5G smarT mObility, media and e-health for toURists and citizenS

Deliverable D7.1
Evaluation methodology

Project Details

Call	H2020-ICT-19-2019
Type of Action	RIA
Project start date	01/06/2019
Duration	36 months
GA No	856950

Deliverable Details

Deliverable WP:	WP7
Deliverable Task:	Task T7.2
Deliverable Identifier:	5G-TOURS_D7.1
Deliverable Title:	Evaluation methodology
Editor(s):	Evangelos Kosmatos (WINGS)
Author(s):	Alfredo Palagi (ERI-IT), Mara Piccinino (ERI-IT), Giancarlo Sacco (ERI-IT), Marco Gramaglia (UC3M), Marianna Legaki (UC3M), Simos Symeonidis (ACTA), Panayiotis Verrios (ACTA), Vlasis Barousis (NOK-G, R), Eduardo Garro (UPV), Felipe Vicens (ATOS), Baruch Altman (LIV), Roman Odarchenko (BLB), T. Doukoglou (OTE), G. Agapiou (OTE), V. Gezerlis (OTE), Nikos Papagiannopoulos (AIA), Evangelos Kosmatos (WINGS), E. Giannopoulou (WINGS), V. Stavroulaki (WINGS), P. Demestichas (WINGS), K. Tsagkaris (WINGS), S. Stavropoulou (WINGS), V. Foteinos (WINGS), P. Vlaheas (WINGS), M. Mitrou (WINGS).
Reviewer(s):	George Kokkinis (KEMEA), Simon Fletcher (RW), Eleni Rigani (NOK-GR), Linas Maknavicius (NOK-FR)
Contractual Date of Delivery:	29/2/2020
Submission Date:	28/2/2020
Dissemination Level:	PU
Status:	Completed
Version:	1.0
File Name:	5G-TOURS_D7.1_Evaluation_methodology_v1.0.docx

Disclaimer
The information and views set out in this deliverable are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Deliverable History

Version	Date	Modification
V1.0	*28/2/2020*	Initial version, submitted to EC through SyGMa

Table of Contents

List of Acronyms and Abbreviations

3GPP	3rd Generation Partnership Project
5G EVE	5G European Validation platform for Extensive trials
5G-MoNArch	5G Mobile Network Architecture for diverse services, use cases, and applications in 5G and beyond
5G-NORMA	5G NOvel Radio Multiservice adaptive network Architecture
5GS	5G System
5QI	5G QoS Identifier
AMF	Access and Mobility management Function
AR	Augmented Reality
BBU	Based Band Unit
BTS	Base Transmission System
CDF	Cumulative Distribution Function
CN	Core Network
CP	Control Plane
CPF	CP Function
CUM	Cumulative
DER	Discrete Event Registration
DL	Downlink
DRB	Dedicated Radio Bearer
DST	Destination
eMBB	enhanced Mobile Broadband
eNB	eNodeB
EPC	Evolved Packet Core
ETSI	European Telecommunication Standards Institute
EU	European Union
E-UTRA	Evolved Universal Terrestrial Radio Access
gNB	gigabit eNodeB
GNSS	Global Navigation Satellite System
HARQ	Hybrid Automatic Repeat Request
IPsec	Internet Protocol security
ISD	Inter Site Distance
ITU	International Telecommunication Union
KPI	Key Performance Indicator
MANO	Management and Orchestration
MEC	Mobile Edge Computing
METIS	Mobile and wireless communications Enablers for the Twenty-twenty Information Society
mMTC	massive MTC
MTBF	Mean Time Between Failures
MTC	Machine Type Communications
NF	Network Function
NFV	Network Function Virtualization
NFVI	NFV Infrastructure
NFVO	NFV Orchestrator
NGMN	Next Generation Mobile Networks
NR	New Radio
NR-RAN	NG Radio Access Network
NS	Network Slice
NSA	Non-Standalone
NSI	NS Instance
NSS	Network Slice Subnet
NST	Network Slice Template
NSVLD	Network Slice Virtual Link Descriptor
OTT	One way Time Trip
OWAMP	One-way Active Measurement Protocol
PDU	Protocol Data Unit
PNF	Physical Network Function
QCI	QoS Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RLC	Radio Link Control
RRC	Radio Resource Control
RRH	Remote Radio Head
RTT	Round Trip Latency
SDN	Software Defined Networking
SDU	Service Data Unit

SIM	Subscriber Identity Module
SMF	Session Management Function
SNR	Signal-to-Noise Ratio
S-NSSAI	Single – Network Slice Selection Assistance Information
SRC	Source
TBS	Terrestrial Beacon Systems
TCP	Transmission Control Protocol
TN	Termination Node
TWAMP	Two-Way Active Measurement Protocol
UC	Use Case
UDM	Unified Data Management
UDP	User Datagram Protocol
UL	Uplink
UP	User Plane
UPF	UP Function
URLLC	Ultra-Reliable Low Latency Communication
VM	Virtual Machine
VNF	Virtual Network Function
VR	Virtual Reality
XR	X Reality

List of Figures

List of Tables

Executive Summary

The 5G-TOURS project aims at deploying three full end-to-end trials involving real end-users (volunteers who consent to participate) and vertical operational services in three different European cities (Turin, Rennes, and Athens). On the three pilot sites[1], 13 use cases related with the themes of the touristic city (5 use cases), the safe city (4 use cases) and the mobility-efficient city (4 use cases) will be deployed. The ultimate goal of this approach is to trial them in a real environment by continuously collecting network, service and vertical KPIs and evaluate them against a set of predefined vertical-oriented criteria. Toward this direction, the definition and design of a clear evaluation methodology that can guide the pilot sites throughout the whole evaluation process is essential.

This deliverable is the first document produced by WP7 - System integration and evaluation, which focuses to deliver the integrated 5G-TOURS ecosystem that will allow for the realisation of pilots in all three nodes and drive the evaluation of the results of the trials. It presents the basic version of the 5G-TOURS evaluation methodology which will guide the pilot sites during all the steps of KPI validation process including: pilot test case design and definition, test case execution, data collection, data analysis and finally evaluation against predefined KPI targets. It provides the basic building blocks for the 5G-TOURS evaluation methodology, dealing with the evaluation of the network KPIs: latency, throughput, reliability, density, mobility, coverage, slice deployment time, security and location accuracy.

The first step toward the evaluation is to understand what the Vertical needs concerning trial execution in 5G-TOURS pilot sites and design a set of pilot tests. Although some tools exist related to test specification, normally these are more focused on making the life of the test design technician easier rather than on permitting a third party to express particular needs. Therefore, in order to cope with the aforementioned complexity and fill this gap, in 5G TOURS we will design a well-defined evaluation methodology in order to map the vertical requirements in an efficient way into the set of pilot tests. Then, the next step is the actual execution of the pilot tests on the three pilot sites. This step includes the preparation of the sites in terms of deploying the appropriate technologies, deploying the required network functionalities and also prepare the environments for the test execution. Then, the pilot tests are executed and the required metrics are collected. Finally, the collected metrics are analysed and evaluated against a set of predefined criteria.

The roadmap toward the final 5G-TOURS evaluation methodology includes the extension of the basic version presented in this document by evaluating the level of satisfaction of end-users and verticals players with the use cases deployed. This includes users' Quality of Experience (QoE) as well as the feedback from the vertical players on how the technology provided can improve their business' operations. The evaluation of end-users satisfaction will deal with QoE metrics derived from network KPIs as well as QoE metrics derived from Mean Opinion Score (MOS) approaches. The second step includes the interaction with T3.5 and will provide the final evaluation methodology fine-tuned to the network capabilities available in the pilot sites (e.g. MEC, SDN, MANO).

The 5G-TOURS evaluation methodology presented in the document comprised two parts:

- a general part that can be applied to all the project use cases;
- a specialised part that defines evaluation procedures that are specific to some use cases and involve particular KPIs.

The latter is useful firstly because only a subset of all the project KPIs is critical per each use case and secondly in order to capture the different service requirements and constraints per use case which may need particular network characteristics. In this way, the 5G-TOURS evaluation methodology is both generic (in order to support all the use cases of the project) and at the same time extendable (in order to be customised to support the special needs of each use case).

[1] Sites where trials are executed

1 Introduction

The 5G-TOURS project goal is to demonstrate the benefits of 5G technology in the pre-commercial environment for real users, tourists, citizens and patients by implementing 13 representative use cases in 3 different types of cities: (i) Turin, the touristic city (5 use cases) ii) Rennes, the safe city (4 use cases) iii) Athens, the mobility-efficient city (4 use cases).

5G-TOURS has devised a thorough evaluation plan to scrutinise the viability of the use cases, addressing technical performance by analysing network service and application-level KPIs, economic impact by analysing the estimated generated revenues and, ultimately, the satisfaction of the vertical customers. While economic analysis is part of WP8 - Business validation and exploitation, the technical performance and the satisfaction of the vertical customers are addressed within WP7.

In order to achieve the aforementioned target, the definition and design of a well-defined evaluation methodology that can guide the pilot sites throughout the whole evaluation process is essential.

The main targets of the methodology are:

- to provide a set of guidelines to the pilot sites based on wide acceptable standards and methodologies, for the practical evaluation of their use cases
- to provide a set of general procedures for KPI validation in order the evaluation to be realised in all use cases in a common way (when possible).

The current deliverable presents the 5G-TOURS evaluation methodology. The presented methodology targets guiding the pilot sites during all the steps of KPI validation process including: pilot test case definition and execution, data collection, data analysis and finally evaluation against predefined KPI targets. In the current document, the description of the basic evaluation methodology for 5G-TOURS is presented, being the main activity in Task 7.2. The final version of the 5G-TOURS methodology will be reported in Deliverable D7.2 (M22).

The 5G-TOURS evaluation methodology comprised two parts:

- a general part that can be applied to all the project use cases;
- a specialised part that defines evaluation procedures in a per KPI manner.

The latter is useful firstly because only a subset of all the project KPIs is critical per use case and secondly in order to capture the different service requirements and constraints per use case which may need particular network characteristics. In this way, the 5G-TOURS evaluation methodology is both generic (in order to support all the use cases of the project) and at the same time extendable (in order to be customised to support the special needs of each use case).

1.1 Role of WP7 in 5G-TOURS evaluation activities

The role of WP7 in the evaluation activities of 5G-TOURS as well as its interactions with the other WPs are illustrated in Figure 1. Specifically, regarding the interaction with WPs 4/5/6:

- Initially, WPs 4/5/6 provides WP7 with a description of the use cases and relevant metrics.
- WP7 based on this input, it defines a methodology to collect the KPIs and evaluate the level of satisfaction of the involved stakeholders and provide this methodology to WPs 4/5/6.
- WPs 4/5/6 perform the trials and provide the results back to WP7.
- WP7 analyses the results and provide the outcome of the evaluation.
- WP7 provide the outcome to WPs 4/5/6 for further enhancements and more trials, in order several iterations for the above steps will take place.

In addition, WP7 interacts with WPs 2/3/8 as follows:

- WP7 interacts with WP2 on the continuous refinement of the targeted KPIs, also from the QoE point of view.
- WP7 interacts with WP3 on the specific network capabilities required by the evaluation methodology. The interaction with T5.3 is important since it will provide the final evaluation methodology fine-tuned to the net-work capabilities available in the pilot sites (e.g. MEC, SDN, MANO).
- WP7 interacts with WP8 in order to provide inputs used by WP8 on the tecno-economic analysis of the achieved KPIs.

