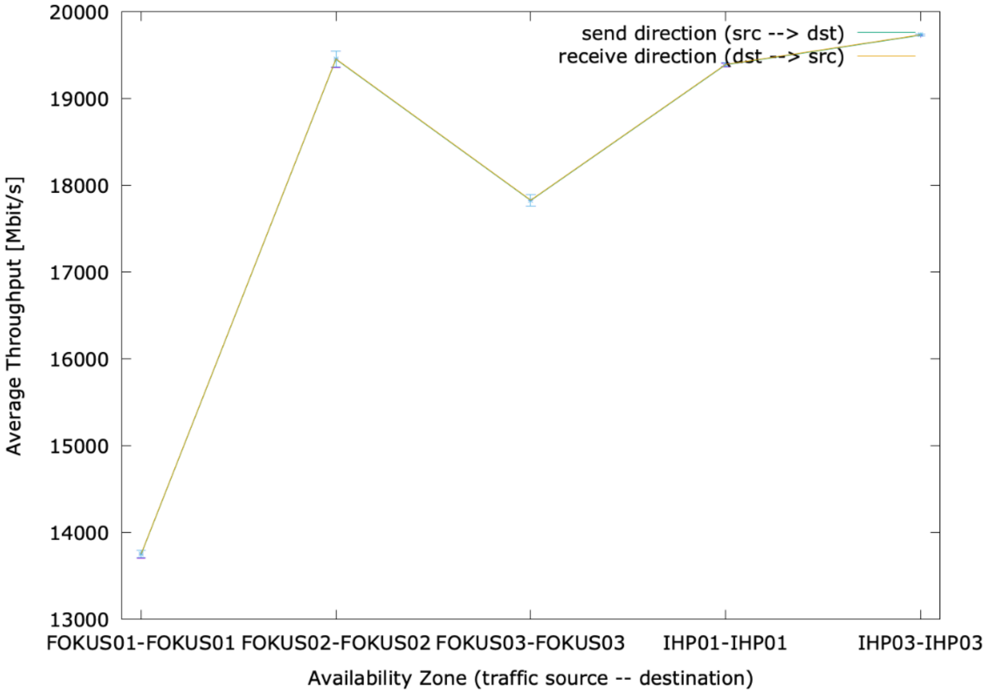


Figure 38 Average Throughput between VNFs/VMs deployed in the same availability zone

For all experiments, the observed average throughput is well above 13.5 Gbps. The precision of the result is high; the 95% confidence value is below 1% of th measured average throughput and as such, the confidence intervals are almost not visible in Figure 38.

The throughput, especially on the compute hosts deployed in the FOKUS availability zones with their limited compute power, we observe a significant variation in the measurement results, which is – similar to the variation in RTTs observed previously – caused by the OpenStack controller and orchestration tool running on the FOKUS01 compute host. Even though the compute hosts deployed in the IHP01 and IHP03 availability zone are significantly more powerful as compared to the ones deployed at the FOKUS' availability zones, we observe for both at most an average throughput of 19.7 Gbps. This opens the question if that upper limit is given by the installed OpenStack deployment or if it can be further increased using more powerful compute nodes. Despite this open item to investigate, the existing equipment seems – regarding its compute power – to be well cabable of evaluating 5G KPIs, which aim at demonstrating peak data rates between 1 and 10 Gbps for specific 5GENESIS use cases. As well the anticipated value of demonstrating 10 Gbps per RRH in the access domain is feasible with the deployed compute power.

The second set of experimnts evaluates the **influence of intermediate switches connecting availability zones** within one premise.[21] Corresponding compute hoste are directly attached to the swicth via Ethernet cables, i.e. the performance of the pure, cable-based testbd infrastructure is evaluated. We observe that the intermediate switch deployed at the FOKUS site limites the average throughput to approximately 769.88 Mbps (send direction, and 796.90 Mbps in the receive direction) at a very high precision value (95% confidence values below 1%). This throughput is well below the achivable throughput of 1 Gbps as advertised by the manufacturer of the deployed switch and the acievalbe capacity of the network card included in the compute host. This result emphasizes the importance of running calibration tests on the deployed network infrastructure before conduncting performance evaluations of 5G E2E KPIs on a testbed's network infrastructure. As the conducted testcase can be re-run in a fully automatized manner once the procurement process of the new switching infrastructure for the Berlin platform is completed, such evaluation will be repeated in the following trial phases to assure the suitablily of the final hardware.

The third set of experimnts evaluates the **influence of the wide-area, GEANT-based connection between the facilities of the Berlin platform.** Corresponding compute hoste are directly attached to the switch via Ethernet cables and the switches are directly attachted to the backbone network of each premises which are then interconnected via GEANT. The inter-site connection provides a throughput of 108.29 Mbps (send direction; 108.31 for the receive direction); again, the precision of the measurement is at with 95% confidence values being below 1%).

In summary, we observe that the placement of measurement endpoints in different availability zones has a significant impact on the suitability of the Berlin Platform on conducting 5G E2E evaluations that involve throughput measurements and evaluations of 5G capacity. In particular, we conclude:

- The compute power of existing compute nodes is sufficient for evaluating enhanced Mobile Broadband (eMBB) use-cases.
- Existing compute nodes need to be upgraded towards devices having 10GBase-T interfaces.[22]
- The switching infrastructure of the platform's backbone network should support preferably 100 Gbps to simultaneously support several 10 Gbps streams originating over several RRHs.
- The existing GEANT wide-area connection is not suitable for eMBB evaluations in which VNFs/VMs are deployed within availability zones at different sites. For eMBB evaluations, all 5G network components – including the 5G packet core – should be deployed in immediate (physical and geographical) proximity of all gNBs involved in a test case.

Further, we conclude from the calibration tests for subsequent experiments in Phases 2 and 3:

---

[21] Note that only the performance of the switches at the FOKUS premise could be evaluated in an isolated manner as the IHP premesis did not provide as of now two compute nodes, which are immediately connected to local switches via Ethernet cables.

[22] Note: a corresponding procurement process for the Berlin Platform is ongoing as part of WP4.

- Calibration tests have to be (re-)run for all subsequent tests in order to assess the impact of the underlying testbed infrastructure on the achievable 5G E2E KPI; this is especially the case for eMBB-related test cases (e.g. assessing the capacity KPI).
- Fully automated calibration tests are essential to cope with the complexity involved in assessing the underlying network infrastructure before conducting any 5G KPI evaluations on top of it. For that, existing TAP-based test plans should be improved to not only produce the target numbers for a specific metric, but also to include a “pass / fail criteria” in order to simply test the infrastructure for its suitability (e.g. passing the criteria “average throughput > 10 Gbps).

11.1.2.2. E2E throughput between two VNFs running on the same compute node

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>07.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS01 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in FOKUS01 is a FOKUS-01</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>13706.57793700</td><td>13751.54600000</td><td>13796.51406300</td><td>send</td></tr>
<tr><td>13709.72359800</td><td>13754.69000000</td><td>13799.65640200</td><td>receive</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>09.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS02 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in FOKUS02 is a FOKUS-01</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>19360.015805</td><td>19453.4276</td><td>19546.839395</td><td>send</td></tr>
<tr><td>19364.356581</td><td>19457.7972</td><td>19551.237819</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>14.07.2019</td></tr>
</table>

<table>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the FOKUS03 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in FOKUS03 is a FOKUS-03</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Throughput</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Throughput
<table>
<tr><th colspan="3">Average Throughput [Mbps]</th><th>Traffic direction</th></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>17762.033209</td><td>17826.986</td><td>17891.938791</td><td>send</td></tr>
<tr><td>17766.210444</td><td>17831.2004</td><td>17896.190356</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>15.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the IHP01 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in IHP01 is a IHP-01</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Throughput</i></td></tr>
</table>

<table>
<tr><td>Additional tools involved</td><td>TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Throughput
<table>
<tr><th colspan="3">Average Throughput [Mbps]</th><th>Traffic direction</th></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>19369.578584</td><td>19389.702</td><td>19409.825416</td><td>send</td></tr>
<tr><td>19374.116891</td><td>19394.2444</td><td>19414.371909</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on a single compute node in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the compute node and the performance of the virtualization layer (SDN) as the entire communication resides within a single compute node, i.e. never leaves the node and hence never transverses a physical network interface, nor intermediate physical switches.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>09.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are both placed within the IHP03 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used in IHP03 is a IHP-03</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Throughput
<table>
<tr><th colspan="3">Average Throughput [Mbps]</th><th>Traffic direction</th></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>19718.438803</td><td>19730.6344</td><td>19742.829997</td><td>send</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td><table><tr><td>19722.957972</td><td>19735.1556</td><td>19747.353228</td><td>receive</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.2.3. E2E throughput between two VNFs located in different availability zones interconnected via a single physical switch

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on the same network segment, i.e. two compute nodes connected by a single physical switch.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the local network connection through the physical switch and the performance of the virtualization layer (SDN).</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>08.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are placed within the availability zones FOKUS01 and FOKUS02.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute nodes used in FOKUS01 and FOKUS02 are both of type FOKUS-COMPUTE-NODE-F01</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Throughput</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Throughput<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>796.72562600</td><td>796.88074000</td><td>797.03585400</td><td>send</td></tr>
<tr><td>796.90536200</td><td>797.06198000</td><td>797.21859800</td><td>receive</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.2.4. E2E latency (RTT) of the wide area inter-data-center

Test Case ID	TC-Thr-001
General description of the test	The test assesses the average Throughput between two communication end-points (VNF) deployed in different data centers.

<table>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the inter-data-center network connection and the performance of the virtualization layer (SDN).</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>09.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs) are within the FOKUS02 and IHP03 availability zones</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in FOKUS02 is of type FOKUS-COMPUTE-NODE-F01 and the compute node used in IHP03 is of type IHP-COMPUTE-NODE-I03</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Throughput</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>107.92475800</td><td>108.28680000</td><td>108.64884200</td><td>send</td></tr>
<tr><td>107.94972500</td><td>108.31180000</td><td>108.67387500</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.3. Service creation time calibration test

11.1.3.1. Summary and discussion of results of the virtualizd packet core creation time

This initial evaluation aims at assessing the service creation time required to successfully deploy an Ubuntu-based Unix system (without running any additional services) in the testbed. The experiment assesses the baseline performance of the compute host used to deploy the service at as well as the influence of in which availability zone the compute host is located.

Deploying the VM at the compute host where the OpenStack and Open Baton controller resides on results in the lowest average service creation time of 69 s. Therein, the VM does not need to be deployed via any intermediate network components; all traffic is local to the host which naturally causes the lowest service creation time. When deploying to a different availability zone at the same site via wired Ethernet connections (here FOKUS02), the service creation time almost doubles to 134 s. For the former two results, precision is very high: 95% confidence intervals are ± 2 s, i.e. below 2%.

When placing the service at compute hosts connected via wireless 60 GHz backbone links and across the wide-area, GEANT-based inter-site connection, creation times further increase at a simultaneous decrease in precision of the observation. A graphical representation of these results is shown in Figure 39.

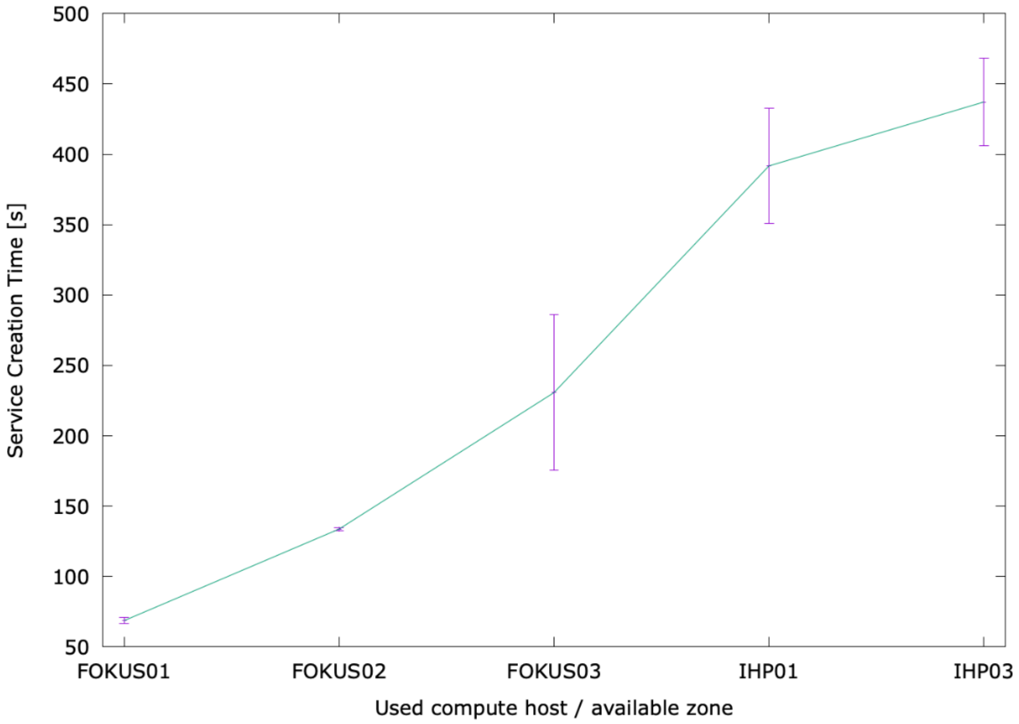


Figure 39 Service creation time calibration test – influence of VM placement

We **conclude** that the **underlying network infrastructure** has a **major influence on the service creation time**. It is anticipatd that adding an additional layer, i.e. a full E2E 5G system including a packet core, add an additional, potentially constant delay. These initial calibration results show the importance of re-running calibration test whenever network components and compute hosts involved in an experiment change in order to differentiate the effects of the testbed infrastructure vs. the 5G system on the observations.

11.1.3.2. Service Creation Time calibration test

Test Case ID	TC-Ser-001			
General description of the test	The test assesses the average service creation time of a generic Ubuntu 16.04 VM on a compute host.			
Purpose	The test acts as a calibration test to assess the performance of the compute host used for deployment and the performance of the underlying network components.			
Executed by	Partner:	FOKUS	Date:	17.07.2019
Involved Partner(s)	FOKUS, IHP, KAU			
Scenario	The scenario deploys a running Unix host – without any additional services – on top of an existing OpenStack deployment. The test assesses all overhead involved, from triggering and coordinating the deployment via Open Baton, downloading the image of the system to the compute host, and starting at the latter the service, i.e. virtualized machine running Unix.			
Slicing configuration	An exclusive slice / software-defined network is used for the test.			