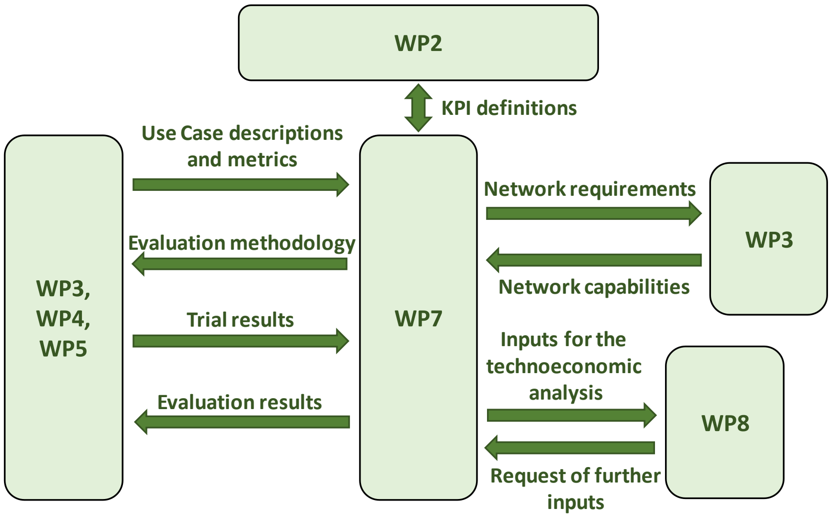

Figure 1. WP7 role and interactions

1.2 Structure of the document

The main structure of this deliverable is summarized as follows. Section 2 presents the 5G-TOURS general evaluation methodology. The proposed methodology is inspired on the methodologies used in two projects, building on these methodologies and expanding them:

- the model-based testing and validation methodology of the ongoing Phase 3 project 5G EVE [21];
- the evaluation methodology of the Phase 2 project 5G-MoNArch [22].

Section 3 describes the specialised evaluation procedures that will be used at the pilot sites for the evaluation of the use cases in a per KPI manner. The list of KPIs for which, one (or more) evaluation procedures are defined, is the outcome of a preliminary analysis on the list of project use cases. Our target in this analysis is to generate for each use case a list of KPIs that are of high importance for this specific use case. The outcome of the analysis, the definition of the selected KPIs and the procedure for their evaluation are described in Section 3. Section 4 reports the main conclusions of the document and gives summary directions on future work.

2 5G-TOURS General Evaluation Methodology

A variety of different testing and evaluation methodologies can be applied for the testing and validation of 5G infrastructures and services, many of which are coming from various 5G PPP projects of Phase 2 and Phase 3, while some initial approaches are coming from 5G PPP projects of Phase 1 (e.g. 5G-NORMA[19] and METIS II [20]). The requirement for the selection of a model-based testing and evaluation methodology is that the methodology would sufficiently encapsulate a model that is representative of the desired behaviour of a System Under Test (SUT) and the related testing and evaluation procedures and environment. In addition, this methodology should be general enough in order to be sufficiently flexible to be applied to all 5G-TOURS use cases, whilst being flexible enough to be customised or extended to address the needs of each specific 5G-TOURS use case.

To meet the aforementioned requirement the 5G-TOURS project shall adopt as the baseline methodology the Model-based Testing and Validation Methodology designed and developed in the 5G PPP Phase 3 project 5G EVE, given that the 5G-TOURS infrastructure is based on 5G EVE infrastructure. This approach will allow 5G-TOURS to exploit the facilities provided by the 5G EVE infrastructure. The methodology is presented in detail in the deliverables D5.1 [1] and D5.2 [2] of the 5G EVE project.

The 5G EVE testing and validation methodology provides a baseline validation methodology. The baseline will then be enhanced to build the general 5G-TOURS testing and evaluation methodology and then customised to meet specific evaluation procedures dealing with the relevant 5G KPIs for all the project use cases.

The 5G-TOURS general evaluation methodology describes the steps that should be taken starting from collecting the vertical objectives and requirements, collecting information related to:

- the services or applications the vertical will provide on top of the network infrastructure;
- the network topology and deployment;
- the pilot test scenarios that will be useful for the vertical to be evaluated, and finally;
- the list of KPIs together with their minimum requirements.

Then, the methodology translates all the aforementioned information into a set of technical specific test cases that the pilot sites should execute in order to evaluate the selected KPIs per Use Case (UC) against the preferable defined criteria.

The 5G-TOURS general testing and evaluation methodology is presented in the current section, while in the next section (Section 3) the 5G-TOURS evaluation procedures are presented. The term "evaluation methodology" is used to describe the general approach that will be followed by the project involving all the steps starting from collecting the requirements and ending by providing the evaluation results. The term "evaluation procedure" is used to describe the step-by-step procedure that should be followed for the evaluation of a specific KPI including the test definition, execution and metrics collection.

2.1 Workflow

The current section presented the workflow of evaluation methodology. In 5G-TOURS, the terminology used in the 5G EVE for testing and validation methodology will be reused. The main target of the evaluation methodology is to provide the appropriate guidelines to the pilot sites in order to design the test cases, to prepare the tests, to execute and finally evaluate the test results. Independently of low-level conditions, requirements, etc., the high-level workflow for the evaluation process, consists of the four phases presented in Figure 2. In this section, the terminology of 5G EVE is used.

<u>Test Design</u>

The main goal of the Test Design phase is to permit the developers of the trials to understand the objectives of the different experiments, how they can be executed, and what KPIs are critical for the evaluation process. This phase is key, as the whole interaction between verticals and pilot site providers will be influenced by how well they are able to align different perspectives from an early stage.

The main external input to this phase is the "Pilot test plan template" which is used by the Verticals in order to convey "in their own language" what are their intentions from the evaluation process. The Pilot test plan template is described in Section 0. The generated Pilot test plans are used by the Experiment Developers (technical experts) of the pilot sites in order to generate the "Low-level test plan templates". The low level templates translate the Vertical requirements into specific service descriptors, test descriptor and test scripts. These descriptors in 5G EVE terminology are called Blueprints, and they are defined and described in 5G EVE D5.2 [2]. In 5G EVE different type of Blueprints are defined, including Blueprints for Vertical Service/Context/Test Case/Experiment aspects [2]. In addition, other components are defined including test scripts, required VNFs/PNFs, etc.

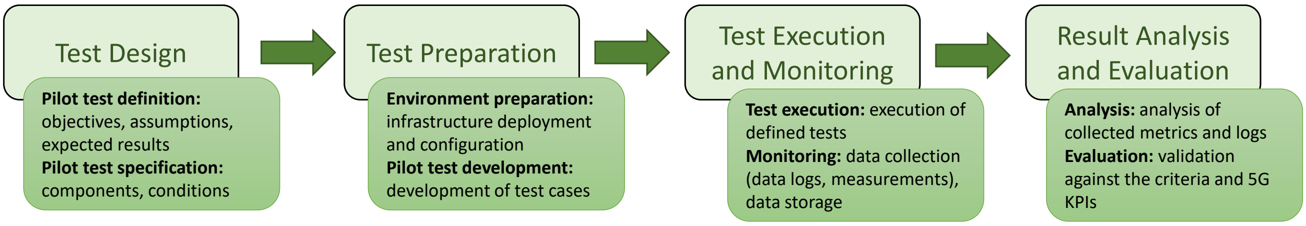

Figure 2. 5G-TOURS evaluation process workflow

The "Pilot test plan" includes a set of the test cases. These test cases define the experiments at the Experimenter-level and must be translated into a more technical and execution-oriented language. By the term "Experimenter" we define the person that will actually execute the pilot tests on the 5G-TOURS site (e.g. the vertical in many use cases). In other words, a meaningful test plan is to be generated, including not only the inputs from the experimenters, but also the testing procedures, the test configurations, the measurable KPIs (vertical- and network-related), the required supporting infrastructure, etc.

In order to accomplish this translation, and in order to generate an appropriate set of pilot site tests, the Experiment Developers will use the evaluation procedures described in Section 3. The main target of this translation is to create a set of low-level test procedures that if executed are enough to evaluate the KPIs defined in the high-level pilot test plans.

For example, if a Vertical requires to test a service deployed as Virtual Network Function (VNF) on the 5G EVE platform, it will have to provide the correct executable image file and descriptors so it can actually be deployed. The final outcome of this phase is an execution schedule, which will determine the time frame when the experimenter will be able to conduct tests.

Finally, it is also expected that VNF providers upload their packages into the pilot sites during this phase. The 5G EVE mechanisms which are implemented to make these packages available automatically across the pilot sites once they are correctly approved and validated will be reused by 5G-TOURS.

Test Preparation

The objective of this phase is to get everything ready to execute the desired trials. All trials have some initial conditions that must be set up by Experiment Developers, and most of them have some variables and threshold values for validation that Verticals may want to particularize. This customization enables the generation of the required descriptors from the blueprints.

From there, it should be ensured that the testing environment is ready to host the trial tests.

Test Execution and Monitoring

Once the entire infrastructure is ready, the virtual environment (e.g. virtual machines, containers, VNFs) to run the pilot tests is built and configured. This should be done automatically, and when ready, the Experimenter can start executing the different trial test cases. Both application and infrastructure metrics are gathered and made available both for monitoring (visualization) and for validation purposes.

Result Analysis and Evaluation

Results are collected and analysed, and the final evaluation report is generated. There are many levels of analysis, depending on many occasions on the KPIs which are of interest for each specific Vertical. A set of network

KPIs will be measured by default by 5G-TOURS as it will be based on the 5G EVE platform, in order to validate the results against the 5G PPP KPIs and as a way to ensure that the obtained results are valid (i.e. obtained under correct network conditions). The list of 5G EVE network KPIs are depicted in Table 1 (5G EVE D5.2 [2]). In this table the default KPIs considered by 5G-EVE are presented. For these KPIs we will reuse 5G-EVE methodology to evaluate them. Instead for other KPIs that are not in this list, we will develop a specific 5G-TOURS methodology.

Finally, and based on the gathered results, it is checked that the values of the KPIs are fulfilling the success criteria; if so, the final test report will indicate that the test has been passed; otherwise, it will have failed.

Table 1: 5G EVE list of Network KPIs

5G EVE KPIs	ITU-R M.24010-0 (11/2017)	Minimum requirements
Speed	DL User Experienced Data Rate (Mbps)	100 Mbps
	UL User Experienced Data Rate (Mbps)	50 Mbps
Broadband Connectiv-ity	DL Peak Data Rate (Gbps)	20 Gbps
	UL Peak Data Rate (Gbps)	10 Gbps
Capacity	Area Traffic Capacity (Mbit/s/m2)	10 Mbit/s/m2
Latency	UP Latency (ms)	1ms (URLLC), 4ms (eMBB)
	CP Latency (ms)	<20ms
Device Density	Connection Density (devices/km2)	1 M devices/ km2 (mMTC)
Mobility	Stationary (km/h)	0 km/h
	Pedestrian (km/h)	0 km/h to 10 km/h

In the evaluation methodology, among the four phases, the most critical and cumbersome phase is the initial "Test Design" phase. The "Test Design" phase requires a high level of interaction with the Vertical Customers and high effort in order the Vertical Customers' requirements and envisage scenarios to be mapped to a set of clearly defined test cases. In order to ease this process, we used a template, called "Pilot test plan" to be filled by the Vertical Customers. This template is explained in details in Section 2.2. In addition, after be filled by the Vertical Customers, this template is further translated to a set of "Low level test plans" which are the detailed technical descriptions of the test cases that should be executed during the "Test execution and monitoring" phase. The "Low level test plans" are described in details in Section 2.3. Based on the evaluation methodology, one "Low level test plan" may include one or several evaluation procedures. The high-level description of them is presented in Section 2.4. This high-level description is further specialized to a set of KPI specific evaluation procedures, presented in Section 3.