<table>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>The compute node used for deploying the service is of type FOKUS-01.
OpenStack controllers are run on FOKUS01.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td>Service Creation Time</td></tr>
<tr><td>Additional tools involved</td><td>TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Service Creation Time
<table>
<tr><td colspan="4">Service Creation Time [s]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>66.39</td><td>68.66</td><td>70.67</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a generic Ubuntu 16.04 VM on a compute host.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to assess the performance of the compute host used for deployment and the performance of the underlying network components.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>16.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The scenario deploys a running Unix host – without any additional services – on top of an existing OpenStack deployment. The test assesses all overhead involved, from triggering and coordinating the deployment via OpenBaton, downloading the image of the system to the compute host, and starting at the latter the service, i.e. virtualized machine running Unix.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used for deploying the service is of type FOKUS-01
OpenStack controllers are run on FOKUS01.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Service Creation Time</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results</td><td colspan="4">Service Creation Time
<table>
<tr><td>Service Creation Time [s]</td></tr>
</table></td></tr>
</table>

<table>
<tr><td><i>(those included in the test case definition)</i></td><td><table><tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr><tr><td>Mean</td><td>132.68</td><td>133.69</td><td>134.70</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a generic Ubuntu 16.04 VM on a compute host.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to assess the performance of the compute host used for deployment and the performance of the underlying network components.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>17.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The scenario deploys a running Unix host – without any additional services – on top of an existing OpenStack deployment. The test assesses all overhead involved, from triggering and coordinating the deployment via Open Baton, downloading the image of the system to the compute host, and starting at the latter the service, i.e. virtualized machine running Unix.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used for deploying the service is of type FOKUS-03.
OpenStack controllers are run on FOKUS01.</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Service Creation Time</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Service Creation Time<table><tr><td colspan="4">Service Creation Time [s]</td></tr><tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr><tr><td>Mean</td><td>175.49</td><td>230.68</td><td>285.87</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td>TC-Ser-001</td></tr>
</table>

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a generic Ubuntu 16.04 VM on a compute host.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to assess the performance of the compute host used for deployment and the performance of the underlying network components.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>17.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The scenario deploys a running Unix host – without any additional services – on top of an existing OpenStack deployment. The test assesses all overhead involved, from triggering and coordinating the deployment via OpenBaton, downloading the image of the system to the compute host, and starting at the latter the service, i.e. virtualized machine running Unix.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used for deploying the service is of type IHP-01.
OpenStack controllers are run on FOKUS01.</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Service Creation Time</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Service Creation Time
<table>
<tr><th colspan="4">Service Creation Time [s]</th></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>350.95</td><td>391.72</td><td>432.49</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a generic Ubuntu 16.04 VM on a compute host.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to assess the performance of the compute host used for deployment and the performance of the underlying network components.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>17.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The scenario deploys a running Unix host – without any additional services – on top of an existing OpenStack deployment. The test assesses all overhead involved, from triggering and coordinating the deploy-</td></tr>
</table>

<table>
<tr><td></td><td>ment via OpenBaton, downloading the image of the system to the compute host, and starting at the latter the service, i.e. virtualized machine running Unix.</td></tr>
<tr><td>Slicing configuration</td><td>An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td>The compute node used for deploying the service is of type IHP-01.
OpenStack controllers are run on FOKUS01.</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td><i>Service Creation Time</i></td></tr>
<tr><td>Additional tools involved</td><td>TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td>Service Creation Time
<table>
<tr><th colspan="4">Service Creation Time [s]</th></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>406.12</td><td>437.14</td><td>468.17</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.4. Evaluation of mmWave-based Backhaul for 5G networks

11.1.4.1. Summary and discussion of results for mmWave backhauling

The first set of **experiments related to mmWave-based backhauling**, conducted in Phase 1, **aim at evaluating the achievable delay (RTT) and throughput** of the SUT. Evaluations involve two 60 GHz-based systems deployed at the two sites of the Berlin Platform, i.e.:

- MetroLinq 60 GHz system[23], consisting of:
  - MetroLinq 60-LW, 60GHz + 5GHz + 2.4 GHz PTP/PTMP client device.
  - MetroLinqTM 10G Tri-Band Omni, 3x120° 60 GHz, 4x90° 2.4 GHz & 4x90° 5 GHz BS.
- IHP development of a 60 GHz backhaul system.

For the MetroLinq 60 GHz backhaul system, the average (2.6 ms), maximum (126 ms), and minimum (1 ms) **RTTs** are invariant against the ICMP packt size (95% confidence intervals overlap for the respective measurements despite a slight increast in RTT when increasing the packet size) as depicted in Figure 40. While the precision of the observed minimum and maximum RTT is less than ± 0.5 ms, we measure rather rather high variations for the maximum resulting in rather low precision (95% confidence values of ± 20 ms). The observed maximum RTT is in the order of 100 ms, which prevents the MetroLinq system from being suitable for providing backhauling connectivity for URLLC use cases.

---

[23] Note, the 2.5 and 5 GHz links were disabled for the tests.

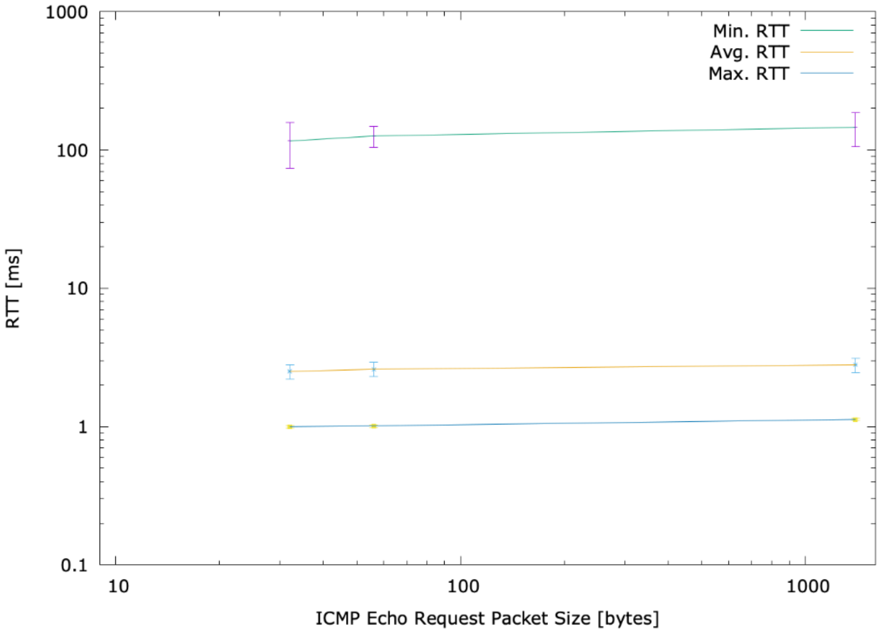


Figure 40 Round-Trip-Time observed for the MetroLinq 60 GHz backhaul

For the IHP-Protoype 60 GHz Backhaul system, we also observe a slight increase for the minimum, average, and maximum RTT when increasing the ICMP Echo Requst packet size. In contrast to the MetroLinq system, precision values for the observations is so good that the increase of the RTT for 1024-byte packts is statistically significant, though still below 1 ms total. Observed average, minimum, and maximum RTTs are 2.4 ms, 0.4 ms, and 4.7 ms respectively.

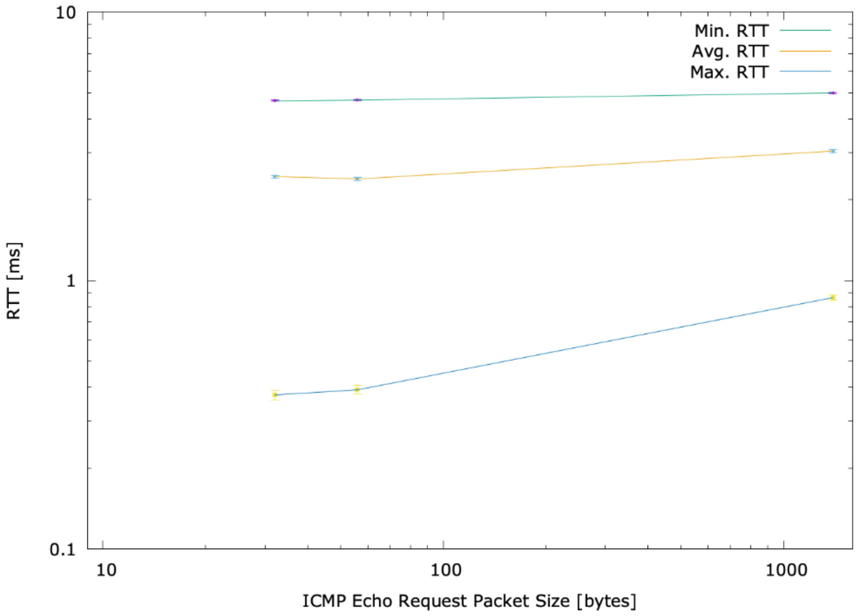


Figure 41 Round-Trip-Time observed for the IHP-Prototype 60 GHz backhaul

Comparing the two backhaul systems, we conclude that the IHP-Prototype system is suitable for URLLC use-cases: even the average RTT (two-way delay) is with 2.4 ms well below the 5 ms one-way delay target for 5G; and the target is also held for the observed maximum RTT.

The measured average **Throughput** varies significantly for each system under test (c.f. Figure 42). While the commercially available MetroLinq system provides 530 Mbps, we observe 870

Mbps throughput for the IHP prototype 60 GHz backhaul system. For both SUTs, up- and downlink (send and receive direction) provide the same throughput. Also, the precision for all reported results is extremely high (95% confidence intervals are below 0.2% of the observation value). As anticipated, the wide-area GEANT-based link limits the throughput when the two 60 GHz systems located at the FOKUS and at the IHP facility are both used in the same E2E connection over GEANT.

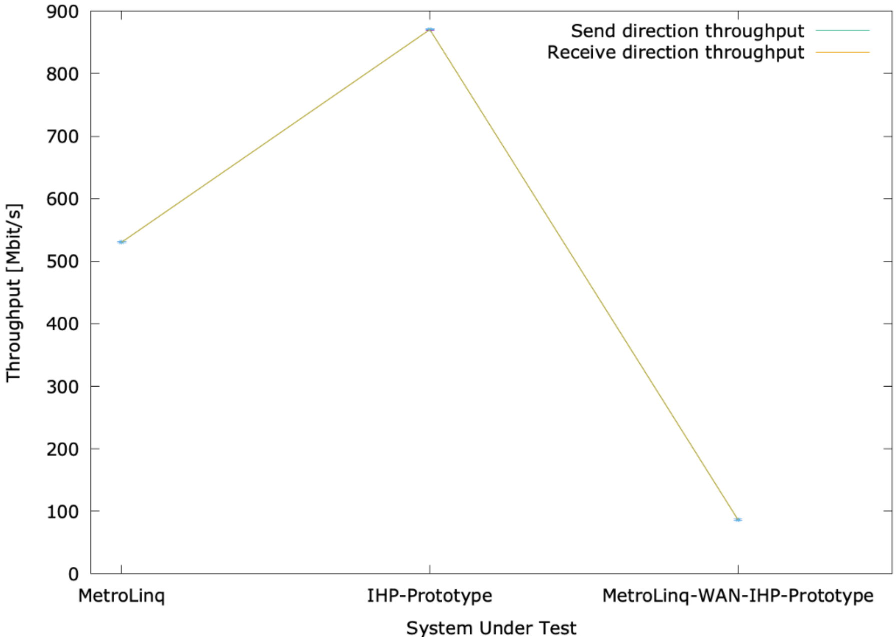
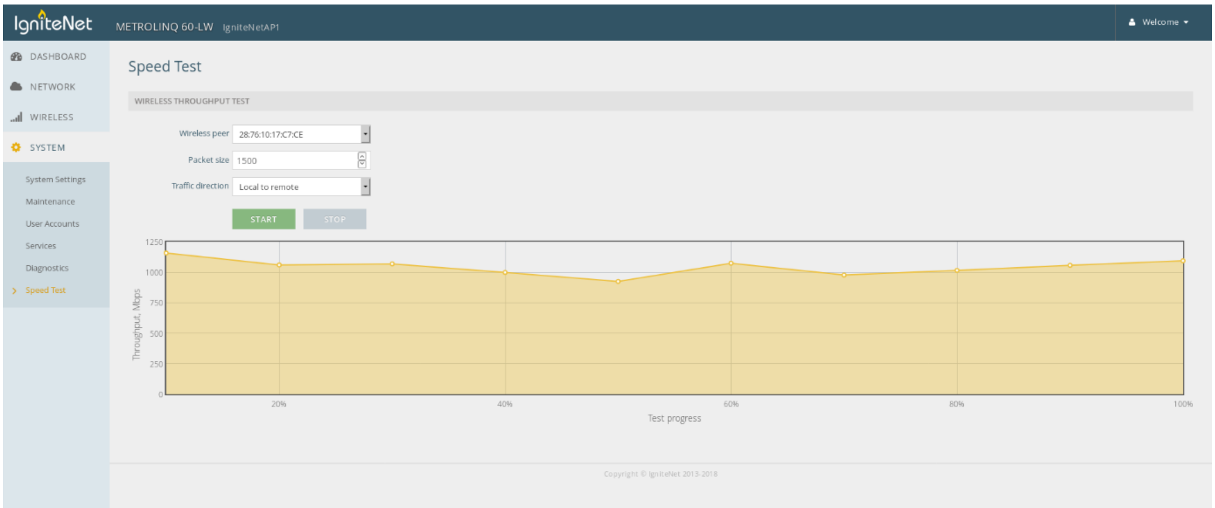


Figure 42 Throughput of the IHP 60 GHz backhaul system

It should be noted that the Metrolinq system is attached to a 1 Gbps Ethernet switch and that the observed PL via the 60 GHz link is rather high. As the switch, as shown by previous calibration throughput tests is well cabable of supporting throughput rates of 770 Mbps at zero packet loss, we conclude that the switch is not a limiting factor in the experiment. Low throughput rates are rather due to the packet losses causing a limitation of the TCP transmission window. As a control measurement, a throughput measurement is directly taken between the MetroLinq BS and client via an internal line speed tool. As such, the effects of any network equipment apart from the 60 GHz system itself is removed from the set-up. The result is shown in the screenshot in Figure 43. The observed throughput of approx. 1.2 Gbps is still below the – according to the used modulation and coding scheme – theoretically achievable throughput of 2.4 Gbps, which supports the assumption of having a rather poor 60 GHz link or misaligned antennas.



Speed Test

WIRELESS THROUGHPUT TEST

Wireless peer 28:76:10:17:C7:CE

Packet size 1500

Traffic direction Local to remote

Figure 43 60 GHz backhaul systems – internal throughput test result

In **summary**, we can conclude that the IHP prototype of a 60 GHz backhaul system is well capable of supporting URLLC use-cases. Despite the limited usability of the MetroLinq system for eMBB and URLLC use cases, it is still a suitable solution for providing management links in a testbed set-up as it is capable of simultaneously using 2.4, 5, and 60 GHz links (the former two being disabled for the previous tests), which increases the reliability of the system.

11.1.4.2. E2E latency (RTT) between two VNFs interconnected via a single 60 GHz backhaul

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003			
General description of the test	The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on both sides of a 60GHz backhaul.			
Purpose	The test acts as a calibration test to primarily assess the performance of the intermediate 60GHz backhaul link.			
Executed by	Partner:	FOKUS	Date:	14.07.2019
Involved Partner(s)	FOKUS, IHP, KAU			
Scenario	The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the FOKUS03 and FOKUS02 availability zone.			
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved *(e.g. HW components, SW components)*	The compute node used in FOKUS03 is a FOKUS-03 The compute node used in FOKUS02 is a FOKUS-01 The connecting 60 GHz backhaul is realized by the MetroLinq nodes.			
Metric(s) under study (*Refer to those in Section 4*)	*Latency (RTT)*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points			
Primary measurement results	Round Trip Time RTT [ms]			

(those included in the test case definition)		Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval
	Mean	2.20049700	2.50252000	2.80454300
	Minimum	0.97464000	0.99812000	1.02160000
	Maximum	73.46218500	116.01004000	158.55789500
Complementary measurement results	n/a			

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003			
General description of the test	The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) deployed on both sides of a 60 GHz backhaul.			
Purpose	The test acts as a calibration test to primarily assess the performance of the intermediate 60 GHz backhaul link.			
Executed by	Partner:	FOKUS	Date:	04.07.2019
Involved Partner(s)	FOKUS, IHP, KAU			
Scenario	The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the IHP01 and IHP03 availability zone.			
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved *(e.g. HW components, SW components)*	The compute node used in IHP01 is a IHP-01. The compute node used in IHP03 is a IHP-01. The connecting 60GHz backhaul is realized by IHP's prototype.			
Metric(s) under study (*Refer to those in Section 4*)	*Latency (RTT)*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points			
Primary measurement results *(those included in the test case definition)*	Round Trip Time			
	RTT [ms]			
		Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval
	Mean	2.412704	2.4398	2.466896
	Minimum	0.359593	0.37516	0.390727
	Maximum	4.643428	4.66688	4.690332
Complementary measurement results	n/a			

11.1.4.3. E2E latency (RTT) between two VNFs interconnected via two mmWave (60 GHz) backhauls and a wide-are inter-site connection

Test Case ID	TC-Rtt-001, TC-Rtt-002, TC-Rtt-003

<table>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points (VNF) in different data centers connected to the network via a 60 GHz backhaul.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the combined 60 GHz backhaul inter-data-center links.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>04.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the FOKUS03 and IHP01 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used in FOKUS03 is a FOKUS-03
The compute node used in IHP01 is a IHP-01</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Latency (RTT)</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip Time
<table>
<tr><th colspan="4">RTT [ms]</th></tr>
<tr><th></th><th>Lower bound of 95% confidence interval</th><th>Value</th><th>Upper bound of 95% confidence interval</th></tr>
<tr><td>Mean</td><td>17.36704200</td><td>17.53880000</td><td>17.71055800</td></tr>
<tr><td>Minimum</td><td>14.75233700</td><td>14.79356000</td><td>14.83478300</td></tr>
<tr><td>Maximum</td><td>75.62825300</td><td>113.47292000</td><td>151.31758700</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

11.1.4.4. E2E Throughput between two VNFs interconnected via a single 60 GHz backhaul

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average Throughput between two communication end-points (VNF) deployed on both sides of a 60GHz backhaul.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the intermediate 60GHz backhaul link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>12.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the FOKUS02 and FOKUS03 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved</td><td colspan="4">The compute node used in FOKUS03 is a FOKUS-03</td></tr>
</table>

(e.g. HW components, SW components)	The compute node used in FOKUS02 is a FOKUS-01 The connecting 60GHz backhaul is realized by the MetroLinq nodes.
Metric(s) under study (*Refer to those in Section 4*)	*Throughput*
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points
Primary measurement results *(those included in the test case definition)*	Throughput
Complementary measurement results	n/a

Throughput

Average Throughput [Mbps]			Traffic direction
Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval	
529.604483	530.335962	531.06744	send
529.706667	530.439192	531.171718	receive

Test Case ID	TC-Thr-001			
General description of the test	The test assesses the average Throughput between two communication end-points (VNF) deployed on both sides of a 60 GHz backhaul.			
Purpose	The test acts as a calibration test to primarily assess the performance of the intermediate 60 GHz backhaul link.			
Executed by	Partner:	FOKUS	Date:	11.07.2019
Involved Partner(s)	FOKUS, IHP, KAU			
Scenario	The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the IHP01 and IHP03 availability zone.			
Slicing configuration	An exclusive slice / software-defined network is used for the test.			
Components involved *(e.g. HW components, SW components)*	The compute node used in IHP01 is a IHP-01. The compute node used in IHP03 is a IHP-01. The connecting 60 GHz backhaul is realized by IHP's prototype.			
Metric(s) under study (*Refer to those in Section 4*)	*Throughput*			
Additional tools involved	TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points			
Primary measurement results *(those included in the test case definition)*	Throughput			

Throughput

Average Throughput [Mbps]			Traffic direction
Lower bound of 95% confidence interval	Value	Upper bound of 95% confidence interval	
869.887106	870.5158	871.144494	send

<table>
<tr><td></td><td><table><tr><td>870.091016</td><td>870.72</td><td>871.348984</td><td>receive</td></tr></table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.4.5. E2E Throughput between two VNFs interconnected via two mmWave (60 GHz) backhauls and a wide-are inter-site connection

<table>
<tr><td>Test Case ID</td><td>TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td>The test assesses the average Throughput between two communication end-points (VNF) on different premises and connected to the inter-site connection via a 60 GHz backhaul.</td></tr>
<tr><td>Purpose</td><td>The test acts as a calibration test to primarily assess the performance of the combined 60 GHz backhaul inter-data-center links.</td></tr>
<tr><td>Executed by</td><td><table><tr><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>16.07.2019</td></tr></table></td></tr>
<tr><td>Involved Partner(s)</td><td>FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td>The communication endpoints of the service (VNFs), i.e. the virtual instruments, are placed within the FOKUS03 and IHP01 availability zone.</td></tr>
<tr><td>Slicing configuration</td><td>An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>The compute node used in IHP01 is a IHP-01
The compute node used in FOKUS03 is a FOKUS-F03
The connecting 60GHz backhauls are realized by IHP’s prototype and the MetroLinq devices on IHP’s and FOKUS’ end respectively.</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td>*Throughput*</td></tr>
<tr><td>Additional tools involved</td><td>TAP for automated testing, Debian-based virtual instruments (VNFs) acting as communication end-points</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Throughput<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>85.251507</td><td>86.36556</td><td>87.479613</td><td>send</td></tr>
<tr><td>85.268584</td><td>86.38296</td><td>87.497336</td><td>receive</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.5. Packet Core (Open5GCore Rel.3) Evaluations

11.1.5.1. Summary and discussion of results of the Open5GCore Rel.3 packet core evaluations

The experiments provide an **initial E2E network layer evaluation of the Open5GCore Rel.3** with respect to achievable **delay** (RTT), **throughput**, and **service creation time.** As 5G NR equipment was not available at the time, an Airspan LTE Femtocell as attached to the core; a notebook with a Huawei 4G USB dongle was used as the UE. Though the access technology is LTE, the Open5GCore implements already an initial split of the data and control path while at the same time supporting the LTE packet core interfaces towards the access network. Also, it should be noted that due to constrains of available spectrum licenses, the bandwidth of the LTE system had to be reduced to 5 MHz.

The **observed RTT** is on average 45 ms (approx. 22 ms one-way latency) with minimums at 20 ms and maximums at 125 ms (c.f. Figure 44). Increasing the packet size yields to increasd RTTs of approx. 20% when going from 32-byte to 1400-byte packets.

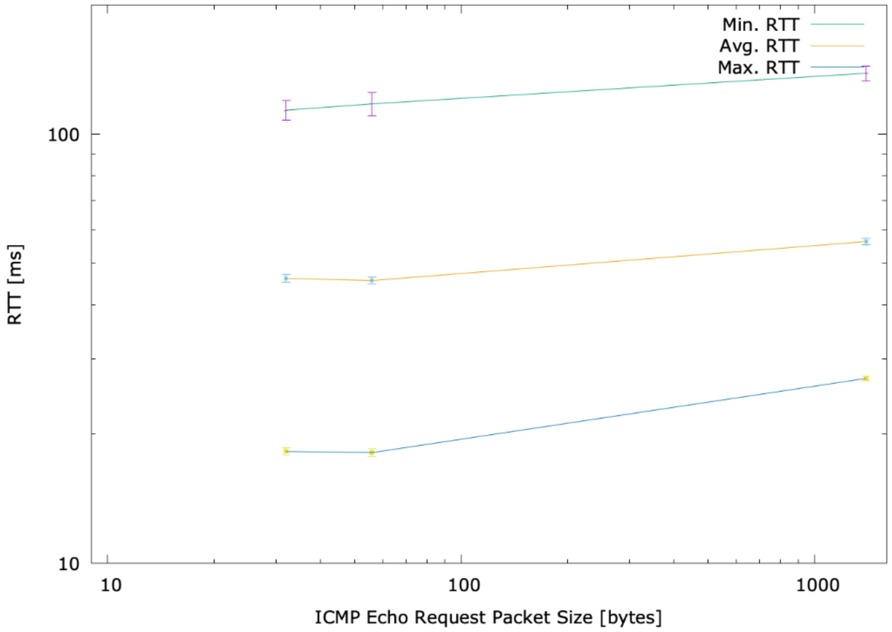


Figure 44 E2E network layer RTT evaluation of the Open5GCore Rel.3 packet core

The observed **throughput** is asymmetrical for the up- and downlink. The rather limited throughput rates of 8 Mbps and 21 Mbps are explained by the very narrow 5 MHz channel available for the test. Precision is excellent for all observations resulting in 95% confidence values of less than 0.1% and as such are not visible in Figure 45.

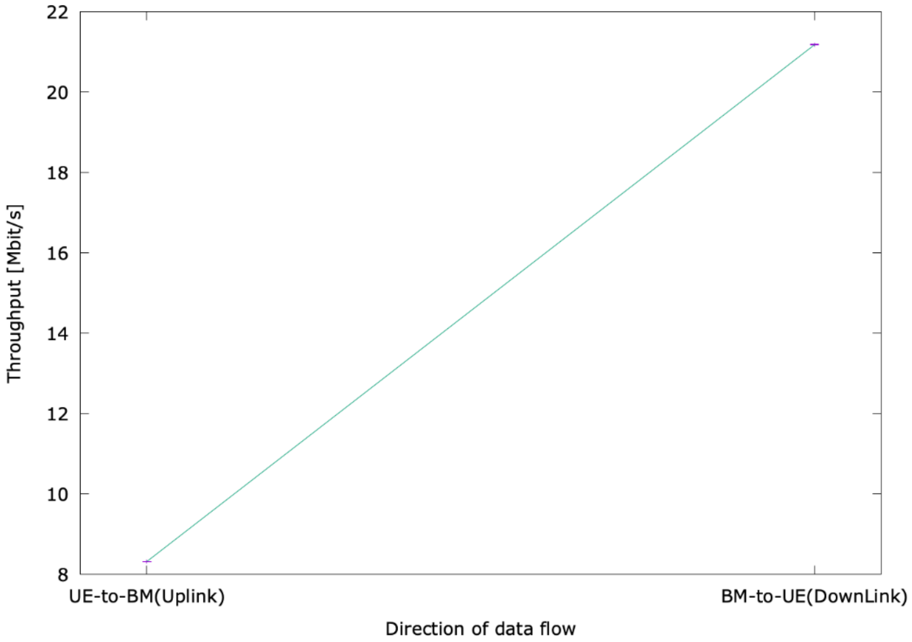


Figure 45 E2E network layer Up- and Down-Link Throughput evaluation of the Open5GCore Rel.3 packet core

Figure 46 shows the influence of the placemnt of the placement of the virtualizd packet core on the associated deployment / **service creation time**, which is in the order of 200 s. The observed service creation times need to be put in relation to the service creation times observed during the calibration tests. While deploying a simple Ubuntu-based VM on FOKUS01 (local deployment, no underlying switching infrastructure needed) requires only 67 s, deploying a full 5G packet core on the same system requires 209 s. Notably, the observed service creation time on FOKUS02 – a host for which required the deployment across the local network infrastructure at the site – are lower than the ones observed at FOKUS01. Also, high variations in observed measurements cause extremely large confidence intervals, especially for FOKUS01. This is likely due to the fact, that FOKUS01 hosts additionally the OpenStack Controler and Open Baton, causing an increased compute load and memory usage as compared to FOKUS02 which exclusively hosts the virtualized packet core in this experiment.