2.2 Pilot test plan

As described above, the main goal of the Test Design phase is to understand what the Vertical needs concerning trial execution in 5G-TOURS pilot sites. Although some tools exist related to test specification, normally these are more focused on making the life of the test design technician easier (e.g. time management applications, equipment catalogues, etc.) rather than on permitting a third party to express particular needs, and for those to be translated into tests, which is the focus of 5G-TOURS. Furthermore, the process is even more complex if we consider that not all Verticals have the technical knowledge related to 5G networks, MEC, SDN, MANO, etc., as to be able to prepare a test plan on their own. In 5G TOURS in order to cope with the aforementioned complexity and fill this gap, we will retain a high level of interaction between the technical persons and the verticals in order the vertical requirements to be mapped in an efficient way into the set of pilot tests.

In that sense, the 5G-TOURS project has prepared a pilot test description template, based on the 5G EVE high-level test plan [2], which is used as a first input gathered from Verticals willing to validate their trials using the project infrastructure. One important fact in the interaction between Verticals and Telcos (or Manufacturers) is that the “language” is not always the same. Telcos tend to focus on network parameters, conditions, SLAs, etc., while the implications of these concepts are not always fully understood by Verticals. Therefore, a lot of effort has been put in the template to make it understandable for experimenters, hiding the network related components as much as possible. The current template is a word document, which can be easily shared with Verticals that are going to execute the trials on the pilot sites. Afterwards, the results obtained through the vertical test plan template discussed here will be analysed through the lenses of the 5G MoNArch evaluation methodology, as discussed in Section 2.3.

2.2.1 Pilot test plan template

Two sections in the current template need to be filled by Vertical Customers. Both sections are required with specific instructions, dealing also with issues like the naming criteria.

2.2.1.1 Pilot test description

In the first section, the Vertical Customers is asked to include a brief description of the pilot tests that they would like to execute, and to outline their expectations from the 5G-TOURS pilot site (Figure 3).

Test plan template

A) Pilot test description

1) Please provide a brief description of the pilot test to be executed, including what are your expected capabilities from the 5G infrastructure (3-5 lines).

Figure 3. Pilot test plan template – Experiment description

The Vertical Customers is also asked to provide a list of the components that it will be bringing into pilot sites as part of the experiment, together with their deployment requirements and how they interact with each other (Figure 4). This is important, on the one hand to help determining hosting capabilities for the different sites, and on the other for the pilot site owners to understand the required connectivity among components.

2) Please include a brief description or figure preferably on the topology/deployment of the pilot infrastructure (how the different components interact with each other). This way we can derive the connectivity requirements for your pilot experiments. You can provide more than one deployment (e.g. one with server in the Cloud, the other with server in the ME. One without the 5G-Tours innovation, the other including the 5G-Tours innovation).

Figure 4. Pilot test plan template – Experiment components

Finally, the Vertical Customers is asked to list the KPIs which are meaningful for them (including the business-related KPIs). If these KPIs can be measured during experimentation, and for those which are not network-related, if there is a known relationship with network parameters on which we can leverage for the analysis of results (Figure 5).

3) Please, include a list of all the KPIs (5G network related or internal to your application) that are of interest for you during the pilot tests. Some of these KPIs will be used to determine the validation conditions (PASS/FAIL) of the experiment; some others will be used as variable parameters during execution. Please, list them all so it is clear which KPIs should be included in your test reports.

However, for KPIs internal to your application, please determine if they can be empirically measured during experimentation (either by your own components or by external ones), or if they can be mathematically derived from 5G network related KPIs. If there is such a known mathematical relationship with network parameters like latency, jitter, throughput, etc., please include it in the table below.

5G related KPIs	**Comments**	
High level KPIs	**Is it measureable during experimentation?**	**Can it be mathematically derived from 5G related KPIs?**

Figure 5. Pilot test plan template – Experiment KPIs

The 5G-EVE evaluation methodology that is at the basis of the 5G-TOURS one focuses on the specific KPI in isolation, due to its intrinsic test-based nature. However, the aim of 5G-TOURS is to go beyond this particular assumption, trying also to capture finer interactions across groups of KPIs as done, for instance, by the 5G MoNArch evaluation procedures (see Section 2.4.2). So, in order to help the integration with the 5G MoNArch evaluation methodology, in the "Comments" column, the vertical will be asked to indicate possible joint interactions in the KPI (e.g., different throughput / latency requirement levels).

This process will be reiterated, possibly done in consultation with a technical expert (e.g. from Vertical Solutions) if required.

2.2.1.2 Test plan

The second section of the template includes a list of test cases, ideally one per target KPI (Figure 6). Beyond the test name, and target KPI, Verticals are asked about the measurement method, about the test variables (e.g. parameter values that will be used in order to emulate specific network/application conditions) they would like to have available to ensure that the test is executed under appropriate circumstances, and what the validation conditions are (e.g. threshold for a specific variable).

B) Test plan

1) Please describe the set of pilot test cases that you intent to execute and validate. There should be (ideally one and only) one test case for each network KPI which is meaningful for the pilot site running the experiment. For each Test Case provide:

- Test Name
- Target KPI
- Measurement Method
- Parameters
- Validation Conditions

Measurement method: should include the procedure the pilot site uses to measure the target KPI E.g. component A is measuring KPI A, putting results in a database in component B, from where they can be accessed.

Parameters: should include which variables (external conditions) the pilot site uses to ensure the experiment conditions match those of the production environment. E.g. number of users, background traffic, etc.

Validation conditions: should include the conditions the target KPI should meet to consider the test as passed. E.g. KPI A should be below a certain threshold during the whole experiment. If the test is failed, the test report will give back the value of the parameter at which the Target KPI surpassed the threshold.

Test Case 1	
Test Name:	
Target KPI:	
Measurement Method:	
Parameters:	
Validation Conditions:	

Figure 6. Pilot test plan template – Test cases

Verticals can include as many test cases as required. Validation conditions will play a fundamental role in the overall understanding of the joint KPI fulfilment and Quality of Experience (QoE) validation process.

2.3 Low level test plan

Based on the Vertical inputs on the previous questionnaire, the consortium (specifically technical participants in each use case in cooperation with use case owner and pilot owners) can then derive a lower level test plan, valid to start the interactions with the pilot site owner. For that purpose, a second template (Figure 7) has been created, including all the information that typically populates the required test plans. Here it is important to mention that several lower level test plans can derive from a single Vertical test plan. Beyond fields like the test pre-requisites, topology, procedure, and so on, the field on “Required Capabilities” is very important since it will permit matching the needs of the experiment with the features offered by the different pilot sites. In case there are limitations, it will be possible to launch specific trials only at specific pilot sites. The generation of such test cases will be obtained by applying also the evaluation methodology discussed in Section 2.3.

B) Low level test plan

- There should be (ideally one and only) one test case for each high-level KPI which is meaningful for the Vertical running the experiment.
- If derived from the external conditions, the "expected results" change, then this can be reflected in different Test Sub-cases. Ideally, the Test Objective, Prerequisites, and Required Capabilities will be common, while the rest of the fields will be completed on a per sub-case basis.
- The fields on each Test Case are:
  - Test Objective, to be derived mainly from "Target KPI" in Vertical Test Plan
  - Test Prerequisites, to be derived mainly from "Experiment Description" and/or "Measurement Method" in Vertical Test Plan. This reflects if any initial conditions are to be met before launching the test
  - Required Capabilities, to be derived mainly from "Experiment Description" and/or "Measurement Method" in Vertical Test Plan. This reflects the requirements imposed to the potential hosting sites, to make sure they can support the experiment.
- The fields associated with each Test Sub-case are:
  - Test Topology, to be derived mainly from "Experiment Description"
  - Test Variables, to be derived mainly from "Parameters" in Vertical Test Plan
  - Test Procedure, to be derived mainly from "Experiment Description" and "Measurement Method" in Vertical Test Plan
  - Expected Results, to be derived mainly from "Expected Results" in Vertical Test Plan

Test Case 1		
Test Name:		
Test Objective:		
Test Prerequisites:		
Required Capabilities:		
Sub-Case 1	**Test Topology:**	
	Test Variables:	
	Test Procedure:	
	Expected results:	

Figure 7. Pilot test plan template – Low level test plan

It may happen that, depending on several internal or external conditions, for the same KPI the validation conditions change. For example, a specific service deployed in a container of a VM may support a maximum number of clients depending on the number of assigned virtual cores. The Vertical providing the service may want to deploy three flavours of the services, with 1, 2 and 4 cores respectively, and check how they behave depending on the delay. While it may be argued that these are three different use cases, we will consider that, since the target KPI is the same, it is easier to split them in different "sub-cases" (as depending on validation condition). The service flavours can be reflected based on slightly different test topologies, and each "sub-case" will have its own validation conditions (e.g. the VM with 1 core will PASS the test if it serves 1k subscribers, while the VM with 4 cores will PASS if it serves 4k subscribers).

In other words, "sub-cases" represent tests which are essentially the same, measuring the same KPI, but in which the validation threshold is different for whatever cause.

The test cases derived from the low-level test plan will be those which will be available for verticals to select the set of tests in the trial that will be actually executed, very similarly to what "step-by-step" GUI do in commercial traffic generation tools: these test cases represent very common scenarios, which may be useful for different trial executions.

2.4 High level description of testing and evaluation procedures

2.4.1 5G EVE evaluation procedures

In 5G EVE, the main procedure in all vertical applications testing is to vary the testing variable until the validation parameter is outside of its expected values [1]. The main result included in the Test Report, therefore, is that the test can be considered as PASSED until a certain value of the testing variable (KPI thresholds) is retained. In other words, the test FAILS when a certain value of the testing variable is reached.

On occasions, the validation parameter may not be directly measurable, but there is a well-established relationship with a measurable network parameter. This should be reflected directly in the low-level test plan, so that when the test is deployed, mechanisms to measure such network parameter are considered.

In some cases, it is requested only to present the values and not to be validated. In this cases, the 5G EVE platform only presents the measured KPIs versus the testing variable, without determining the final result of the experiment; this, if desired, shall be done by the vertical.

Finally, a very interesting practice consists of validating the deployment scenario before the actual testing is executed. Doing a preliminary benchmarking test to ensure that the network is in the appropriate conditions may simplify troubleshooting, in the sense that if the network is already experiencing uncontrolled impairments, it makes no sense to execute any test on top of it. Having said this, a high-level procedure for Vertical applications testing is depicted in Figure 8.

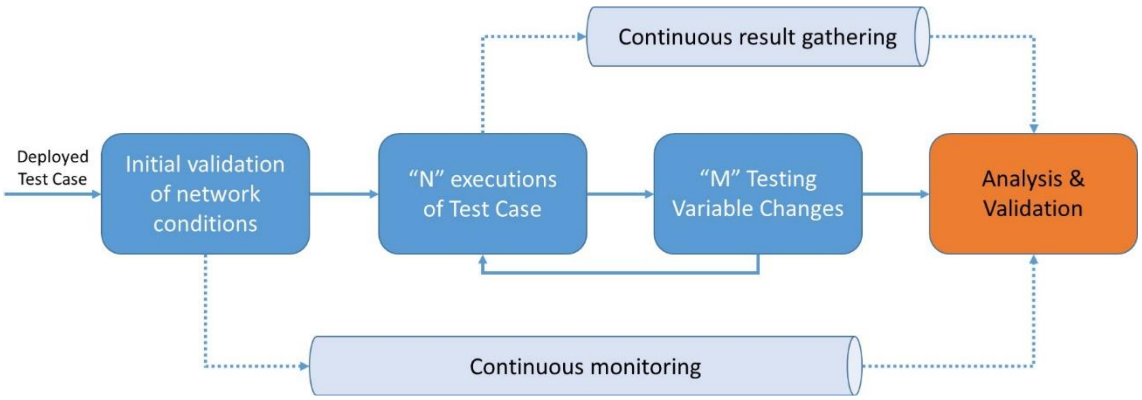

Figure 8: Vertical applications testing high-level procedure

2.4.2 5G-MoNArch evaluation procedures

The 5G MoNArch evaluation methodology [8] is a compound procedure that has been used by the project to validate the novel technologies proposed there (i.e. resiliency and elasticity) and evaluate the impact they have in realistic scenarios such as a port or a museum.