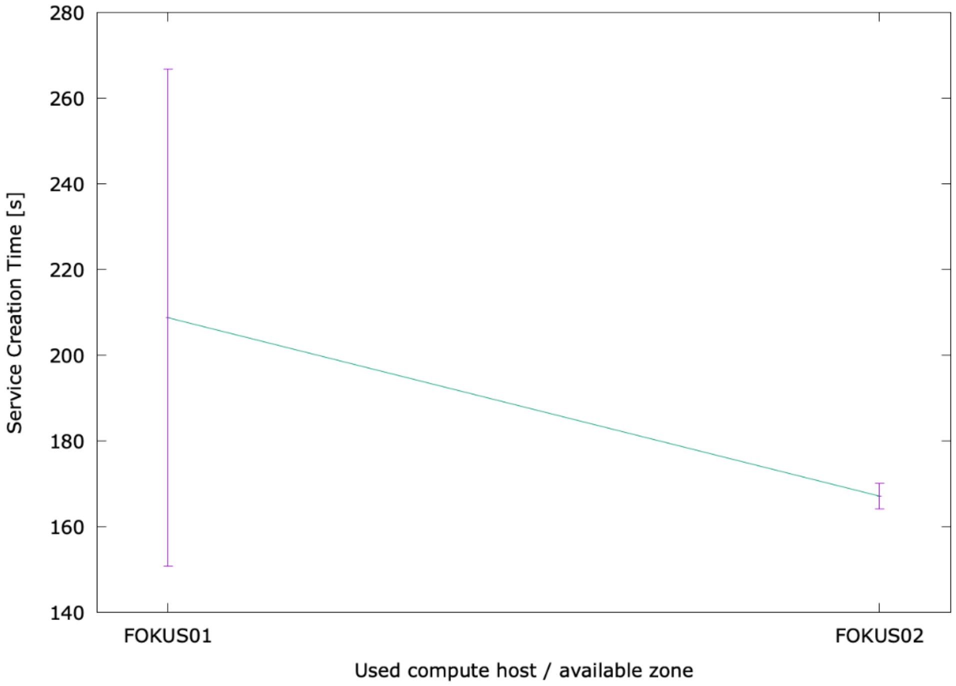


Figure 46 Service creation time of the Open5GCore Rel.3 packet core

11.1.5.2. E2E network layer RTT test (LTE Rel.14 Core and RAN) tests

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Rtt-001, TC-Rtt-002, TC-Rtt-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points, namely the UE and an additional host connected to the core network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>10.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints are the UE and the additional host.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">UE host of type FOKUS-05.
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(*Refer to those in Section 4*)</td><td colspan="4">*Latency (RTT)*</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Round Trip time
<table>
<tr><td colspan="4">RTT [ms]</td><td rowspan="2">ICMP Echo Request Packet Size [byte]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><td>Mean</td><td>45.092822</td><td>46.15144</td><td>47.210058</td><td>32</td></tr>
<tr><td>Minimum</td><td>17.865116</td><td>18.23716</td><td>18.609204</td><td>32</td></tr>
<tr><td>Maximum</td><td>107.668148</td><td>113.82</td><td>119.971852</td><td>32</td></tr>
<tr><td>Mean</td><td>44.801215</td><td>45.61996</td><td>46.438705</td><td>56</td></tr>
<tr><td>Minimum</td><td>17.75253</td><td>18.12644</td><td>18.50035</td><td>56</td></tr>
<tr><td>Maximum</td><td>110.510235</td><td>117.71328</td><td>124.916325</td><td>56</td></tr>
<tr><td>Mean</td><td>55.213249</td><td>56.25152</td><td>57.289791</td><td>1400</td></tr>
<tr><td>Minimum</td><td>26.68384</td><td>26.9914</td><td>27.29896</td><td>1400</td></tr>
<tr><td>Maximum</td><td>133.147867</td><td>138.73984</td><td>144.331813</td><td>1400</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Rtt-001, TC-Rtt-002, TC-Rtt-003</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average, minimum, and maximum RTT between two communication end-points, namely the UE and the host running the Open5GCore Rel.3 in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>11.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints are placed within the UE and the host running the Open5GCore Rel.3.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">UE host of type FOKUS-05
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Latency (RTT)</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Round Trip time
<table>
<tr><td colspan="4">RTT [ms]</td><td rowspan="2">ICMP Echo Request Packet Size [byte]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>44.939519</td><td>45.95244</td><td>46.965361</td><td>32</td></tr>
<tr><td>Minimum</td><td>17.797566</td><td>18.20576</td><td>18.613954</td><td>32</td></tr>
<tr><td>Maximum</td><td>109.585446</td><td>114.80176</td><td>120.018074</td><td>32</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><td>Mean</td><td>45.571879</td><td>46.29628</td><td>47.020681</td><td>56</td></tr>
<tr><td>Minimum</td><td>17.861913</td><td>18.29896</td><td>18.736007</td><td>56</td></tr>
<tr><td>Maximum</td><td>121.940436</td><td>127.77504</td><td>133.609644</td><td>56</td></tr>
<tr><td>Mean</td><td>54.815661</td><td>55.93992</td><td>57.064179</td><td>1400</td></tr>
<tr><td>Minimum</td><td>25.869996</td><td>26.20252</td><td>26.535044</td><td>1400</td></tr>
<tr><td>Maximum</td><td>134.361821</td><td>138.87752</td><td>143.1393219</td><td>1400</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.5.3. E2E network layer Throughput test (LTE Rel.14 Core and RAN)

Downlink

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average throughput between two communication end-points, namely the UE and an additional host connected to the core network</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the downlink performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>10.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints are placed within the UE and the additional host.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">UE host of type FOKUS-05.
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4"></td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>21.176394</td><td>21.18268</td><td>21.188966</td><td>send</td></tr>
<tr><td>21.17937</td><td>21.18568</td><td>21.19199</td><td>receive</td></tr>
</table>
</td></tr>
</table>

Complementary measurement results	n/a

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average throughput between two communication end-points, namely the UE and the host running the 5GCore in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the downlink performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>11.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints are placed within the UE and the host running the 5GCore.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">UE host of type FOKUS-05.
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4"></td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>21.182999</td><td>21.18828</td><td>21.193561</td><td>send</td></tr>
<tr><td>21.19096</td><td>21.19644</td><td>21.20192</td><td>receive</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

Uplink

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average throughput between two communication end-points, namely the UE and an additional host connected to the core network</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the uplink performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>10.07.2019</td></tr>
</table>

<table>
<tr><td>Involved Partner(s)</td><td>FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td>The communication endpoints are placed within the UE and the additional host.</td></tr>
<tr><td>Slicing configuration</td><td>An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td>UE host of type FOKUS-05
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td>Throughput</td></tr>
<tr><td>Additional tools involved</td><td></td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>8.323764</td><td>8.32448</td><td>8.325196</td><td>send</td></tr>
<tr><td>8.329213</td><td>8.32984</td><td>8.330467</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Thr-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average throughput between two communication end-points, namely the UE and the host running the Open5GCore Rel.3 in the network.</td></tr>
<tr><td>Purpose</td><td colspan="4">The test acts as a calibration test to primarily assess the uplink performance of the LTE link.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>11.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The communication endpoints are placed within the UE and the host running the Open5GCore Rel.3.</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">UE host of type FOKUS-05
Core Network host of type FOKUS-04 running Open5GCore Rel.3.
AirSpan Femto Cell and LTE dongle for wireless connectivity.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Throughput</td></tr>
<tr><td>Additional tools involved</td><td colspan="4"></td></tr>
</table>

<table>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td>Throughput
<table>
<tr><td colspan="3">Average Throughput [Mbps]</td><td>Traffic direction</td></tr>
<tr><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Upper bound of 95% confidence interval</td><td></td></tr>
<tr><td>8.322223</td><td>8.32332</td><td>8.324417</td><td>send</td></tr>
<tr><td>8.327357</td><td>8.32836</td><td>8.329363</td><td>receive</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td>n/a</td></tr>
</table>

11.1.5.4. Service Creation Time for a 5G packet core deployment

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a fully virtualized packet core (Open5GCore Rel.3).</td></tr>
<tr><td>Purpose</td><td colspan="4">The test assesses the time required to deploy a fully functional instantiation of the Open5GCore Rel.3 packet core.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>17.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The VM containing the virtualized core is downloaded to the compute host, deployed there, and all components of the packet core are powered up therein. IP addresses of base stations attached to the network are pre-configured</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
(e.g. HW components, SW components)</td><td colspan="4">The compute node used for deploying the service is FOKUS02
OpenStack controllers are run on FOKUS01.
Open5GCore Rel.3 VNF package
AirSpan Femto Cell attached to the network.</td></tr>
<tr><td>Metric(s) under study
(Refer to those in Section 4)</td><td colspan="4">Service Creation Time</td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing</td></tr>
<tr><td>Primary measurement results
(those included in the test case definition)</td><td colspan="4">Service Creation Time
<table>
<tr><td colspan="4">Service Creation Time [s]</td></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Lower bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>164.18</td><td>167.16</td><td>170.13</td></tr>
</table>
</td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

<table>
<tr><td>Test Case ID</td><td colspan="4">TC-Ser-001</td></tr>
<tr><td>General description of the test</td><td colspan="4">The test assesses the average service creation time of a fully virtualized packet core (Open5GCore Rel.3).</td></tr>
<tr><td>Purpose</td><td colspan="4">The test assesses the time required to deploy a fully functional instantiation of the Open5GCore Rel.3 packet core.</td></tr>
<tr><td>Executed by</td><td>Partner:</td><td>FOKUS</td><td>Date:</td><td>18.07.2019</td></tr>
<tr><td>Involved Partner(s)</td><td colspan="4">FOKUS, IHP, KAU</td></tr>
<tr><td>Scenario</td><td colspan="4">The VM containing the virtualized core is downloaded to the compute host, deployed there, and all components of the packet core are powered up therein. IP addresses of base stations attached to the network are pre-configured</td></tr>
<tr><td>Slicing configuration</td><td colspan="4">An exclusive slice / software-defined network is used for the test.</td></tr>
<tr><td>Components involved
<i>(e.g. HW components, SW components)</i></td><td colspan="4">The compute node used for deploying the service is FOKUS01
OpenStack controllers are run on FOKUS01.
Open5GCore Rel.3 VNF package
AirSpan Femto Cell attached to the network.</td></tr>
<tr><td>Metric(s) under study
(<i>Refer to those in Section 4</i>)</td><td colspan="4"><i>Service Creation Time</i></td></tr>
<tr><td>Additional tools involved</td><td colspan="4">TAP for automated testing</td></tr>
<tr><td>Primary measurement results
<i>(those included in the test case definition)</i></td><td colspan="4">Service Creation Time
<table>
<tr><th colspan="4">Service Creation Time [s]</th></tr>
<tr><td></td><td>Lower bound of 95% confidence interval</td><td>Value</td><td>Lower bound of 95% confidence interval</td></tr>
<tr><td>Mean</td><td>150.76</td><td>208.78</td><td>266.81</td></tr>
</table></td></tr>
<tr><td>Complementary measurement results</td><td colspan="4">n/a</td></tr>
</table>

12. SUMMARY AND CONCLUSIONS

This deliverable describes the trials and experimentation results from the first integration cycle of 5GENESIS. Upcoming versions of this deliverable will describe the trials and experimentation results from the second integration cycle (D6.2, M21) and the third integration cycle (D6.3, M36).

It describes eighteen baseline tests for nine different areas[24] and one application level test.

Based on the test case descriptions, 56 experiments have been performed by the project partners on the five test platforms represented in the project (Malaga, Athens, Limassol, Surrey and Berlin) with an initial strong focus on verifying the testing methodology developed in 5GENESIS.

For all these tests, numerical metrics are provided, as well as 95 % confidence intervals for all metrics, based on multiple repetitions of measurements during each experiment.

In addition to providing measurement results, this first integration cycle shows that all five platforms are operational and that the fully automated measurement tools and methology are working. It has also been established that the platform and infrastructure are well capable of handling the upcoming E2E evaluation and use case trials.

Upcoming deliverables 6.2 and 6.3 will add further test case descriptions and provide the results of additional experiments performed in subsequent integration cycles.

---

[24] Placeholders for three additional areas, namely location accuracy tests, reliability tests and speed tests, have already added to the baseline tests section, but these test were not described and performed during the first cycle of testing and will be added in documents describing later cycles (D6.2, D6.3).

REFERENCES

[1] https://5g-ppp.eu/wp-content/uploads/2017/10/Euro-5G-D2.6_Final-report-on-programme-progress-and-KPIs.pdf

[2] 5G PPP Test, Measurement and KPIs Validation Working Group White Paper, https://5g-ppp.eu/wp-content/uploads/2019/06/TMV-White-Paper-V1.0-03062019.pdf, June 2019

[3] IEEE P802.11.2/D1.01 – Draft Recommended Practice for the Evaluation of 802.11 Wireless Performance. February 2008.

[4] Jain, R.: The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation, and Modelling. Wiley Professional Computing, 1991. ISBN 978-0471503361.

[5] IEEE 100 -- The Authoritative Dictionary of IEEE Standards Terms. Seventh edition, December 2000. ISBN 0-7381-2601-2.

[6] 3GPP TS 28.554 – Management and orchestration; 5G end to end Key Performance Indicators (KPI). Release 16.0.0, March 28, 2019.

[7] 3GPP TR 22.872 – Study on positioning use cases; Stage 1. Release 16.1.0, September 2018

[8] 5G Research in Horizon 2020, European Commission, http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=5204

[9] 5G PPP PROGRESS MONITORING REPORT – 2017, https://5g-ppp.eu/wp-content/uploads/2018/10/5G-PPP-Progress-Monitoring-Report-2017.pdf

[10] 5GENESIS Deliverable D2.1, "Requirements of the Facility", 2018, https://5GENESIS.eu/wp-content/uploads/2018/11/5GENESIS_D2.1_v1.0.pdf

[11] 5GENESIS Deliverable D2.2, "Initial overall facility design and specifications", December 2018, https://5GENESIS.eu/wp-content/uploads/2018/12/5GENESIS_D2.2_v1.0.pdf

[12] 5GENESIS Deliverable D2.3, "Initial planning of tests and experimentation", February 2019, https://5GENESIS.eu/wp-content/uploads/2019/02/5GENESIS_D2.3_v1.0.pdf

[13] "Festival of Lights"[Online], https://festival-of-lights.de/en, retrieved FEB/2019

[14] OpenStack [Online], https://www.openstack.org

[15] OpenBaton [Online], https://openbaton.github.io

[16] "MONROE Project" [Online], https://www.monroe-project.eu/access-monroe-platform/, retrieved FEB/2019

[17] "Keysight TAP" [Online], https://www.keysight.com/en/pc-2873415/test-automation-platform-tap?cc=US&lc=eng, retrieved FEB/2019

[18] Open5GCore [Online], https://open5gcore.org

[19] ETSI NFV MANO [Online], http://etsi.org/technologies/nfv

[20] Florian Kaltenberger, Guy de Souza, Raymond Knopp, Hogzhi Wang, "The OpenAirInterface 5G New Radio Implementation: Current Status and Roadmap", [Online], http://www.eurecom.fr/publication/5822

[21] "ETTUS USRP X310" [Online], https://www.ettus.com/product/details/X310-KIT, retrieved FEB/2019.