5G TOURS aims at different objectives (i.e., the validation of trials executed in the pilot sites), however the methodology drafted by the 5G-MoNArch project will be leveraged to define the 5G-TOURS one, as discussed in Section 2.1. In the following, we discuss in a nutshell the 5G-MoNArch evaluation methodology, leaving the interested reader to [8] for a comprehensive detailed description of the overall methodology.

The key objective of the 5G-MoNArch evaluation methodology was to verify in a quantitative and qualitative way the feasibility (both from a technical and economical perspective) of the proposed innovations of the project (this task has been called *verification*). Also, the evaluation methodology was employed to check that the innovation fit the stakeholder needs (or, in the project terminology, *validation*). In 5G TOURS these two concepts may naturally merge, as many of the verification tasks will be performed while validating them since the beginning by the different verticals that participate to the project.

To verify technical and economic feasibility, a network scenario has been selected in the Hamburg study area. This test scenario encloses the seaport and part of the Hamburg city. So, this environment has been used to verify that the network could satisfy the service-related requirements through the different KPIs by defining a set of evaluation cases and use cases. Clearly, in 5G TOURS the use cases are the ones envisioned by the project and discussed in Section 3.2.

In the 5G MoNArch evaluation methodology, the dependency of performance on the network parameters as well as the trade-offs between KPIs has been explored: this capability will be less relevant for the 5G-TOURS goals as the project aims to provide working trials of the 5G services through the 5G technology.
Finally, the 5G MoNArch evaluation methodology focused on three evaluation cases, that involved the usage of eMBB and URLLC connectivity in the same infrastructure deployment, which delivered smart cities, industrial and pure mobile Internet access services.

Some of the evaluation was performed through simulation, using ad-hoc tools, while other insights were obtained directly from the testbeds deployment in Turin and Hamburg.

Overall, the 5G-MoNArch evaluation methodology provides an excellent knowledge base about the advantages introduced by a flexible virtualized 5G network platform when delivering novel mobile services for vertical industries, which is exactly the goal of 5G TOURS. Also, a rather unique insight of this methodology has been to evaluate the overall impact of the technology not only from a technical perspective but also from a socio-economic point of view. In 5G-TOURS, these aspects will be studied in depth in WP8.

In particular, the project testbeds (which share many commonalities with the 5G TOURS test site and use cases), which involved vertical players and end-users together with the telco industry, were used to demonstrate the actual benefits of 5G technology in real environments (as 5G-TOURS targets).

2.4.3 5G-TOURS evaluation procedures

As discussed above, 5G-TOURS will integrate and extend the 5G EVE and 5G MoNArch evaluation methodologies and procedures. The iterative nature of 5G EVE evaluation procedures will be adopted in 5G-TOURS for the vertical application testing and validation, while the 5G MoNArch evaluation methodology [8] will be reused, in the aspects that are relevant for the development of the use cases.

In particular 5G MoNArch evaluation methodology will be complemented through the 5G EVE evaluation procedures. As a 5G-PPP phase 2 project, 5G MoNArch did not have a full set of implementation details that could be employed to gather all the possible KPIs, but it was rather focusing on specific KPIs that were relevant for the specific enablers implemented in the project. Thus, in order to evaluate advanced aspects such as the joint effect of different enablers, the project resorted to techniques such as system level simulators, leaving real world measurements coming from the testbed just for very specific KPIs (e.g., service creation time).

In 5G-TOURS there will be the ability to measure all the relevant KPIs for the specific UCs. This will be accomplished by leveraging the format specified by the 5G EVE evaluation procedures, so as to increase the compatibility with the underlying environment.

The evaluation methodology put in place by 5G-TOURS shall necessarily take into account the additional requirements that are imposed to the network. Each use case will need to foresee more heterogeneous inputs depending on the particulars of the application. These new metrics are strictly related to the devices type (e.g. Robot in the UC 2 and UC3 of WP4, XR visor in UC13), and their requirements to implement the specific application (for instance the mobility as metric of UC 9 of WP5 for the Ambulance or UC5 of the WP4 for Itinerant Orchestra).

Besides the verification of the technical KPIs, the extent of the use cases proposed by 5G-TOURS apart of the validation of the services from technical and business perspective will also examine the use case context from the ethical and societal point of view. By leveraging on the KPIs verification activities, the 5G-TOURS partners will evaluate the holistic impact of the proposed use cases.

This will be realized on the acceptability of the services perceived by the users (human experience in some context, and machine or connected things in others) and it is the definition of the QoE, as discussed in several Telecommunications fora.

For each use case some methodologies that aim to get feedback from the relevant End-User will be applied involving them in a direct discussion or, for instance, ask them to provide feedback through detailed questionnaires.

3 Evaluation procedures

In the currents section, initially the definitions of the KPIs identified by the 5G-TOURS project are presented. Then a categorization of the KPI are done in a per use case manner, by putting together requirements that were described in the initial deliverables of the project (D4.1[9], D5.1[10] and D6.1[11]). Then, for each KPI, one or more evaluation procedures are defined based on standards, past projects or widely accepted methodologies. The main objective of the current section is to specialize the general methodology of the previous section, in a per KPI manner. In that sense, the evaluation procedures presented here will be used as the main building blocks for the translation of "Pilot test plan" described in Section 0 into the "Low level test plan" described in Section 2.3.

3.1 5G KPI definition

5G-TOURS has identified a set of networks KPIs to be evaluated in the pilot tests and across the different pilot sites. That these KPIs comprise metrics related to the service provided to the end-users (such as latency, data rate, etc.) as well as others related to the operation of the network (such as deployment time and scalability). The definition of each KPI are presented below.

Table 2: 5G-TOURS KPI definitions

5G-TOURS KPIs	Definition
Latency	Latency is the time it takes to transfer a first/initial packet in a data burst from one point to another. (TS. 28.552).
Throughput	Throughput is the amount of information transmitted per unit of time. Throughput is usually measured in bits per second (bit/s or bps). User experienced data rate (bps) is a minimum achievable data rate for a user in real network environment. Peak data rate (bps) is a maximum achievable data rate per user.
Reliability	Reliability is the ability of the network to perform assigned tasks in certain operating conditions.
Density	Density is a total number of connected devices per unit area (Total connected devices (human & machines communications) within range of 5G signal divided by the total land (average #devices/Km2)). It is the ability to support the successful delivery of a message of a certain size within a certain time.
Mobility	Mobility(km/h) is a relative speed between receiver and transmitter under certain performance requirement.
Coverage	Coverage is a total land area covered by 5G signal divided by total land area.
Slice deployment time	Slice deployment time means the period that begins with the first procedures of deployment and ends when the slice is already deployed.
Security	Security consists of the policies and practices adopted to prevent and monitor unauthorized access, misuse, modification, or denial of a computer network and network-accessible resources.
Location accuracy	Location accuracy refers to the closeness of a measured location to the real location of the device at the time of the measurement.

3.2 Use cases KPI requirements

In the following table we list the network KPIs that all the 5G-TOURS use cases need for their correct operation. We performed a categorization of the KPI by putting together requirements that were described in a different way in D4.1[9], D5.1[10] and D6.1[11]. Not all the KPIs are relevant for all the use cases, so we left in the table a N/A where the KPI is not relevant (or cannot be evaluated) for the specific use case.

UC#	UC Name	Latency	Reliability	Density	Mobility	Coverage	Slice Deployment time	Throughput DL	Throughput UL	Security	Location accuracy
1	Augmented Tourism Experience	<15ms E2E	99,999%	~50 per Km2	< 10 Km/h	0.5 Km2	<90 min	200 Mbps per device	> 20 Mbps	Carrier grade	<1m
2	Telepresence	<10ms E2E	99,999%	< 5 per Km2	< 5 Km/h	0.5 Km2	N/A	>15 Mbps per device	>15 Mps per robot	Carrier grade	<1m
3	Robot-assisted museum guide and monitoring	<10ms E2E	99,999%	< 5 devices	< 5 Km/h	0.5 Km2	N/A	>10 Mbps	>15 Mps	Carrier grade	<1m
4	High quality video services distribution	<10ms E2E	99,999%	Not important if broadcast is used. ~15 devices with unicast.	250 Km/h	Turin city	< 60 min	25-100 Mbps	N/A	Baseline	Not important
5	Remote and distributed video production	<10ms E2E	99,999%	Very High	< 5 Km/h	N/A	Few Minutes.	N/A	>25 Mbps per camera	N/A	N/A
6	Remote health monitoring and emergency situation notification	<10ms E2E	99,999%	N/A	100 Km/h	Whole city	N/A	N/A	N/A	N/A	<200m
7	Teleguidance for diagnostics and intervention support	<10ms E2E	99,999%	6 per Km2	150 Km/h	Whole city	< 1 sec	>150 Mbps	>200 Mbps	Medical grade	< 10 m

8	Wireless operating room	<5 ms E2E	99,999%	<10 devices	Static	100m2	<90 min	>150 Mbps	>150 Mbps	Carrier grade	NA
9	Optimal ambulance routing	<10 ms E2E	99,999%	N/A	=>100 Km/h	3 Km2	N/A	N/A	N/A	Carrier grade	<100m
10	Smart airport parking management	<10 ms E2E	99,999%	100 devices	<=30 Km/h	Whole city	N/A	<=26 Kbps per device	<=66 Kbps per device	N/A	<5m
11	Video-enhanced ground-based moving vehicles	<10 ms E2E	99,999%	tens of videos	High speed	N/A	N/A	N/A	25 Mbps	N/A	N/A
12	Emergency airport evacuation	<10ms E2E	99,999%	Several users per m2	Low speed	N/A	N/A	~5 Mbps	N/A	N/A	N/A
13	Excursion on an Augmented Reality (AR)/Virtual Reality (VR)-enhanced bus	<10ms E2E	99,999%	Several users per m2	~80Lm/h	N/A	~ 60 min	500 Mbps	N/A	N/A	N/A

3.3 Evaluation procedures for 5G KPIs

The current subsection presents the procedures that will be adopted by the 5G-TOURS pilot sites for the evaluation of the network KPIs. These procedures are presented below in a per KPI manner. The evaluation procedures presented here will be used as the main building blocks for the translation of "Pilot test plan" into the "Low level test plan" as described in Section 0.

3.3.1 Latency

For the evaluation of latency KPI, widely accepted procedures described in standards and well-known protocols will be used, as described below. The selection of one or the other approach is based on the actual environment of the pilot sites that the use cases will be actually deployed.

Therefore, in cases, that latency is measured between two nodes (source node and destination node) that have clocks that are very closely synchronized the evaluation procedure based on RFC2679 [34] will be used. If the clocks are not synchronised, then we will measure Round Trip Latency (RTT) using the procedure described in 3GPP TS 28.554 [31]. In the latter case the one-way latency is defined as the RTT latency divided by 2. In cases, that in addition, to the actual latency value, we need also to evaluate other metrics derived from latency (e.g. jitter) then we will use specialised protocols for the evaluation of the latency. An indicative specialized protocol (TWAMP) [15] is described in Section 3.3.1.3 supporting both one-way and two-way evaluation capabilities.

3.3.1.1 Evaluation procedures in RFC2679

In RFC2679, the latency evaluation methodology is defined as below.

As with other Type-P-* metrics, the detailed methodology will depend on the Type-P. The Type-P describes the characteristics of the connection and the packets (e.g., protocol number, UDP/TCP port number, size, precedence) used throughout the tests.

Generally, for a given Type-P, the methodology would be as follows:

1. Arrange that Src (source node) and Dst (destination node) are synchronized; that is, that they have clocks that are very closely synchronized with each other and each fairly close to the actual time.
2. At the Src host, select Src and Dst IP addresses, and form a test packet of Type-P with these addresses. Any 'padding' portion of the packet needed only to make the test packet a given size should be filled with randomized bits to avoid a situation in which the measured delay is lower than it would otherwise be due to compression techniques along the path.
3. At the Dst host, arrange to receive the packet.
4. At the Src host, place a timestamp in the prepared Type-P packet, and send it towards Dst.
5. If the packet arrives within a reasonable period of time, take a timestamp as soon as possible upon the receipt of the packet. By subtracting the two timestamps, an estimate of one-way delay can be computed. Error analysis of a given implementation of the method must take into account the closeness of synchronization between Src and Dst. If the delay between Src's timestamp and the actual sending of the packet is known, then the estimate could be adjusted by subtracting this amount; uncertainty in this value must be taken into account in error analysis. Similarly, if the delay between the actual receipt of the packet and Dst's timestamp is known, then the estimate could be adjusted by subtracting this amount; uncertainty in this value must be taken into account in error analysis.
6. If the packet fails to arrive within a reasonable period of time, the one-way delay is taken to be undefined (informally, infinite).

3.3.1.2 Definition and evaluation procedure in 3GPP TS 28.554

In 3GPP TS 28.554 standard, the latency evaluation procedure presents the definition, description, method of measurement and collection of the latency samples. The findings are summarized in the below Table 3.

Table 3: Latency evaluation in 3GPP TS 28.554

KPI parameters	Description
Long name	End-to-end latency of 5G network
Description	This KPI describes the end to end packet transmission latency through the RAN, CN, and TN part of 5G network and is used to evaluate utilization performance of the end-to-end network.
Logical formula definition	This KPI is the RTT end to end latency of UE IP packets transmitted from UE to the N6 interface in the 5G network. The N6 interface is the reference point between UPF and DN.
Physical formula definition	E2ELatency
Measurement names used for the KPI	End-to-end latency
KPI Object	5GS
KPI category	Integrity
Unit of the KPI	Time interval (millisecond)
Type of the KPI	MEAN

3.3.1.3 Evaluation of latency-jitter-packet loss using the Two-Way Active Measurement Protocol

Latency [in miliseconds] is measured as either one-way (the time from the source sending a packet to the destination receiving it), or round-trip delay time (the one-way latency from source to destination plus the one-way latency from the destination back to the source) [12].

Jitter [in unit intervals (UI)] is the deviation from true periodicity of a presumably periodic signal, Jitter is a significant and undesired factor in the design of almost all communications links [13].

Packet loss [percentage %] occurs when one or more packets of data travelling across a computer network fail to reach their destination [14]. Packet loss can be caused by errors in data transmission or network congestion.

These three KPIs can be measured using the two-Way Active Measurement Protocol (TWAMP) [15]. TWAMP uses the methodology and architecture of the One-way Active Measurement Protocol (OWAMP). The OWAMP, specified in RFC4656, provides a common protocol for measuring one-way metrics between network devices. OWAMP can be used bi-directionally to measure one-way metrics in both directions between two network elements. However, it does not accommodate round-trip or two-way measurements. TWAMP is an open protocol for measurement of two-way or round-trip metrics, in addition to the one-way metrics of OWAMP and allows continuous measurements (24h basis) with traffic covering fully all the use case trial periods. In this case TWAMP is going to be used for measurements over ethernet.

TWAMP employs time stamps applied at the echo destination (reflector) to enable greater accuracy. TWAMP consists of two inter-related protocols: TWAMP-Control and TWAMP-Test. TWAMP-Control is used to initiate, start, and stop test sessions, and TWAMP-Test is used to exchange test packets between two TWAMP entities.

The TWAMP-Control and TWAMP-Test protocols accomplish their testing tasks as outlined below:

- The Control-Client initiates a TCP connection on TWAMP's well-known port, and the Server responds with its Greeting message, indicating the security/integrity mode it is willing to support.
- The Control-Client responds with the chosen mode of communication and information supporting integrity protection and encryption, if the mode requires them. The Server responds to accept the mode and give its start time. This completes the control-connection setup.
- The Control-Client requests a test session with a unique TWAMP-Control message. The Server responds with its acceptance and supporting information. More than one test session may be requested with additional messages.
- The Control-Client initiates all requested testing with a Start-Sessions message, and the Server acknowledges.
- The Session-Sender and the Session-Reflector exchange test packets according to the TWAMP-Test protocol for each active session.
- When appropriate, the Control-Client sends a message to stop all test sessions.

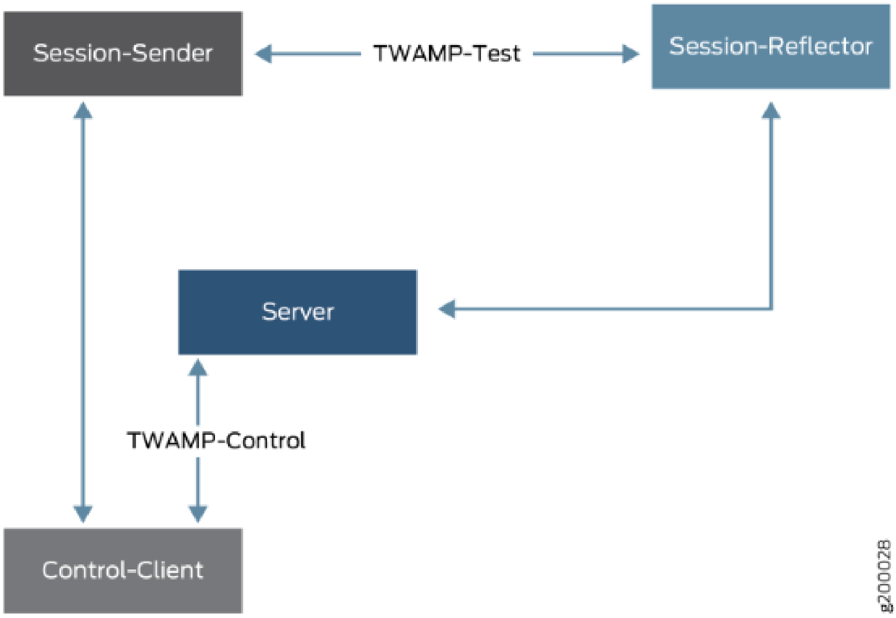

Figure 9: The four elements of TWAMP

3.3.2 Throughput

For the evaluation of throughout KPI, 5G-TOURS will apply evaluation procedures described in 3GPP TS 28.552 [30] and 3GPP TS 28.554 [31] standards. The selection of one or the other protocol is based on the granularity of the throughout measurements. In case, that we would validate the throughput in a per UE basis, then the 3GPP TS 28.552 standard will be adopted. In cases, that the accumulated throughout in a the network or in a network slices will be evaluated, the 3GPP TS 28.554 will be used.

3.3.2.1 Evaluation procedures in 3GPP TS 28.552

Two methods for the evaluation of throughout in 3GPP TS 28.552 are presented in the below tables. The first evaluates the “Average UE throughput in downlink”, while the second evaluates the “Average UE throughput in uplink”. In the tables, the template described in 3GPP TS 32404 is followed.

Table 4: Throughput evaluation in 3GPP TS 28.552 (Average UE throughput in downlink)

<table>
<tr><th>KPI parameters</th><th>Description</th></tr>
<tr><td>Description</td><td>This measurement provides the average UE throughput in downlink. This measurement is intended for data bursts that are large enough to require transmissions to be split across multiple slots. The UE data volume refers to the total volume scheduled for each UE regardless if using only primary- or also supplemental aggregated carriers. The measurement is optionally split into subcounters per QoS level (mapped 5QI or QCI in NR option 3).</td></tr>
<tr><td>Collection Method</td><td>DER(N=1)</td></tr>
<tr><td>Condition</td><td>This measurement is obtained according to the following formula based on the "ThpVolDl" and "ThpTimeDl" defined below. It is optionally split into subcounters for each QoS level.

$If \sum_{UEs} \sum \text{ThpTimeDl} > 0, \quad \frac{\sum_{UEs} \sum \text{ThpVolDl}}{\sum_{UEs} \sum \text{ThpTimeDl}} \text{x}1000\,[kbits/s]$ For small
$\text{If} \sum_{UEs} \sum \text{ThpTimeDl} = 0, \quad 0\ [kbits/s]$
data bursts, where all buffered data is included in one initial HARQ transmission, $ThpTimeDl = 0$, otherwise $ThpTimeDl = T1 - T2\,[ms]$

<table>
<tr><td>ThpTimeDl</td><td>The time to transmit a data burst excluding the data transmitted in the slot when the buffer is emptied. A sample of "ThpTimeDl" for each time the DL buffer for one DataRadioBearer (DRB) is emptied.</td></tr>
<tr><td>$T1$</td><td>The point in time after T2 when data up until the second last piece of data in the transmitted data burst which emptied the RLC SDU available for transmission for the particular DRB was successfully transmitted, as acknowledged by the UE.</td></tr>
<tr><td>$T2$</td><td>The point in time when the first transmission begins after a RLC SDU becomes available for transmission, where previously no RLC SDUs were available for transmission for the particular DRB.</td></tr>
<tr><td>$ThpVolDl$</td><td>The RLC level volume of a data burst, excluding the data transmitted in the slot when the buffer is emptied. A sample for ThpVolDl is the data volume, counted on RLC SDU level, in kbit successfully transmitted (acknowledged by UE) in DL for one DRB during a sample of ThpTimeDl. (It shall exclude the volume of the last piece of data emptying the buffer).</td></tr>
</table></td></tr>
<tr><td>Measurement Result (measured value(s), Units)</td><td>Each measurement is a real value representing the throughput in kbit per second. The number of measurements is equal to one. If the optional QoS level measurement is performed, the number of measurements is equal to the number of mapped 5QIs</td></tr>
</table>

Measurement Type	The measurement name has the form DRB.UEThpDl, or optionally DRB.UEThpDl.*QOS*, where *QOS* identifies the target quality of service class.
Measurement Object Class	NRCellDU
Switching Technology	Valid for packet switched traffic
Generation	5GS
Purpose	One usage of this measurement is for performance assurance within integrity area (user plane connection quality).