[22] Docker [Online], https://www.docker.com

[23] Zabbix [Online], https://www.zabbix.com

[24] Iperf [Online], https://iperf.fr

[25] Ixia [Online], https://www.ixiacom.com/Keysight-technologies

[26] IxLoad [Online], https://www.ixiacom.com/products/ixload

[27] IxLoad-Ve [Online], https://www.ixiacom.com/products/ixload-ve

[28] Hesk [Online], https://www.hesk.com

[29] Osticket [Online], https://www.osticket.com Slack [Online], https://www.slack.com

[30] Github [Online], https://www.github.com

[31] 3GPP TR 32.972 V16.0.0 (2018-12): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Study on system and functional aspects of energy efficiency in 5G networks (Release 16).

[32] ETSI ES 202 706-1 V1.5.1 (2017-01): Environmental Engineering (EE); Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power Consumption - Static Measurement Method.

[33] ETSI ES 102 706-2 V1.5.1 (2018-11): Environmental Engineering (EE); Metrics and Measurement Method for Energy Efficiency of Wireless Access Network Equipment; Part 2: Energy Efficiency - dynamic measurement method.

[34] ETSI ES 203 228 (V1.2.1) (2017-04): "Environmental Engineering (EE); Assessment of mobile network energy efficiency".

[35] ETSI TS 132 425 (V12.0.0): "LTE; Telecommunication management; Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (3GPP TS 32.425 version 12.0.0 Release 12)".

[36] ETSI TS 132 425 V15.2.0 (2019-04): LTE; Telecommunication management; Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (3GPP TS 32.425 version 15.2.0 Release 15).

[37] ETSI TS 132 412 V15.0.0 (2018-07): Digital cellular telecommunications system (Phase 2+) (GSM);Universal Mobile Telecommunications System (UMTS); LTE; Telecommunication management; Performance Management (PM) Integration Reference Point (IRP): Information Service (IS) (3GPP TS 32.412 version 15.0.0 Release 15).

[38] ETSI TS 123 203 V15.4.0 (2018-09): Digital cellular telecommunications system (Phase 2+) (GSM);Universal Mobile Telecommunications System (UMTS); LTE; Policy and charging control architecture (3GPP TS 23.203 version 15.4.0 Release 15).

[39] ETSI TS 136 314 V15.2.0 (2019-04): LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Layer 2 - Measurements (3GPP TS 36.314 version 15.2.0 Release 15).

[40] ITU-T Recommendation L.1330 (03/2015) : Energy efficiency measurement and metrics for telecommunication networks.

[41] ITU-T Recommendation ITU-T L.1310 (07/2017): Energy efficiency metrics and measurement methods for telecommunication equipment.

[42] Recommendation ITU-T L.1331 (04/2017): Assessment of mobile network energy efficiency.

[43] ETSI TR 103 117 V1.1.1 (2012-11): Environmental Engineering (EE); Principles for Mobile Network level energy efficiency.

[44] ETSI ES 202 706-1 V1.5.1 (2017-01): Environmental Engineering (EE);Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power Consumption - Static Measurement Method.

[45] Recommendation ITU-R M.2083-0 (09/2015): IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond.

[46] NGMN White Paper 2015 - https://www.ngmn.org/5g-white-paper/5g-white-paper.html

[47] 3GPP TR 38.913 V15.0.0 (2018-06): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Scenarios and Requirements for Next Generation Access Technologies; (Release 15).

[48] 3GPP TR 21.866 V15.0.0 (2017-06): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Study on Energy Efficiency Aspects of 3GPP Standards (Release 15).

[49] ETSI ES 201 554 V1.2.1 (2014-07): Environmental Engineering (EE);Measurement method for Energy efficiency of Mobile Core network and Radio Access Control equipment.

[50] ETSI ES 202 336-1 V1.1.2 (2008-09): Environmental Engineering (EE);Monitoring and Control Interface for Infrastructure Equipment(Power, Cooling and Building Environment Systems used in Telecommunication Networks)Part 1: Generic Interface.

[51] 3GPP TR 36.927 V15.0.0 (2018-07): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Potential solutions for energy saving for E-UTRAN (Release 15).

[52] 3GPP TR 32.856 V15.0.0 (2017-06): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Study on Operations, Administration and Maintenance (OAM) support for assessment of energy efficiency in mobile access networks (Release 15).

[53] ETSI ES 202 336-12 V1.2.1 (2019-02) : Environmental Engineering (EE); Monitoring and control interface for infrastructure equipment (power, cooling and building environment systems used in telecommunication networks); Part 12: ICT equipment power, energy and environmental parameters monitoring information model.

[54] 3GPP TS 28.550 V16.0.0 (2019-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; Performance assurance (Release 16).

[55] 3GPP TS 28.552 V16.1.0 (2019-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; 5G performance measurements (Release 16).

[56] 3GPP TS 28.304 V15.0.0 (2018-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Control and monitoring of Power, Energy and Environmental (PEE) parameters Integration Reference Point (IRP); Requirements (Release 15).

[57] 3GPP TS 28.305 V15.1.0 (2018-12): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Control and monitoring of Power, Energy and Environmental (PEE) parameters Integration Reference Point (IRP); Information Service (IS) (Release 15).

[58] 3GPP TS 28.306 V15.0.0 (2018-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Control and monitoring of Power, Energy and Environmental (PEE) parameters Integration Reference Point (IRP); Solution Set (SS) definitions (Release 15).

[59] ETSI ES 203 539 V1.1.0 (2019-03): Environmental Engineering (EE); Measurement method for energy efficiency of Network Functions Virtualisation (NFV) in laboratory environment.

[60] 3GPP TS 28.622 V15.2.0 (2018-12): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Telecommunication management; Generic Network Resource Model (NRM) Integration Reference Point (IRP); Information Service (IS) (Release 15).

[61] 3GPP TS 28.531 V16.1.0 (2019-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; Provisioning; (Release 16).

[62] 3GPP TS 28.554 V16.0.0 (2019-03): 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; 5G end to end Key Performance Indicators (KPI) (Release 16).
[63] ITU-T Recommendation ITU-T L.1320 (03/2014): Energy efficiency metrics and measurement for power and cooling equipment for telecommunications and data centres.
[64] https://www.celticnext.eu/project-soogreen
[65] 3GPP TS 22.261 V16.1.0 (2017-09) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Service requirements for the 5G system; Stage 1 (R16)
[66] 3GPP, TR 22.861, "Feasibility Study on New Services and Markets Technology Enablers for massive Internet of Things; Stage 1"
[67] 3GPP, TR 22.862, "Feasibility study on new services and markets technology enablers for critical communications; Stage 1"
[68] 3GPP, TR 22.863, "Feasibility study on new services and markets technology enablers for enhanced mobile broad-band; Stage 1"
[69] 3GPP, TR 22.864, "Feasibility study on new services and markets technology enablers for network operation; Stage 1"
[70] 3GPP, TR 22.891, "Feasibility study on new services and markets technology enablers; Stage 1"
[71] Recommendation ITU-R M.2083-0 (09/2015): IMT Vision - Framework and overall objectives of the future devel-opment of IMT for 2020 and beyond.
[72] NGMN 5G Initiative Team, "A Deliverable by the NGMN Alliance", NGMN 5G White Paper, Feb.17, 2015 - https://www.ngmn.org/5g-white-paper/5g-white-paper.html
[73] Euro-To5G project, Deliverable D2.6, "Final report on programme progress and KPIs", https://5g-ppp.eu/wp-content/uploads/2017/10/Euro-5G-D2.6_Final-report-on-programme-progress-and-KPIs.pdf
[74] 5G-MoNARCH project, Deliverable D6.1, "Documentation of Requirements and KPIs and Def-inition of Suitable Evaluation Criteria", https://5g-monarch.eu/wp-content/uploads/2017/10/5G-MoN-Arch_761445_D6.1_Documentation_of_Requirements_and_KPIs_and_Definition_of_Suitable_Evaluation_Criteria_v1.0.pdf
[75] METIS project, Deliverable D1.1, "Scenarios, requirements and KPIs for 5G mobile and wireless system", https://metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf
[76] Euro-To5G project, Deliverable D2.6, "Final report on programme progress and KPIs", https://5g-ppp.eu/wp-content/uploads/2017/10/Euro-5G-D2.6_Final-report-on-programme-progress-and-KPIs.pdf
[77] OFCOM Spectrum Review, Joint response from ESOA and GVF, 2015. https://www.ofcom.org.uk/__data/assets/pdf_file/0019/57520/gvf_and_esoa.pdf
[78] OFCOM, Delivering the Broadband Universal Service Proposals for designating providers and applying conditions, 2019, https://www.ofcom.org.uk/__data/assets/pdf_file/0024/129408/Consultation-Delivering-the-Broadband-Universal-Service.pdf
[79] Smart-me - Smart energy solutions. Available: https://www.smartme.com/
[80] Law of large numbers, https://en.wikipedia.org/wiki/Law_of_large_numbers

ANNEX 1

Overview of specifications on EE KPIs and metrics for mobile networks

Telecommunication networks energy efficiency KPIs are defined by various SDOs / organizations and are of various natures [30] . They can be applied to either:

- whole networks (i.e. E2E), or to
- sub-networks (e.g. the RAN), or to
- single network elements, or to
- telecommunication sites, which contain network elements and site equipment.

Moreover, EE KPIs can also be categorized according to the operator's network life cycle phase they may apply to, e.g.:

- during the equipment procurement phase, mobile network operators may be willing to compare network elements from various vendors from an EE standpoint. Some EE KPIs and measurement methods have been specified for this purpose.
- during the Design / Build phase, mobile network operators are always faced to several design options and may be willing to compare them from an EE standpoint. This may happen for the whole network, sub-networks and for telecom sites. For telecom sites, EE KPIs have been specified.
- during the Run phase, mobile network operators need to assess the energy efficiency of the live network, as a whole (i.e. end-to-end), or for sub-networks, or for single network elements or telecom sites. Some EE KPIs and measurement methods have also been specified for this purpose.

Generally, EE KPIs for network elements are expressed in terms of Data Volume divided by the Energy Consumption of the considered network elements. In the case of radio access networks, an EE KPI variant may also be used, expressed by the Coverage Area divided by the Energy Consumption of the considered network elements.

In the remainder of this ANNEX, an overview of the main standards /recommendations addressing EE KPIs and metrics for mobile networks is provided. The list includes:

- **ETSI ES 202 706-1 (2017)**
- **ETSI ES 102 706-2 (2018)**
- **ETSI ES 203 228 (2017)**
- **ITU-T** Recommendation **L.1330 (2015)**
- **ITU-T** Recommendation **L.1331 (2017)**
- **ITU-R** Recommendation **M.2083 (2015)**
- **NGMN 5G whitepaper (2015)**
- **3GPP TR 38.913 (2018)**
- **3GPP TR 21.866 (2017)**
- **3GPP TR 32.972 (2018)**
- **3GPP TR 32.856 (2017)**

ETSI ES 202 706-1 (2017)

The ETSI ES 202 706-1 specification [31] defines methods to **evaluate the power consumption of base stations** (for GSM, UMTS and LTE) in **static mode**. These methods can be used by i) telecom equipment manufacturers in their labs. Measured KPIs are generally captured in product specification datasheets; and ii) MNOs may use such measurements to compare equipment from different vendors from an EE point of view. They can also make their own measurements in their own labs in order to check if they have the same results. The specification describes methods for:

- Average power consumption of BS equipment under static test conditions: the BS average power consumption is based on measured BS power consumption data under static condition when the BS is loaded artificially in a lab for three different loads, low, medium and busy hour under given reference configuration.
- **Daily average power consumption of the base station.**

Sections 7.2 & 7.3 of the specification describe calculation methods of **average static power consumption for integrated and distributed BS configurations** respectively. In [31] under static test conditions, the Base Station (BS) average power consumption is based on measured BS power consumption data when the BS is loaded artificially in a lab for three different loads (low, medium and busy hour) under given reference configuration.

The **power consumption of integrated BS equipment in static method** is defined for three different load levels as follows:

- P_{BH} is the power consumption [Watts] with busy hour load.
- P_{med} is the power consumption [Watts] with medium term load.
- P_{low} is the power consumption [Watts] with low load.

The load levels are defined differently for different radio systems. The model covers voice and/or data hour per hour. The models are provided in the annexes D, E, F of [31].

The **power consumption of distributed BS equipment in static method** is defined for three different load levels as follows (for details of load levels see the annexes D, E and F in [31]):

- $P_{BH,C}$ and $P_{BH,RRH}$ are the power consumption [W] of central and remote parts of BS with busy hour load.
- $P_{med,C}$ and $P_{med,RRH}$ are the power consumption [W] of central and remote parts of BS with medium term load.
- $P_{low,C}$ and $P_{low,RRH}$ are the power consumption [W] of central and remote parts of BS with low load.
  - **Note that ETSI ES 202 706-1 defines daily average power consumption of GSM/WCDMA/ LTE/WI-MAX base stations, defined for three different load levels, in a lab-based test setup (Measurement Lab Setup for STATIC power consumption measurements is provided in section 6.1.1 in [31]). The templates for test reporting are provided in ANNEX A, and Reference parameters for LTE system in ANNEX F of [31] respectively.**

ETSI ES 102 706-2 (2018)

The ETSI ES 102 706-2 [32] document defines the **dynamic measurement method** (section 6.2) and defines **base station energy efficiency KPI** (section 6.2.11). Under dynamic test conditions, the BS capacity is measured under dynamic traffic load provided within a defined coverage area and the corresponding power consumption is measured for given reference configurations. Dynamicity of measurements may be achieved thanks to dynamic load, activation / deactivation of radio network features, various user terminals performance and distribution. The results can be used to assess and compare the energy efficiency of base stations.