Table 5: Throughput evaluation in 3GPP TS 28.552 (Average UE throughput in uplink)

KPI parameters	Description
Description	This measurement provides the average UE throughput in uplink. This measurement is intended for data bursts that are large enough to require transmissions to be split across multiple slots. The UE data volume refers to the total volume scheduled for each UE regardless if using only primary- or also supplemental aggregated carriers. The measurement is optionally split into subcounters per QoS level (mapped 5QI or QCI in NR option 3).
Collection Method	DER(N=1)
Condition	This measurement is obtained according to the following formula based on the "ThpVolUl" and "ThpTimeUl" defined below. It is optionally split into subcounters for each QoS level. $If \sum_{UEs} \sum \mathrm{ThpTimeUl} > 0, \quad \frac{\sum_{UEs} \sum \mathrm{ThpVolUl}}{\sum_{UEs} \sum \mathrm{ThpTimeUl}} \mathrm{x}1000 [kbits/s]$ $\mathrm{If} \sum_{UEs} \sum \mathrm{ThpTimeUl} = 0, \quad 0 \ [kbits/s]$ For small data bursts, where all buffered data is included in one initial HARQ transmission $ThpTimeUl = 0$ otherwise: $ThpTimeUl = T1 - T2\ [ms]$ <table><tr><td>ThpTimeUl</td><td>The time to transmit a data burst excluding the data transmitted in the slot when the buffer is emptied. A sample of "ThpTimeUl" for each time the UL buffer for one DataRadioBearer (DRB) is emptied.</td></tr><tr><td>$T1$</td><td>The point in time when the data up until the second last piece of data in data burst has been successfully received for a particular DRB</td></tr><tr><td>$T2$</td><td>The point in time when transmission is started for the first data in data burst for a particular DRB.</td></tr></table>

	ThpVolUl	The RLC level volume of a data burst, excluding the data transmitted in the slot when the buffer is emptied. A sample for ThpVolUl is the data volume counted on RLC SDU level in kbit received in UL for one DRB during a sample of ThpTimeUl, (It shall exclude the volume of the last piece of data emptying the buffer).
Measurement Result (measured value(s), Units)	Each measurement is a real value representing the throughput in kbit per second. The number of measurements is equal to one. If the optional QoS level measurement is performed, the number of measurements is equal to the number of mapped 5QIs.	
Measurement Type	The measurement name has the form DRB.UEThpUl, or optionally DRB.UEThpUl.*QOS*, where *QOS* identifies the target quality of service class.	
Measurement Object Class	NRCellDU	
Switching Technology	Valid for packet switched traffic	
Generation	5GS	
Purpose	One usage of this measurement is for performance assurance within integrity area (user plane connection quality).	

3.3.2.2 Evaluation procedures in 3GPP TS 28.554

In 3GPP TS 28.552, the evaluation procedures focus on the network slice perspective (accumulated traffic from many users) as already explained. Two methods for the evaluation of Network Slice throughout are described in presented in TS 28.552 and summarised in the the below tables. The first evaluates the upstream throughput, while the second evaluated the downstream throughout. In the tables, the template described in 3GPP TS 28.554 is followed.

Table 6: Throughput evaluation in 3GPP TS 28.554 (Upstream throughput for Network Slice Instance)

KPI parameters	Description
Long name	Upstream throughput for network and network slice instance
Description	This KPI describes the upstream throughput of one single network slice instance by computing the packet size for each successfully transmitted UL IP packet through the network slice instance during each observing granularity period and is used to evaluate integrity performance of the end-to-end network slice instance.
Logical formula definition	This KPI is obtained by upstream throughput provided by N3 interface from NG-RAN to all UPFs which are related to the single network slice instance.
Physical formula definition	$UTSNSI = \sum_{AMF} GTP.InDataOctN3UPF$
Measurement names used for the KPI	GTP.InDataOctN3UPF
KPI Object	5GS
KPI category	Integrity

Unit of the KPI	kbit/s.
Type of the KPI	CUM

Table 7: Throughput evaluation in 3GPP TS 28.554 (Downstream throughput for Network Slice Instance)

KPI parameters	Description
Long name	Downstream throughput for network and network slice instance
Description	This KPI describes the downstream throughput of one single network slice instance by computing the packet size for each successfully transmitted DL IP packet through the network slice instance during each observing granularity period and is used to evaluate integrity performance of the end-to-end network slice instance.
Logical formula definition	This KPI is obtained by downstream throughput provided by N3 interface from all UPFs to NG-RAN which are related to the single network slice instance.
Physical formula definition	$UTSNSI = \sum_{UPF} GTP.OutDataOctN3UPF$
Measurement names used for the KPI	GTP.OutDataOctN3UPF
KPI Object	5GS
KPI category	Integrity
Unit of the KPI	kbit/s.
Type of the KPI	CUM

3.3.3 Reliability

The requirement of Reliability can be defined and measured in term of whole Telecommunication Network, or for a particular Network Service (i.e. Telephony) or it can even be defined only for a particular network segment.

A general definition of the Network Reliability is the capability of the network to offer the same services even during a failure. Single failures (node or link) are usually considered since they account for the vast majority of failures.

It should not be confused with Availability that measures the amount of uptime in a network system over a specific time interval. Uptime refers to the amount of time a network is fully operational. Network availability is measured as a percentage and is monitored to ensure the service being provided is consistently kept running for end-users. Analytically, this can be expressed as:

$$\text{Availability} = \frac{Uptime}{Total\ time\ (Uptime+Downtime)}$$

Reliability relates to Availability as follows [7]:

Network reliability is similar to availability; however, instead of measuring the amount of uptime in a system, reliability is the measured likelihood of a failure occurring in a system. Reliability will track how long a network's infrastructure is functional without interruption. Network reliability is also measured in percentages, where a fully reliable system has 100% availability. Reliability can be calculated through either dividing the total time in service by the number of failures (known as mean time between failures- MTBF) or by dividing the number of failures by total time in service (known as failure rate). More specifically, Reliability relates to the capability of transmitting a given amount of traffic within a predetermined time duration with high success probability.

Reliability refers to the continuity in the time domain of correct service and is associate with a maximum latency requirement. More specifically, reliability accounts for the percentage of packets properly received within the given maximum E2E latency (One way Time Trip (OTT) or RTT (Round Trip Time) depending on the service). For its evaluation dynamic simulations are needed, and realistic traffic models are encouraged.

Reliability (based on [25]) is the Percentage (%) of the amount of sent network layer packets successfully delivered to a given system node (incl. the UE) within the time constraint required by the targeted service, divided by the total number of sent network layer packets.

Two points can be stressed here:

- The reliability is evaluated only when the network is available.
- Dependent on the targeted service the RTT latency instead of the E2E (OTT) latency may be applied.

The RAN reliability can be evaluated by the success probability of transmitting X bytes within a certain delay of *t ms*, which is the time it takes to deliver a data packet from the radio protocol layer 2/3 SDU ingress point to the radio protocol layer 2/3 SDU egress point of the radio interface. The target communication range and reliability requirement is dependent on the selected deployment and operation scenario, i.e., by taking into account a certain channel quality (e.g., at the coverage edge). Link level evaluation with deployment scenario specific operating point and system level simulations are to be performed (e.g., Indoor Hotspot and Urban Macro for eMBB; Highway and Urban grid for connected cars/URLLC).

In classical resilience theory, the reliability of a system/component over time is directly related to its MTBF. In the simplified case that the MTBF will not change over the system's/component's lifetime, it can be calculated as follows:

$$R(t) = exp(-t/MTBF).$$

It should be noted, however, that the MTBF of most systems/components (respectively their *failure rate = 1/MTBF*) will change significantly over time ([26]).

The chosen Reliability KPI will be continuously verified during the project work. The type of analysis that can be followed will potentially be of three types:

a) *Analytical evaluation:* The verification process is performed through an analytical evaluation based on theoretical assumptions and values of the considered system.

b) *Simulation:* The verification process is performed through a SW simulation of the considered system that is modelled according to the goals of the verification.

c) *Testbed measurements:* The verification process is performed through experimental measurements during trials in the testbeds. The collected data is processed statistically according to the goals of the verification. Data can be objective (collected from systems) or subjective (collected from users).

The different levels of Reliability are the following:

- Low: < 97%
- Medium: 97 – 99%
- High: > 99,9% (Enterprise-Grade Reliability)
- Ultra High: > 99,999% (Carrier-Grade Reliability)

Concerning the 5G-TOURS KPIs measurements, two testbed measurement methods have been proposed for the Turin and Rennes trial sites, based on measuring the collected results from a) "ping" and b) "iperf" commands. In addition to that, for the Athens trial site "probes" can be used for the measurement of Reliability KPI as well as other KPIs. Probes can be installed between antennas, BBUs and central servers. Then, data and measurements will be collected for a certain period (6 months of 1 year) and using the corresponding analysis of data, the Reliability KPI for the Athens node 5G network will be measured.

3.3.4 Density

The November 2019 Ericsson mobility report estimated that there are over 8 billion active mobile devices around the world, with a steady yearly growth of 10%. 5G is about to enable a fully connected world. During the third quarter of 2019, service providers continued to switch on 5G and 13 million 5G subscriptions were estimated for the end of 2019. Over the next six years, 5G subscription uptake is expected to be significantly faster than that of LTE. In particular, by 2025, 2.6 billion 5G subscriptions are forecasted, carrying nearly half of the world's mobile data traffic. In addition, 5G is also expected to address massive IoT vertical use cases (also called mMTC), consisting of wide-area use cases, connecting massive numbers of low-complexity, low-cost devices with long battery life and relatively low throughput. With respect to this vertical, the number of massive IoT connections were estimated to have increased by a factor of 3x, reaching close to 1.3 billion at the end of 2019 with a CAGR of 25%, leading to another 5 billion IoT devices by the end of 2025 [24]. With this level of growth, it is agreed that massive IoT in high-density industrial plants shall support more than 1million devices per km2 according to IMT-2020 framework. These figures confirm that massive connection density is one of the main KPIs that 5G should fulfil and depending on the use case needs, lighter or more challenging requirements should be guaranteed. In the context of the project, the estimated values have been provided by 5G-TOURS vertical experts according to the table on Section 3.2, the most constrained but realistic requirements are foreseen in UC1, UC4, UC10, UC11, UC12, and UC13[2], and extracted in the table below for better readability.

Table 8: Connection density in UCs

Use Case	Density requirement
UC1: Augmented Tourism Experience	~50 devices per km2
UC4: High quality video services distribution	Unlimited per broadcast ~15 devices with unicast.
UC10: Smart airport parking management	100 devices
UC11: Video-enhanced ground-based moving vehicles	tens of videos
UC12: Emergency airport evacuation	Several users per m2
UC13: Excursion on an AR/VR enhanced bus	Several users per m2

As it can be seen the commonly agreed unit for measuring this KPI is (users/devices per km2). Based on this, several evaluation methodologies can be obtained from technical specifications as well as from 5G-EVE evaluation methodology which are briefly summarized below. The use of one or other methodology will depend on vertical and pilot owners test plan. In addition, it can also be noticed that the connection density requirements of 5G-TOURS use cases may be seen as not very challenging compared to the IMT-2020 5G 1 million devices per km2. Nevertheless, those massive number of devices requirements are particularly foreseen on Industry 4.0 use cases, which are out of the scope of 5G-TOURS project.

3.3.4.1 ITU-R M.2410-M.2412 (IMT-2020 framework)

In ITU, the standards ITU-R M.2410 [32] and M.2412 [33] describe the methodology for evaluating density using system level simulation, based on the methodology described in Table 9.

[2] Other 5G-TOURS use cases will require high throughputs, but the number of independent connections to be served will be less relevant than the highlighted above.