The TR defines the dynamic measurement method for evaluation energy efficiency:

- **BS EE** under dynamic load conditions: the BS capacity under dynamic traffic load provided within a defined coverage area and the corresponding energy consumption is measured for given reference configurations.
- ETSI ES 202 706-1 [31] defines **daily average power consumption** of the base station.

The base station energy efficiency KPI is an indicator for showing how a base station in an energy efficient way is doing work in terms of delivering useful bits to the UEs served by the base station. A base station is more energy efficient when doing more work with the same energy, doing the same work with less energy or in the best case doing more work with less energy. **The base station energy efficiency KPI is the ratio of delivered bits and consumed energy (reported in units of bits/Wh)** and is denoted by:

$$BSEP = \frac{DV_{total}}{E_{equipment}^{total}} \tag{1}$$

Where DV_{total} is the total delivered bits during the measurement for all three traffic levels according to section 6.2.6 in [32] and $E_{equipment}^{total}$ is the total consumed energy during the measurement period for delivering DV_{total} according to section 6.2.8 in [32].

Data Volume Measurement

All received data by the UEs during each measurement period for each traffic level shall be measured. The measured data is the net data volume and shall not contain any duplicated or retransmitted data. The data shall be generated as described in section 6.2.3 and annex C in [32]. The measured data will be used for calculation of BS efficiency KPI and is in bits.

Since the time period for the three load levels in a real network under a 24-hours period is different, three weighting factors are applied to the measurement results to reflect the time ratio of low load, medium load and busy-hour load levels in a 24-hours period respectively.

These weighting factors are denoted as W_{low} for low traffic, W_{medium} for medium traffic and $W_{busy\text{-}hour}$ for busy-hour traffic level and they are defined in annex C in [32].

The measured data volume in bits for low load level is denoted as $DV_{measured\text{-}low}$.

The measured data volume in bits for medium load level is denoted as $DV_{measured\text{-}medium}$.

The measured data volume in bits for busy-hour load level is denoted as $DV_{measured\text{-}busy\text{-}hour}$.

The total data volume for 24-hours period is calculated as following:

$$DV_{total} = \left(DV_{low} \times \frac{W_{low}}{T_{measurement\ low}}\right) + \left(DV_{medium} \times \frac{W_{medium}}{T_{measurement\ medium}}\right) + \left(DV_{busy\ hour} \times \frac{W_{busy\ hour}}{T_{measurement\ busy\ hour}}\right)[bits] \tag{2}$$

The three load levels shall be measured at middle frequency channel.

EC Measurement

The energy consumption of the base station under test shall be calculated during the whole measurement period. The total energy consumption of the base station will be the sum of weighted energy consumption for each traffic level i.e. low, medium and busy-hour traffic. Since the time period for the three load levels in a real network under a 24-hours period is different, three weighting factors are applied to the measurement results to reflect the low load, medium load and busy-hour load levels in a 24-hours period respectively. These weighting factors are denoted as W_{low} for low traffic, W_{medium} for medium traffic and $W_{busy\text{-}hour}$ for busy-hour traffic level and they are defined in annex C.

To calculate the energy consumption, the power consumption of the BS is sampled continuously (interval time Δt_m: 0,5 seconds or shorter) over the complete measurement period for each traffic level. For the integrated BS, is the measured power value for the i^{th} sampled measurement during the measurement period. The energy which is the energy consumption of the BS during the measurement is calculated as follows:

$$E_{equipment}^{traffic\_scenario\_x} = \sum_{k=1}^{n}\left(\Delta t_m \cdot P_{k,equipment}^{traffic\_scenario\_x}\right) \text{ [Wh]} \qquad (3)$$

For the distributed BS, $E_{C,\ equipment}$ and $_{ERRH,\ equipment}$ [Wh] are the energy consumption of the central and the remote parts in the dynamic method defined as:

$$E_{RRH,equipment}^{traffic\_scenario\_x} = \sum_{k=1}^{n}\left(\Delta t_m \cdot P_{k,RRH,equipment}^{traffic\_scenario\_x}\right) \text{ [Wh]} \qquad (4)$$

$$E_{C,equipment}^{traffic\_scenario\_x} = \sum_{k=1}^{n}\left(\Delta t_m \cdot P_{k,C,equipment}^{traffic\_scenario\_x}\right) \text{ [Wh]} \qquad (5)$$

Where $n = \frac{T_{measurement}}{\Delta t_m}$, and $T_{measurement}$ is the measurement time for each traffic level and Δt_m is the sampling period.

The measured energy consumption in Wh for low load level is denoted as $E_{equipment}^{measured-traffic\_scanrio\_low}$.

The measured energy consumption in Wh for medium load level is denoted as $E_{equipment}^{measured-traffic\_scenario\_medium}$.

The measured energy consumption in Wh for busy-hour load level is denoted as $E_{equipment}^{measured-traffic\_scenario\_busy-hour}$.

The total energy consumption for 24-hours period is calculated as following:

$$E_{total\ equipment} = \left(E_{low} \times \frac{W_{low}}{T_{measurement\ low}}\right) + \left(E_{medium} \times \frac{W_{medium}}{T_{measurement\ medium}}\right) + \left(E_{busy\ hour} \times \frac{W_{busy\ hour}}{T_{measurement\ busy\ hour}}\right) [Wh] \qquad (6)$$

For the calculation of the total energy consumption for distributed BS similar calculation as above for radio remote part and the central equipment part formulas (6.2) to (6.6) can be used. The sum of each part and then summing up these two parts to obtain the total energy consumption for a distributed BS.

- **Note that ETSI ES 102 706-2 defines LTE base station energy efficiency KPI based on total data volume & energy consumption for 24-hour period, defined for three different load levels, in a lab-based test setup (Measurement Lab Setup provided in section 6.2.1 in [3]). The templates for test reporting are provided in ANNEX A, and Data Traffic Models in ANNEX C of [32] respectively.**

ETSI ES 203 228 (2017)

The ETSI ES 203 228 [33] **defines energy efficiency metrics and measurement procedures in operational radio access networks.** Two high-level EE KPIs are defined:

$$EE_{MN,DV} = \frac{DV_{MN}}{EC_{MN}}$$

in which Mobile Network data Energy Efficiency ($EE_{MN,DV}$), expressed in bit/J, is the ratio between the performance indicator (i.e. Data Volume DV_{MN}) and the energy consumption (EC_{MN}), and

$$EE_{MN,CoA} = \frac{\text{CoA\_des}_{MN}}{EC_{MN}}$$

in which $EE_{MN,CoA}$, expressed in m^2/J, is the ratio between the coverage area (CoA-des$_{MN}$) and the energy consumption EC_{MN}. EC_{MN} is the yearly energy consumption and $\text{CoA_des}_{\text{MN}}$is the "coverage area" as defined in section 6.2.3].

This specification/recommendation considered as a point of reference also by 3GPP (SA and RAN) that deals with the methods and metrics to evaluate **EE for mobile radio access networks**, encompassing GSM, UMTS and LTE.

MN EC Measurement

The Mobile Network Energy Consumption (EC_{MN}) is the sum of the energy consumption of equipment included in the MN under investigation (see section 4). The network energy consumption is measured according to the assessment process defined in section 6 such that individual metrics are provided per RAT and per MNO. **The overall EC of the partial network under test** is measured as follows:

$$EC_{MN} = \sum_i \left(\sum_k EC_{BS_{i,k}} + EC_{SI_i}\right) + \sum_j EC_{BH_j} + \sum_l EC_{RC_l} \quad (1)$$

where:

- EC is Energy Consumption.
- BS refers to the Base Stations in the MN under measurement.
- BH is the backhauling providing connection to the BSs in the MN under measurement.
- SI is the site infrastructure (Rectifier, battery losses, climate equipment, TMA, tower illumination, etc.).
- RC is the control node(s), including all infrastructure of the RC site.
- *i* is an index spanning over the number of sites.
- *j* an index spanning over the number of BH equipment connected to the *i* sites.
- *k* is the index spanning over the number of BSs in the *i*-th site.
- l is the index spanning over the control nodes of the MN.

EC_{MN} shall be measured in Wh over the period of measurement T.

In order to allow a more precise assessment of the energy consumption impact of local factors (like location specific site equipment) it is requested to measure and report into the parameter EC_{SI_i} the site equipment consumption into two classes:

- ICT equipment (equipment directly needed to perform the telecom service).
- Support equipment (all equipment installed at the site which are needed to operate the particular site, but which are not directly needed for the telecom service, like air-conditioning, back-up power, lights, etc.).

Moreover, it is requested also to classify the site equipment according to operational temperature range. Based on such a classification the following additional network metric describing the energy consumption of the telecom equipment with reference to the total energy consumption shall be introduced:

$$SEE = EC_{BSs}/(EC_{BSs} + EC_{SI}) \quad (1a)$$

The above site energy efficiency **(SEE) metric gives an INDICATION of site energy efficiency (SEE)** in terms of how big fraction of energy is used for actual telecom equipment (telecommunication service delivery).

NOTE: SEE is defined by the ratio of "IT equipment energy" and "Total site energy", which generally includes rectifiers, cooling, storage, security and IT equipment. For datacentres, the "Total site energy" more globally includes building load, powering equipment (e.g. switchgear, uninterruptible power supply (UPS), battery backup), cooling equipment (e.g. chillers, computer room air conditioning unit (CRAC)) and IT equipment energy.

Data Volume Measurement

The Mobile Network performance metrics is derived from parameters of the MN under investigation (see section 5.1.3) relevant for energy efficiency, in particular the total data volume (DV_{MN}) delivered by all its equipment and its global coverage area (CoA_{MN}).

For packet switched services, DV_{MN} is defined as the data volume delivered by the equipment of the mobile network under investigation during the time frame T of the energy consumption assessment. The assessment process defined in section 5.1.3.1. shall be used:

$$DV_{MN-PS} = \sum_{i,k} DV_{BS_{i,k}-PS} \quad (2)$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T (see section 6). *i* and *k* are defined in formula (1).

For circuit switched services like voice, DV_{MN-CS} is defined as the data volume delivered by the equipment of the mobile network under investigation during the time frame T of the energy consumption assessment:

$$DV_{MN-CS} = \sum_{i,k} DV_{BS_{i,k}-CS} \quad (3)$$

where DV, measured in bit, is the performance delivered in terms of data volume in the network over the measurement period T (see section 6). *i* and *k* are like in formula (1).

Note that by "circuit switched", we mean here all voice, interactive services and video services managed by the MNOs, including CS voice, VoLTE and real-time video services delivered through dedicated bearers. The assessment process defined in section 6 shall be used.

The overall data volume is computed as follows:

$$DV_{MN} = DV_{MN-PS} + DV_{MN-CS} \quad (4)$$

DV_{MN} can be derived from standard counters defined in ETSI TS 132 425 [34] (3GPP TS 32.425) and ETSI TS 132 412 [36] (3GPP 32.412) for LTE or equivalent used for 2G and 3G, multiplying by the measurement duration T. The counters (in [35] and [36]) account also for QoS being reported in QoS Class Identifier (QCI) basis (see [37]).

NOTE 1: DV_{MN} includes data volumes for DL and UL.

NOTE 2: BH supervision and control data volumes are not considered (in order to include only the payload).

DV_{MN} is computed in unit of bit.

Coverage area (CoA_{MN}) is also considered as a mobile network performance metric in the MN designed primarily for coverage goals (and hence especially in RU environments). The assessment process defined in section 6 shall be used. CoA is computed in unit of m^2.

The DV_{MN} shall be measured using network counters for data volume related to the aggregated traffic in the set of BS considered in the MN under test.

For PS traffic, the data volume is considered as the overall amount of data transferred to and from the users present in the MN under test. Data volume shall be measured in an aggregated way per each RAT present in the MN and **shall be measured referring to counters derived from vendor O&M systems.**

For CS traffic (e.g. CS voice or VoLTE), the data volume is considered as the number of minutes of communications during the time T multiplied by the data rate of the corresponding service and the call success rate. The call success rate is equal to 1 minus the sum of blocking and dropping rates, i.e.:

$$Call\ Success\ Rate = (1 - dropping\ rate) \times 100\ [\%] \quad (5)$$

The dropping includes the intra-cell call failure (rate of dropping calls due to all the causes not related to handover) and the handover failure:

$$1 - \text{dropping rate} = (1 - \text{intracell failure rate})(1 - \text{handover failure rate}) \quad (6)$$

In order to include reliability in the measurement the aggregated data volume shall be provided together with the 95^{th} percentile of the cumulative distribution, for each RAT in the MN.

NOTE 1: It is not possible for data services to determine a user related QoS, i.e. to identify for each data connection if a target throughput has been reached using counters**. Such a computation would need the usage of probes** that is out of scope of the present document.

NOTE 2: As soon as the MDT related measurements in [35] are available the data volume may be measured according to the specification given therein (especially referring to section 4.1.8 in [35]). In this case, the per-user information about QoS can be obtained for data services and only connections with good QoS should be considered.

Coverage Area Measurement

The Coverage area is subject to network planning and intended services delivered within a certain geographic area. The coverage area shall be described by the following parameters:

- The total geographical area of a country (CoA_geo). This includes the total geographical area which falls into the network operator responsibility (total network and/or sub-area under investigation). A network might cover the geographical area only to a certain fraction (often defined by the license agreements, for example area coverage of a complete country or of a region).
- The designated coverage area (CoA_des). This area defines the area in which a network coverage is provided by the selected sub-network and is derived by planning models from network design, planned service and geographical data.
- A coverage quality factor (CoA_Qdes). This factor considers measured feedback from user equipment (as described in table 8 in [4]). This coverage quality factor signifies that networks might experience false coverage issues (e.g. inside buildings), load congestions or high interference issues.

Coverage quality

The actual coverage area where UEs can be served might differ from the originally designated coverage area (i.e. false coverage zones within the considered area). The coverage quality is a measure to estimate the actually covered fraction of the planned total coverage area. User equipment reports such as failed call attempts (table 1) shall be used to determine how well the users within the coverage area are covered. The coverage quality indicator shall be provided for network efficiency result evaluations. It is linked to network quality and has to be defined in relation to the quality of service (QoS) definitions.

A coverage map based on signal quality (SINR) could be used to determine the fraction of the total area were a signal quality above a certain minimum value is achieved. However, such maps require a large amount of measurements and usually drive tests. For the sake of an energy efficiency assessment it is not required to have the knowledge of the detailed network conditions such as the actual coverage hole locations. From an Energy Efficiency assessment point of view, it is important to know how many users/sessions or served users/sessions experienced problems because of lack of sufficient quality in relation to the total number of users/sessions or served users/sessions within the considered area. This allows a number of simplifications and an indirect determination of a quality factor.

The coverage quality factor for a base station is based on network failure reports of the UE. The coverage quality factor shall be measured based on coverage failures reported by the appropriate network counters:

CoA_Qdes = 1 - "percentage of users/sessions with coverage failure" (7)

The following indicators shall be used to calculate the coverage failure (details see table 8 [4]):

- RRC setup failure ratio (Call setup failure ratio).
- RAB setup failure ratio (UE-BS radio interface failure).
- RAB release failure ratio (UE-BS radio interface failure).

A further factor which can indicate a coverage issue is the handover drop ratio. However, a handover drop can have multiple reasons (cell overload, UE speed, etc.). Furthermore, the handover drop rate depends on the network structure (number of neighbour cells). Its calculation requires several additional network parameters and complicates the data collection and analysis significantly. This factor is therefore omitted.

The coverage quality factor for a site is defined as follows:

CoA_Qdes = (1 - RRC setup failure ratio) (1 - RAB setup failure ratio) (1 - RAB release failure ratio) (8)

The needed parameters are specified by 3GPP standards and the results can be obtained from the network management and supervision.

The failure ratios are the fraction of failures of the total amount of attempts:

- RRC setup failure ratio = (Σk Failed RRC connection establishmentsk)/(Σk Attempted RRC connection establishmentsk).
- RAB setup failure ratio = (Σk RAB setup failurek)/(Σk RAB setup attemptedk).
- RAB release failure ratio = (Σk RAB release failurek)/(Σk RAB release attemptedk).

where *k* is the index spanning over the number of BSs in the considered site.

Table 1: Measurement parameters required for LTE coverage quality calculation
(source reference : [34])

Parameter	Function	Counter name

RRC connection establishment failures	Radio resource control	RRC.ConnEstabFail.*sum*
RRC connection establishment attempts	Radio resource control	RRC.ConnEstabAtt.*sum*
E-RAB setup failures	Initial E-RAB setup	ERAB.EstabInitFailNbr.*sum*
	Additional E-RAB setup	ERAB.EstabAddFailNbr.*sum*
E-RAB setup attempts	Initial E-RAB setup	ERAB.EstabInitAttNbr.*sum*
	Additional E-RAB setup	ERAB.EstabAddAttNbr.*sum*
E-RAB release failures	E-RAB release	ERAB.RelFailNbr.*sum*
E-RAB release attempts	E-RAB release	ERAB.RelAttNbr.*sum*

The following averaging procedure is then used to obtain an average coverage quality factor (which needs to be reported along with CoA_des$_{MN}$) of the partial network under test:

$$\mathrm{CoA\_Qdes_{MN}} = \sum\nolimits_{\mathrm{i}} \mathrm{CoA\_Qdes_{S_i}} \mathrm{DCA_{S_i}} / \mathrm{CoA\_des_{MN}} \qquad (9)$$

where:

- S refers to the sites in the MN under measurement;
- *i* is an index spanning over the number of sites.

To avoid over counting, the sites designed coverage areas should be defined as the area where the signals from the cells of the site are stronger (Best Server). It holds true that:

$$\mathrm{CoA\_des_{MN}} = \sum\nolimits_{\mathrm{i}} \mathrm{DCA_{S_i}} \leq \mathrm{CoA\_geo} \qquad (10)$$

where:

- S refers to the sites in the MN under measurement;
- i is an index spanning over the number of sites.

Finally, ETSI ES 203 228 [33] defines a method to define sub-networks from which these EE KPIs are calculated and to extrapolate them to the operator's whole radio access network. **The EE measured for sub-network can be extrapolated to larger networks.** The extrapolation approach is discussed in section 7.

For data reporting templates, see ANNEX A in ES 203 228 or equivalent in ANNEX I of Rec. ITU-T L.1331.

- Note that ETSI ES 203 228 defines RAN energy efficiency KPI based on total data volume & energy consumption, over weekly/monthly/yearly periods, defined independently of load levels, in operational networks, for a sub-network (or partial network, denoted as the Mobile Network under investigation) comprising:
  - Base stations (e.g. Wide area BS, Medium range BS, Local Area BS, Home BS).
  - Site equipment (air conditioners, rectifiers/batteries, fixed network equipment, etc.).
  - Backhaul equipment required to interconnect the BS used in the assessment with the core network.
  - Radio Controller (RC).

ITU-T Recommendation L.1330 (2015)

The ITU-T Recommendation L.1330 [39] is considered as a point of reference also by 3GPP (SA and RAN) that deals with the methods and metrics to evaluate EE for mobile radio access networks, encompassing GSM, UMTS and LTE. The recommendation provides principles and concepts of energy efficiency metrics and measurement methods for telecommunication network equipment.

Recommendation ITU-T L.1330 provides a set of metrics for the assessment of energy efficiency (EE) of telecommunication (TLC) mobile networks, together with proper measurement methods. Such metrics are of extremely high importance to operators, given that the optimization of the energy performance of a single piece of equipment does not guarantee the overall maximum energy efficiency of a complex network formed by several interconnected equipment. Hence, through the metrics reported in this Recommendation, a better comprehension of network energy efficiency will be gained, not only for "total" networks, but also for "partial" networks, definable through either geographic or demographic boundaries.

This Recommendation was developed jointly by ETSI TC EE and ITU-T Study Group 5 and published respectively by ITU and ETSI as **Recommendation ITU-T L.1330 [39] and ETSI Standard ETSI ES 203 228 [33],** which **are technically equivalent**. This Recommendation describes the energy consumption (EC) and mobile network (MN) energy efficiency measurements in **operational networks.**

ITU-T Recommendation **L.1310 (2017)**

The ITU-T Recommendation L.1310 [40] specifies the principles and concepts of energy efficiency metrics and measurement methods for telecommunication network equipment. This Recommendation also specifies the principles and concepts of energy efficiency metrics and measurement methods for small networking equipment (Metric for DSLAM, MSAM GPON GEPON equipment) used in the home and small enterprise locations.

ITU-T Recommendation **L.1331 (2017)**

The Recommendation ITU-T L.1331 [41] considers the definition of metrics and methods used to measure energy efficiency performance of mobile radio access networks and adopts an approach based on the measurement of such performance on small networks, for feasibility and simplicity purposes. Such a simplified approach is proposed for approximating energy efficiency evaluations and cannot be considered as a reference for planning evaluation purposes throughout the network operation process. The same approach was introduced in ETSI TR 103 117 [42]; the measurements in testing laboratories of the efficiency of the base stations is the topic treated in ETSI ES 202 706-1 [31].

The Recommendation also provides an extrapolation method to extend the applicability of the assessment of energy efficiency to wider networks. The Recommendation was developed jointly by ETSI TC EE and ITU-T Study Group 5 and published by ITU-T and ETSI as **Recommendation ITU-T L.1331 [41] and ETSI ES 203 228 [33]** respectively, which **are technically equivalent.** This Recommendation describes the energy consumption (EC) and mobile network (MN) energy efficiency measurements in o**perational networks.**

ITU-R Recommendation **M.2083 (2015)**

The Recommendation M.2083 [44] establishes the vision for IMT for 2020 and beyond, by describing potential user and application trends, growth in traffic, technological trends and spectrum implications. With regards to Energy efficiency, the recommendation indicates that EE has two aspects: i) on the network side, energy efficiency refers to the quantity of information bits transmitted to/ received from users, per unit of energy consumption of the radio access network (RAN) (in bit/Joule); ii) on the device side, energy efficiency refers to quantity of information bits per unit of energy consumption of the communication module (in bit/Joule). The recommendation also stipulates that: The energy consumption for the radio access network of IMT-2020 should not be greater than IMT networks deployed today, while delivering the enhanced capabilities. The network energy efficiency should therefore be improved by a factor at least as great as the envisaged traffic capacity increase of IMT-2020 relative to IMT-Advanced for enhanced Mobile Broadband.

NGMN 5G whitepaper (2015)

Next Generation Mobile Networks (NGMN) White Paper [45] was considered as a basis for the development of 5G systems. In the White Paper, NGMN states that "Business orientation and economic incentives with foundational shift in cost, energy and operational efficiency should make 5G feasible and sustainable. "In particular, section 4.6.2 of [45] is thoroughly dedicated to energy efficiency, and it is stated that "Energy efficiency of the networks is a key factor to minimize the TCO, along with the environmental footprint of networks. As such, it is a central design principle of 5G". "Energy efficiency is defined as the number of bits that can be transmitted per Joule of energy, where the energy is computed over the whole network.

3GPP TR 38.913 (2018)

The 3GPP TR 38.913 [46] deals with the KPIs to be used to evaluate the performance of the new network in these scenarios. Among these KPIs, in Section 7 of [46], one paragraph is dedicated to “UE energy efficiency” (7.12), another one (7.14) to “Area traffic capacity” and the “User experienced data rate”. These two latter KPIs are relevant for the Energy Efficiency estimation. Finally, paragraph 7.19 is dedicated to “Network energy efficiency”. In such paragraph, it is clearly stated that “Network energy efficiency shall be considered as a basic principle in the NR design”. Qualitative inspection is suggested, for Energy Efficiency, but also quantitative analysis, in particular for

- comparing different solutions or mechanisms directly related to energy efficiency, when their impact is not obvious from qualitative analysis
- comparing the final NR system design with LTE to evaluate the overall improvement brought in terms of Network EE

The suggested quantitative KPI is defined as

$$\boldsymbol{EE_{global}} = \sum_{\boldsymbol{scenario\ K}} \boldsymbol{b_K EE_{scenario\ K}}$$

where

$$\boldsymbol{EE_{Scenario}} = \sum_{\boldsymbol{load\ level\ 1}} \boldsymbol{a_1 \frac{V_1}{EC_1}}$$

b_k refers to the weights of every deployment scenario where the network energy efficiency is evaluated.

V_1 refers to the traffic per second served by a base station (in bits/s)

EC_1 refers to the power consumed by a base station to serve V_1 (in Watt = Joule/s).

a_1 refers to the weight for each traffic load level.

EC” is the power consumed by a base station to serve V.

The suggested KPIs in this 3GPP TR are for use in simulations. For the calculation of the above KPIs, the following assumptions are made:

- Energy Efficiency Quantitative KPI should be **evaluated by means of system level simulations at least in 2 deployment scenarios**: one coverage limited environment (ex: Rural) AND one capacity limited environment (ex: Urban);
- Evaluation should not be for peak hour but based on a **24-hour daily traffic profile**.
- It is recommended that at least 3 load levels should be evaluated.

3GPP TR 32.972 (2018)

The 3GPP TR 36.927 [50] provides an overview of studies and/or normative works initiated by other SDOs / working groups on pre-5G and/or 5G radio access networks energy efficiency. It also inventories high-level EE KPIs defined by those SDOs / working groups and methods to collect required measurements. The study identifies potential use cases and requirements for i) 5G network energy efficiency assessment (measurement & reporting) and ii) energy efficiency optimization (i.e. energy saving).

3GPP TR 32.856 (2017)

The 3GPP TR 32.856 [51] reports how 3GPP OAM specifications can provide support for the assessment of energy efficiency in radio access networks as defined by ETSI ES 203 228 [33], thanks to measuring both network performance and energy consumption. It provides a gap analysis between [33] and 3GPP OAM Technical Specifications.

Measurement methods

Existing measurement methods

Existing measurement methods [30] for the calculation of EE KPIs for mobile networks (cf. ETSI ES 203 228 [33], ETSI ES 202 706-1 [31] and 3GPP TR 21.866 [47]) are based on the collection, on a per network node basis, of:

- Data Volume measurements, and
- Energy Consumption measurements.

In some deployment scenarios of radio access networks, Coverage Area measurements may be used instead of Data Volume measurements.

Data Volume measurements are collected via OAM, as performance measurements, e.g. as recommended in ETSI ES 203 228 [33] for radio access networks.

Energy Consumption information can be collected:

- using power meters or information from invoices provided by power suppliers,
- via built-in sensors, e.g. in case base stations (cf. ETSI ES 202 336-12 [52]), enabling the collection of energy consumption measurements via OAM,
- via external sensors and XCU/DGU, as specified in ETSI ES 202 336-12 [52].

The methodology may vary depending on whether the measurements are made in live networks or in test laboratories.

- In laboratories, ETSI ES 202 706-1 [31] defines a two-level assessment method to be used to both evaluate power consumption and energy efficiency of base stations. The two levels are:

- Base station equipment average power consumption for which it defines reference base station equipment configurations and reference load levels to be used when measuring base station power consumption.

- Base station equipment energy efficiency, defined as the measured capacity for a defined coverage area, divided by the simultaneously measured energy consumption.

- In live networks, ETSI ES 203 228 [33] recommends to split the total mobile network operator network into a small number of networks with limited size ("sub-networks"). These sub-networks are defined to represent some specific characteristics, for example:

- capacity limited networks representing urban and dense urban networks,
- sub-urban networks with high requirements for coverage and capacity,
- rural networks, which are usually coverage limited.

The size and scale of the sub-networks are defined by topologic, geographic or demographic boundaries.

The measurement method defined in ETSI ES 203 228 [33] for sub-networks provides the basis to estimate energy efficiency for large networks of one mobile network operator or within an entire country, applying extrapolation methods.

Potential measurement methods for 5G networks

The calculation of an EE KPI will rely on the collection of related measurement data of two types:

- Data volumes:
  - the reporting method (reporting Method) of data volumes for 5G network elements / functions will be specified in TS 28.550 [53] for stage 1, stage 2 and stage 3,
  - the measurements (e.g. counters) will be defined in TS 28.552 [54] for the 5G radio access network and the 5G core network.