Table 9: Connection density in ITU-R M.2410 and M.2412

KPI parameters	Description
KPI term	Connection density
Definition	Total number of devices fulfilling a specific quality of service (QoS) per unit area (per km2)
KPI evaluation methodology (using simulations)	1. Set system user number per TRxP as N. 2. Generate the user packet according to the traffic model • The considered traffic model is message size of 32 bytes with either 1 message/day/device or 1 message/2 hours/device. Packet arrival follows Poisson arrival process for non-full buffer system-level simulation 3. Run non-full buffer system-level simulation to obtain the packet outage rate. • The outage rate is defined as the ratio of the number of packets that failed to be delivered to the destination receiver within a transmission delay of less than or equal to 10s to the total number of packets generated in Step 2. 4. Change the value of N and repeat Step 2-3 to obtain the system user number per TRxP N' satisfying the packet outage rate of 1%. 5. Calculate connection density by equation $C = N' /A$, where the TRxP area A is calculated as $A = ISD^2 \times \sqrt{3}/6$, where ISD refers to inter-site distance.

3.3.4.2 3GPP TS 28.552 and TS 28.554

In TS 28.552 and 28.554, the Performance Indicators related with density can be derived from the performance measurements collected at the NFs that belong to the group.

Table 10: Connection density in 3GPP TS 28.552 (RRC connection number)

KPI parameters	Description
KPI term	RRC connection number
Definition	The number of users in RRC connected mode during each granularity period.
KPI evaluation methodology (using simulations)	The number of the users in RRC connected and inactive mode need to be monitored as it reflects the load of the radio network, the operators can use this information for dynamic frequency resource allocation or load balance purpose. Moreover, it is an important factor to be evaluated in the radio network capacity enhancement decision-making. Different parameters can be envisaged 1. *Mean number of RRC Connections*: This measurement is obtained by sampling at a pre-defined interval, the number of users in RRC connected mode for each NR cell and then taking the arithmetic mean. It is provided in the field *RRC.ConnMean*

	2. *Max number of RRC Connections:* This measurement is obtained by sampling at a pre-defined interval, the number of users in RRC connected mode for each NR cell and then taking the maximum. It is provided in the field *RRC.ConnMax*

Table 11: Connection density in 3GPP TS 28.552 (Accessibility)

KPI parameters	Description
KPI term	Accessibility
KPI evaluation methodology	Different testing points can be envisaged 1. *Registered Subscribers of Network and Network Slice Instance through AMF[3]:* The total number of subscribers that are registered to a network slice instance by counting the subscribers in AMF. $$RSSNSI = \sum_{AMF} RegisteredSubNbrMean$$ 2. *Registered Subscribers of Network and Network Slice Instance through UDM:* The total number of subscribers that are registered to a network slice instance corresponding to the measurement RM.RegisteredSubUDMNbrMean that counts subscribers registered in UDM[4]. $$RSSNSI = RegisteredSubUDMNbrMean$$ 3. *DRB Accessibility for UE services:* The success rate for RRC connection setup multiplied by the success rate for NG signaling connection setup multiplied by the success rate for DRB setup. The success rate for RRC connection setup and for NG signaling connection setup shall exclude setups with establishment cause mo-Signaling. 4. *Registration success rate of one single network slice instance:* The ratio of the number of successfully performed registration procedures to the number of attempted registration procedures for the AMF set which related to one single network slice instance and is used to evaluate accessibility provided by the end-to-end network slice instance and network performance. $$RSR = \frac{\sum_{Type} AMF.5GSRegisSuccType}{\sum_{Type} AMF.5GSRegisAttType} * 100\%$$ *5.* PDU session Establishment Success Rate of one network slice (S-NSSAI[5]): The ratio of the number of successful PDU session establishment request to the number of PDU session establishment request attempts for 5G network for the *SMF*, which related to one network slice (S-NSSAI) and is used to evaluate accessibility provided by the end-to-end network slice and network performance. $$PSR.SNSSAI = \frac{SM.PduSessionCreationSuccNSI.SNSSAI}{SM.PduSessionCreationReqNSI.SNSSAI}$$

[3] AMF: Access and Mobility management Function

[4] UDM: Unified Data Management

[5] S-NSSAI: Single – Network Slice Selection Assistance Information

3.3.4.3 Evaluation methodology in 5G EVE

In 5G-EVE, the device density definition and evaluation methodology are described in deliverables D1.4 [23] and D5.2 [2].

Table 12: Connection density as defined in 5G EVE

KPI parameters	Description
KPI term	Device density (devices/km^2)
Definition	Number of simultaneous active connections per km^2 supported for massive sensor deployments, where active refers to those devices that are exchanging data with the network.
KPI evaluation methodology (using simulations)	Number of connected/accessible devices are available in NG-RAN node. In real networks this metric can be derived gathering the number of active devices connected to a radio node in a considered area. This area can also be specified given the covering capacity of the radio nodes or the configured ones. • *Measurement object*: “Number of 5G System Connected Users”: Counter available in NG-RAN node [gNB – (“5G base station”, providing NR access) or ng-eNB (“enhanced 4G base station”, providing E-UTRA access)]. Counter reports average number of connected users in the cell in the reporting period. • *Measurement time granularity*: “Number of 5G System Connected Users”: Counters is reported with a minimum granularity of 15 minutes (reporting period).

The appropriate measure of density for 5G-TOURS will be determined on a use case by use case basis.

3.3.5 Mobility

Providing reliable broadband wireless communications in mobility environments remains one of the main challenges in developing next generation wireless systems, especially in view of the cutting-edge requirement to satisfy strict ultra-reliable and low latency characteristics. Considering mobility, it imposes numerous challenges on the modeling, design, analysis and evaluations of upcoming 5G networks. Hence, the impact of mobility to the performance deserves the respective attention. Key concepts such as time-varying fading, channel estimation errors, intercarrier interference and frequent handovers are constitute main degrading factors (among others) and should be addressed. A comprehensive description of the main challenges that raise due to mobility requirements can be found in [16].

In the scope of the project, however, the impact of mobility to the selected KPIs will be evaluated from QoS point of view and in the boundaries of the respective use cases that will be demoed. In 5G-TOURS the mobility KPI will be evaluated primary by using the metrics already available on the different pilot sites (e.g. in the case of Greek site, the ACTA’s measurement platform and equipment will be used). In these cases, the 5G-EVE’s facilities will be also considered as a complementary solution. In sites, that no facilities for measuring mobility metrics are available, the 5G-EVE’s capabilities and approaches will be used.

3.3.5.1 Evaluation metrics

The objective is to draw the mobility effect on the diverse pre-defined KPIs, as listed below:

Latency
It is divided to control-plane (C-plane) latency and user-plane (U-plane) latency. C-plane latency can be defined as the time for a terminal to switch from idle state, where it is not connected to a radio resource control (RRC), to active state where the terminal is able to send data. Since the performance depends mainly on the U-plane

latency [17], our measurements will focus on user plane that is the dominant factor for low latency communication.
In high level, the one-way U-plane latency can be expressed as [18]:

$$L = L_{radio} + L_{backhaul} + L_{core} + L_{transport} \quad (1)$$

In the equation above, L_{Radio} refers to the latency between the client terminal and the eNB and it mainly concerns the processing time at the terminal and the eNB, plus the impact of physical layer communication (e.g. transmit time, propagation latency, channel estimation, encoding/decoding and retransmission time due to possible packet loss). Then, $L_{backhaul}$ is the time between eNB and core network, L_{core} is the processing time in the core network contributed by the various entities that would be involved therein and $L_{transport}$ is any delay between the core network and internet/cloud.
The two-way or round-trip time (RTT) refers to the time a signal needs to be sent plus the time it takes the ack to be received and can be safely approximated as 2L.

Jitter
Originally, jitter refers to the deviation of true periodicity of a signal. In case of networks it pertains to the variance in packet delay at the receiving side of a communication link, or equivalently it quantifies the fluctuation in delay as packets are being transferred across a network. This effect can be experienced as a disruption in the normal sequence of sending data packets, which eventually may reflect even to network congestion and packet loss. As it becomes understood, in the case of video streaming applications jitter can degrade the user experience and the QoS and therefore it is a major metric that should be evaluated and taken into consideration.

Packet Loss
In data communication, packets are exchanged between a server and a client. When noticeable jitter is present, there will be packets that will not be sent as expected and will arrive at once. This causes an overload for the requesting device, thus leading to congestion and a loss of data packets across the network. Then, the receiving endpoint possibly will try to correct the loss, but exact corrections cannot be made in all cases and the losses become irretrievable. It is therefore understood that jitter may cause packet losses that inevitably will influence the smooth flow experience of the video streaming.

Throughput
Ignoring any hardware imperfections, the upper-bound of the throughput is mainly controlled by the available bandwidth and signal-to-noise ratio (SNR). Hence, for given propagation conditions we could argue that the bandwidth approximately reflects to throughput and vice versa. Consequently, a high throughput implies adequate bandwidth that surely can support smooth and high-resolution video streaming.

3.3.5.2 Evaluation procedure

The objective is to evaluate the metrics above as a function of mobility. For this purpose, a convenient approach is to rely on the platforms and equipment that will be already deployed in the pilot sites (e.g. ACTA's measurement platform and equipment in the Greek site, other site specific solutions or 5G EVE facilities on other sites). The evaluation procedures include the position of pass-through probes that intercept the network in proper points in the network. This equipment can monitor the target KPIs continuously. Indicatively, those probes could be attached:

- At the use case site, just after the terminal device: camera, IoT sensor or IoT Gateway, 5G modem, etc.
- At the interface, between the 5G antenna (BTS) and before the backhauling equipment
- At the input stage of the Application servers (after the IP Core router)

Then an appropriate platform will manage the probes (e.g. ACTA's platform), while all measurements will be plotted by the platform as well. However, the measurements can be exported in a convenient format where each measured value is accompanied by a timestamp. Based on the above, we can do the all measurements while the vehicle is moving at a pre-defined and known speed. Then, we can repeat the same trial and the same route, but with different speed each time. Given that all measurements are uniquely defined by the timestamp, a tracking client application that will be running at vehicle will be used to easily filter out the values of interest and use them for further processing. All extracted values will be processed offline and their empirical cumulative distribution function (CDF) will be drawn to reveal any impact of mobility to the network performance and/or any impact of one KPI to another (e.g. jitter to packet loss). It is worthy noted that the latency measurements possibly will not reflect (1). Equation (1) implies that the latency depends on the network segment we target on, and therefore the extracted values will depend on the points where the probes will be attached.

3.3.6 Coverage

Evaluating the coverage KPI, is a complex task as it heavily depends on the physical infrastructure deployment that will be put in place in each test site. Thus, for the coverage KPI verification we follow an empirical approach, tailored per each UC, in which we will describe the tests performed to guarantee the coverage related to each city. We will devote particular attention to the UC that involve the usage of broadcast solutions, in which the coverage KPI strictly relates to density (as discussed in Section 3.3.3).

Thus, the KPI verification procedure will rely on estimations of the coverage maps that will be provided for each test site (and specifically for each location). Then, we will empirically select sub-areas within the expected coverage among the ones that are more relevant for the UCs development and check that all the rest of the KPIs are verified. For instance, for the UC5 (the itinerant orchestra) coverage shall be available both outside the Palazzo Madama site and inside, while for the UC1 this will be focused on the indoor coverage. Similar considerations apply to other test sites.

3.3.6.1 Turin Test Site

As depicted in Figure 10 below, the Turin tests site consists of 7 test sites, of which 4 will be certainly used for the 5G-TOURS use case development: Palazzo Madama, GAM (the gallery of modern art), the Edulab and the 5G EVE Mile of Technology. While other locations are possible (and fully described in D4.1[9]) the coverage shall be guaranteed in these 4 locations.