- Energy consumption:
  - for non-virtualized parts of base stations, regardless of whether these base stations are equipped with built-in or external sensors, their energy consumption can potentially be collected via XCU/DGU and/or VS-RMS and/or their EM/DM (see Note 1), as specified in TS 28.304 [55], TS 28.305 [56] and TS 28.306 [57]. This potentially applies to non-virtualized core network elements as well;
  - for virtualized parts of base stations, the energy consumption of the Virtualized Network Functions (VNFs) is the energy consumption of the server(s) on which the VNF(s) run, minus the energy consumption of the subject servers when they are in idle mode. When multiple VNFs run simultaneously on a given server, how to measure their respective part in the overall energy consumption of the server is not specified. This potentially applies to virtualized core network functions as well.

NOTE 1: The measurement method described in ETSI ES 203 539 [58] is intended to be used to assess and compare the energy efficiency of VNFs in lab testing and pre-deployment testing; it aims not to define measurement method in operational NFV environment. In particular, it does not specify how the energy consumption of each server is measured individually in an operational environment.

Energy efficiency assessment in 5G

The assessment of energy efficiency of 5G networks may be based on the following potential solutions:

- 5G base stations are assumed to be all equipped with built-in sensors (cf. ETSI ES 202 336-12 [52]);
- Object model definition: the attribute 'peeParametersList' of IOC ManagedFunction, defined in TS 28.622 [59], may be used to model the PEE related parameters;
- Management services for network function provisioning, defined in TS 28.531 [60], may apply to read / write PEE related parameters and notify PEE related parameters value changes;
- Data volume measurements required to calculate DV (Data Volume) are to be defined in TS 28.552 [54]. KPIs may have to be defined in TS 28.554 [61];
- Power, Energy and Environmental (PEE) measurements required to calculate EC (Energy Consumption) are to be defined in TS 28.552 [54];
- The 'Measurement job control services for NFs', defined in TS 28.550 [53], may apply for the collection of DV and EC performance measurements data;
- The 'Performance data file reporting services for NFs, defined in TS 28.550 [53], may apply for the file-based reporting of DV and EC performance measurements data.

Upcoming specifications on EE in 5G mobile networks

The 3GPP SA5 WG, recommends starting normative work on the i) definition of use cases and requirements for 5G network energy efficiency assessment and optimization, ii) definition of performance measurements / KPIs enabling to assess the energy efficiency of 5G networks, for both aspects: data volumes and energy consumption, iii) definition of solutions for energy saving management in 5G networks.

The metrics and methods described in ES 203 228 [33]/ITU-T L.STP 5GEE for the legacy networks are considered valid for 5G Phase 1 (focus on eMBB) and an update will be issued once the 5G Phase 1 details are standardized.

The Phase 2 of 5G (rel. 16 and beyond), will impact heavily the specifications to measure energy efficiency and will require an extensive update of ETSI/ITU specifications, in tight cooperation with the standard bodies that will outline the new systems, especially 3GPP RAN and ITU-R. The objective is for example to leverage 3GPP SA5 work dealing with energy efficiency related analytics.

In summary, for 5G phase 1 [rel.15]:

- For operational networks, EE KPIs for whole or partial RAN, should be evaluated according to ITU-T L.1331 recommendation (or ETSI ES 203 228 which is equivalent).
- For lab-based test networks, EE KPIs for a base station, can be evaluated according to ETSI ES 202 706-1 and ETSI ES 102 706-2 for both static and dynamic operations.
- For simulation-based studies, EE KPIs for a base station, can be evaluated according to 3GPP TR 21.866 and 3GPP TR 38.913.

ANNEX 2

EE assessment reporting templates

The assessment report shall include tables defined below. Items in italics can be considered optional. Further guidelines on the test report can be found in clause 5.10 of ISO/IEC 17025.

Table 12 reports the details of the Network Area under test, representing a sub-network where the measurements are conducted. The Network Area is the area encompassing all the sites under measurement; the $\mathrm{CoA\_des}_{\mathrm{MN}}$ is instead computed starting from the area covered by each site and aggregating for all the sites in the Network Area under test.

For each site reported in Table 28 the details shall be included in Table 29. Table 30 reports the measurements results for each site.

Table 28: "Network Area under test" reporting template

Network Area under test		
Demography class [Dense Urban, Urban, Suburban, Rural, Sparse]		
Topography class		
Climate zone		
Informative classification		
Network Area definition [by Demography, by Geography, by Topology]		
	Number of inhabitants in the Network area [estimate]	
	Network Area dimensions [estimate, km²]	
	Number of sites in the Network Area [same radio controller?]	
Type of sites in the Network Area		
	Number of Wide Area BS sites	
	Number of Medium Range BS sites	
	Number of other sites/equipment (Local Area BS, relay nodes, etc.)	
Sites categorization		
	Number of sites in an MNO local exchange premise	
	Number of sites in buildings not owned by MNO	
	Number of sites in a shelter	
	Number of any other sites	
Multi-MNO sites		
	Number of "single MNO" sites	
	Number of co-located multi-MNOs sites	
	Number of sites in "Network Sharing" mode	
Multi-technology sites		
	Number of 2G only sites	
	Number of 3G only sites	
	Number of LTE only sites	
	Number of 2G+3G sites	
	Other options [indicate]	
Backhauling information		
	Predominant type of backhauling [wireless, fibre, copper…]	
	Number of backhauling links per type	

Network Area under test		
Energy efficiency in the Network Area under test		
	$EE_{MN,DV}$ [b/J]	
	$EE_{MN,CoA}$ [m^2/J]	
Energy efficiency top-down approach results (see note)		
NOTE: In case any alternative EE approach has been conducted on the network under test (i.e. measuring the aggregated energy consumption and the aggregated data volume or coverage area) the results of the evaluation shall be reported here for comparison purposes.		

Table 29: “Sites under test” reporting template

Site(s) under test in the Network Area **(one table per site type to be measured in the Network Area)**		
Measurement duration		
	Time duration of the measurement [T]	
	Measurement start date and time	
	Measurement finish date and time	
	Repetition time	
	Granularity of measurements	
Type of site		
	Site "layer" [Wide Area, Medium Range, other] In case of Wide Area, indicate number of sectors and carriers per sector	
	Site "technology" [2G, 3G, 2G+3G, LTE only, 2G+3G+LTE, other]	
	Site "MNOs" [single MNO, co-location, network sharing, other]	
Site and equipment age • *Initial commission date of the site* • *Commission date of the current equipment in the site*		
Temperature • *Average temperature [over period T]* • *Minimum temperature* • *Maximum temperature*	*Internal °C*	*External °C*
Environmental classTemp. range IC class (for each equipment in the site) *A 0 … 28 °C IP23* *B -20 … 40 °C IP45* *C -40 … 55 °C IP45*		
Site infrastructure		
	Site location [local exchange premise, building, shelter, other]	
	Site composition	
	• Air conditioners	
	• Rectifiers/batteries	
	• Fixed network equipment consumption	
	• Other	
	Estimated percentage of infrastructure consumption in the site (EC_{si})	
Energy consumption of ICT equipment in the site [Wh]		
Energy consumption of all the support equipment in the site [Wh]		

Site(s) under test in the Network Area (one table per site type to be measured in the Network Area)		
Energy efficiency in the site equipment (Energy_ICTe-quipment/Energy_Total_network)		
- Total electrical energy supplied from the grid - Peak power delivered from the grid - Total site energy storage capacity - Peak shaving features available at the site		
Energy Efficiency Enhancement methods affecting the site equipment during the test		
Estimated percentage of presence of this site type in the Network Area		
Electricity sources used in the site		
	Electricity [%]	
	Genset [%]	
	Solar [%]	
	Renewables [%]	
	Others (indicate)	

Table 30: “Site measurement” reporting template

Site measurement		
Measurement duration		
	Time duration of the measurement [T]	
	Measurement start date and time	
	Measurement finish date and time	
	Repetition time	
	Granularity of measurements	
Temperature class and average temperature during the test		
Energy consumption in the site		
	Method of measurement [energy bills/counters, sensors, equipment information, other]	
	Measured energy consumption EC$_{MN}$ [Wh or multiples]	
	• Week energy consumption [per week data/graph]	
	• *Month energy consumption [if T allows]*	
	• *Year energy consumption [if T allows]*	
Traffic offered in the site		
	Method of measurement [operational counters, backhauling data, MDT, other]	
	Measured traffic volume DV[bit or multiples]	
	• Week traffic [per week data/graph]	
	• *Month traffic [if T allows]*	
	• *Year traffic [if T allows]*	
Coverage of the site [data to be reported per each RAT present in the site]		
	CoA_geo: [km^2]	
	CoA_des: [km^2]	
	CoA_Qdes:	
	• Failed RRC connection establishments • Attempted RRC connection establishments • RAB setup failure • RAB setup attempted • RAB release failure • RAB release attempted	
Site Energy efficiency		
	Measured Energy Efficiency EE$_{MN}$ [bit/J] and [m^2/J]	
	• Weekly Energy Efficiency [per week data/graph]	
	• *Monthly Energy Efficiency [if T allows]*	
	• *Yearly Energy Efficiency [if T allows]*	

Table 31 reports an example of computation results of a total Mobile Network Energy Efficiency assessment. The EE values are in the format of tables for Partial network 1, and other values are considered in other Partial networks in the same partial network area (not reported in this example) to come to the average values in the EE columns. The Total EE is evaluated in the measurement period T timeframe (2 weeks) for the DV case, while EC is extrapolated to 1 year as required for CoA EE metric.

Table 31: Total (whole) Mobile Network Energy Efficiency assessment

Demography Class	Percentage of presence (PofP) in the total Network Area of the class	EE_{MN} in the class	
		$EE_{MN,DV}$	$EE_{MN,CoA}$
Dense Urban (DU)	42 %	*200 b/J*	*2,7 m^2/MJ*
Urban (U)	20 %	*40 b/J*	*19 m^2/MJ*
Sub-urban (SU)	15 %	*8 b/J*	*38 m^2/MJ*
Rural (RU)	13 %	*2 b/J*	*115 m^2/MJ*
Unpopulated	10 %	*NA*	*NA*
Overall/total EE		***103,8 b/J***	***28,4 m^2/MJ***

The following equations explain how to compute the Total EE in the cases mentioned above.

$$EE_{total,DV} = \frac{PofP_{DU} * EE_{DU,av} + PofP_{U} * EE_{U,av} + PofP_{SU} * EE_{SU,av} + PofP_{Unp} * EE_{Unp,av}}{PofP_{DU} + PofP_{U} + PofP_{SU} + PofP_{Unp}}$$

$$= \frac{42*200+20*40+15*8+13*2}{42+20+15+13} = 103{,}8\ b/J \tag{A.1}$$

$$EE_{total,CoA} = \frac{PofP_{DU} * EE_{DU,av} + PofP_{U} * EE_{U,av} + PofP_{SU} * EE_{SU,av} + PofP_{Unp} * EE_{Unp,av}}{PofP_{DU} + PofP_{U} + PofP_{SU} + PofP_{Unp}}$$

$$= \frac{42*2{,}7+20*19+15*38+13*115}{42+20+15+13} = 28{,}4\ m^2/MJ \tag{A.2}$$

Note that in the CoA case the extrapolation has been made from T = 14 days to 1 year dividing by 26 the results during period T (365/14~26).

ANNEX 3

Cloud RAN energy efficiency

The aim of this annex is to provide the basic information on definitions and principles to be used for the assessment of energy efficiency of Cloud RAN (CRAN) networks.

As far as energy efficiency assessment is concerned, the generic architecture of CRAN can be divided in 3 domains: central cloud, edge cloud and radio access.

The Radio Access (RA) domain consists of the Remote Access Points (RAP) dedicated to the CRAN under investigation. A typical RAP would include the radio, baseband and optical transport equipment. It would perform real time eNB tasks (e.g. Scheduler) and is installed near the transmitting antennas (e.g. within 1 m to 1 km). The density of RAP's deployed for CRAN would vary with different implementations but would be typically be of a few RAP units per 10 km^2. A typical value of RAP energy efficiency is: SEE_{RAP} = 90 %.

The Edge Cloud (EDC) domain is consisting of small datacentres dedicated to telecom functions, including Virtualized Network Functions (VNF) Servers (VNFS) used by the CRAN under investigation. A typical EDC datacentre would perform non-real time eNB tasks, such as Operations, Administration and Maintenance (O&M). The density of EDC datacentres deployed for CRAN would vary with different network configurations but would typically be of a few units per 100 km^2. A typical value of Edge Cloud site energy efficiency is: SEE_{EDC} = 75 %.

The Central Cloud (CC) domain is consisting of a datacentre (DC) including Central Servers (CS), Switching Equipment (SE) and other Telco Equipment (TE). The IP Core network equipment is not be taken into account in the assessment of CRAN EE. Central Cloud datacentres are usually very far from most of the served EDC. Their density would vary with different network configurations but would typically be of a few units per 100,000 km^2. A typical value of Central Cloud site energy efficiency is: SE_{ECC} = 65 %.

The following formulas can be used in the EE assessment for CRANs [33]:

Data Volume:

$$DV_{CRAN} = \sum_{RAP}(DV_{RAP-DL} + DV_{RAP-UL})$$

where DV_{RAP-DL} and DV_{RAP-UL} are the data volume of the RAP for downlink (DL) and uplink (UL) respectively.

Energy consumption:

$$EC_{CRAN} = \sum_{CC\ sites}\left(\sum_{site}(EC_{CS} + EC_{SE} + EC_{TE})\right)/SEE_{CC} \cdot + \sum_{EDC\ sites}\left(\sum_{Site} EC_{VNFS}\right)/SEE_{EDC} + \sum_{RAP} EC_{RAP}/SEE_{RAP}$$

Energy efficiency:

$$EE_{CRAN} = DV_{CRAN}/EC_{CRAN}$$

ANNEX 4

Confidential Annex of Measurements in Athens Platform

This Annex of the document is ranked as Confidential and as such, it has been delivered separately from this public document. The content refers to achieved values of throughput and latency in a pre-commercial end-to-end 5G system as extracted from a 5G vendor. The results indicate a preliminary internal test prior any potential use in the Athens Platform.

As resulted from the internal confidential tests, the 5G protocol stack reached the target of 1Gbit/s for the downlink rate.