Coverage intrinsically relates to the radio infrastructure deployments available at each test site. Specifically, the Turin testbed will rely on Rel. 15 infrastructure configured in Non-Standalone (NSA) mode in which the control plane will use the LTE technology and just the user plane is provided through 5G technology. The network connectivity will rely on, especially for the UC that need broadcast support on a mixture of low tower-low power configurations and high tower – high power ones. As already discussed above, indoor coverage will be a fundamental part of the Turin test site coverage KPI validation. As a matter of fact, many of the use cases will be developed indoors, so a capillary coverage of the different museums rooms will be fundamental.

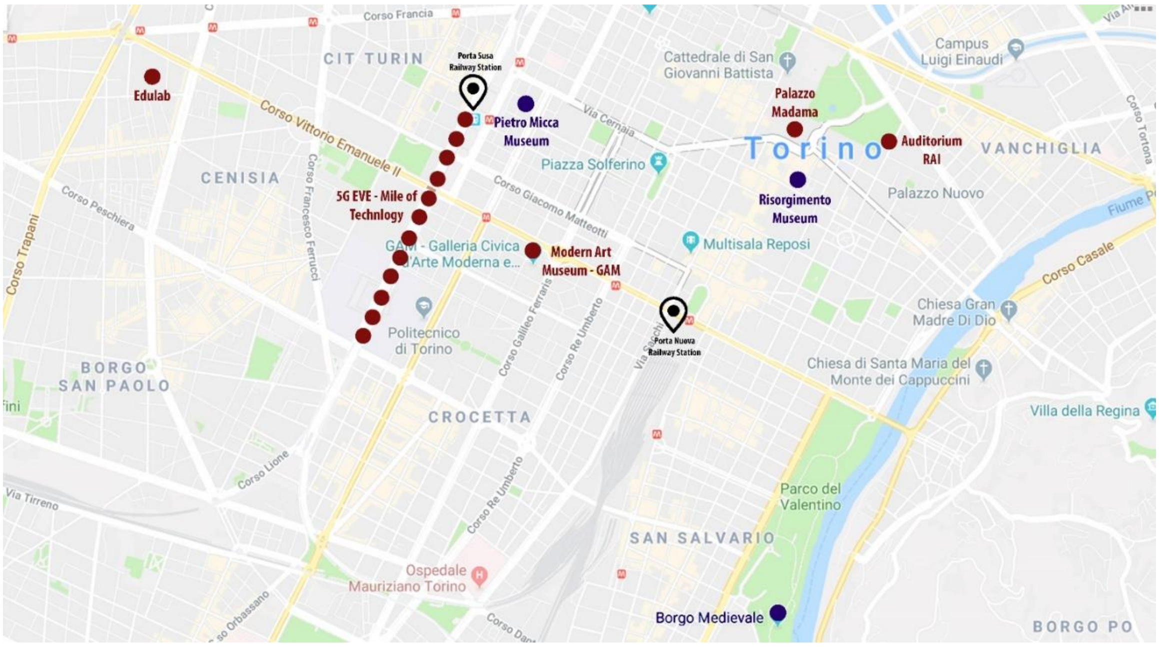

Figure 10: Turin Node “Touristic City”, City of Turin, 2019

3.3.6.2 Rennes Test Site

As already discussed for Turin, the coverage for the Rennes test site includes several radio sites, as depicted in Figure 11 below, deployed at B-Com and Rennes Hospital premises.

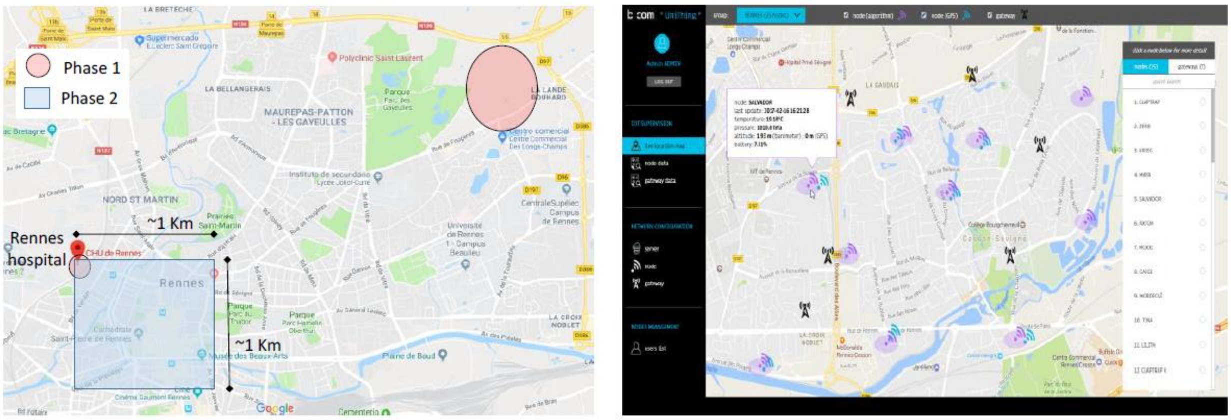

Figure 11: 5G EVE coverage in Rennes

The RAN infrastructure will use bands around 2.6 GHz ad 26 GHz (pending confirmation from the French regulation authority). This deployment both comprises outdoor and indoor base stations to provide coverage for both the use case that foresee in vehicle coverage and the ones such as the wireless operating room that need very high bandwidth in a small environment. Thus, the KPI verification will have specific ad-hoc tests both indoors and outdoors to check that all the required metrics are fulfilled.

3.3.6.3 Athens Test Site

Figure 12. Coverage area at the AIA

The coverage in the Athens test site is split into two subzones: the main one around the Athens airport and the one in the Psalidi area, in the northern part of Athens (see Figure 12 and Figure 13). For the operation of the use cases, the coverage around the area (only the airport has a size of 16 Km^2) different additional nodes will be placed in both the locations to guarantee coverage, including 5G (Rel 15) Remote Radio Heads (RRH) and antennas on the bands 7 and 43. Again, different location tests will be selected to probe the network KPI envisioned for each UC, leveraging the existing testing infrastructure deployed in the Greek site (see D6.1 [11])

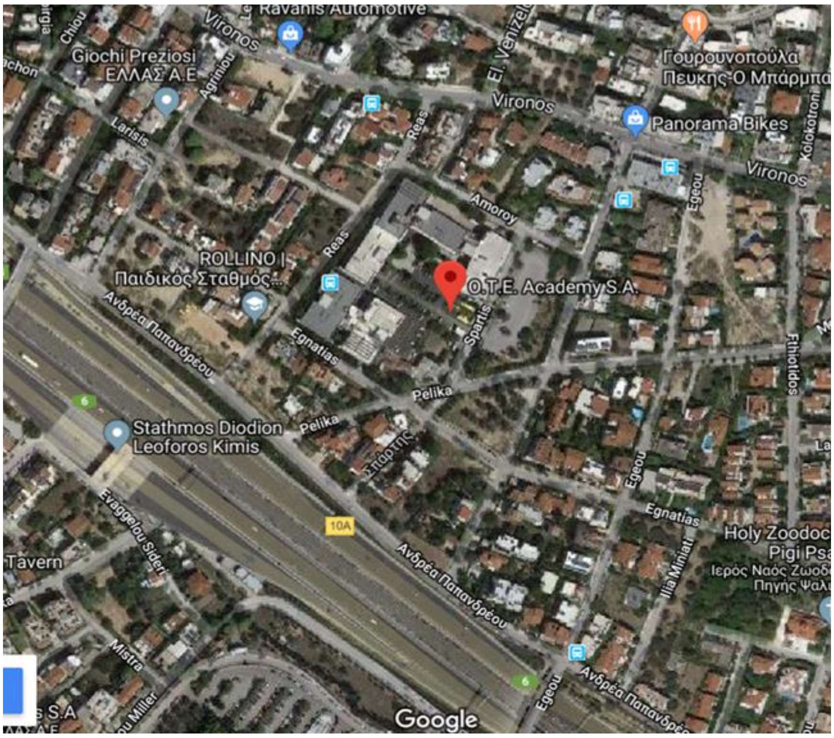

Figure 13. Coverage area in Psalidi (northern Athens, right)

3.3.7 Slice deployment time

Slice deployment time (or Service creation Time) is a very important KPI that roots in the softwarized nature of 5G Networks. By leveraging the ability of the service layer and the enhanced management and orchestration capabilities provided by 5G-TOURS [27] verticals will be able to tailor the requirements imposed to the network slices providing their services and, in turn, operators will be able to deploy slices within a very limited amount of time. Due to the complexity of the system, the slice deployment time is a compound measure of different items that build to the final overall deployment time. Thus, it is relevant to disaggregate this value into different orchestration times. These actions depend on the NFVO and the NFVI used to perform the tests, which may lead to slight variations of the measured times. Hence, in 5G-TOURS, we will measure the slice deployment time, making the following assumptions:

- Network Services (NS) and Network Slice Templates (NSTs) are available in the 5G-TOURS Catalogue ready to be instantiated.
- The NFVO and NFVI used to perform the measurement are per-site hence is expected the execution of the experiment for slice deployment time KPI calculation, in each site.
- The Network Slice Instance will be considered finished when the orchestrator reports the slice in status READY/OPERATIONAL. That means that all Network Slices Subnets are deployed.
- Operations during the development and deployment phase (Day 1) of the network are out of the scope for the slice deployment time KPI measure.

The slice deployment time KPI will follow the formula below:

$$sdt = \sum_{NSVLD=1}^{n} (NSVLDn) + \sum_{NSS=1}^{n} (NSSn)$$

Where:

- sdt is the slice deployment time
- NSS is the Network Slice Subnet
- NSVLD is the Network Slice Virtual Link Descriptor

The slice deployment time is the calculation result of the summation of the time spent by the creation of each the Network Slices VLDs (End-to-end network) plus the time spent by the deployment of each Network Slice Subnet (Network Services). After this, we will have the slice deployed with the network service composition in place.

Measurements will be taken with ad-hoc software running within the service layer and the enhanced management and orchestration modules, which have the full picture of the overall deployed service and the running network slices, respectively.

Phase 1. Onboarding		
Time components	**What**	**Measurement Conditions**
1.01	**Network Slice Template (NEST)**	*Tstart = Request for onboard received in the system Tend = NEST fully onboarded in all the catalogues and available for LCM actions*
1.02	**Network Service Descriptor (NSD)**	*Tstart = Request for onboard received in the system Tend = NSD fully onboarded in all the catalogues and available for LCM actions*
1.03	**VNF package (VNFD)**	*Tstart = Request for onboard received in the system Tend = VND fully onboarded in all the catalogues and available for LCM actions*
1.04	**MEC Application Descriptor (MEC AppD)**	*Tstart = Request for onboard received in the system Tend = APpD fully onboarded in all the catalogues and available for LCM actions*
1.05	**Other applications**	*Tstart = Request for onboard received in the system Tend = Application fully onboarded in all the catalogues and available for LCM actions*
Phase 2. Instantiate, Configure & Activate		
Time components	**What**	**Measurement Conditions**
2.01	**Instantiate Network Slice (NSI)**	Tstart = Request for instantiate received in the system Tend = NSI fully instantiated, 'alive' and functional and available for monitoring actions
2.02	**Instantiate & Activate Network Service (NS)**	Tstart = Request for instantiate received in the system Tend = NS fully instantiated, 'alive' and functional and service performance (QoS) metrics meeting or exceeding the target performance (Ps) of the Network Service. Measure the service performance (QoS) metrics periodically (recommended once every 100 ms)
2.03	**Instantiate & Configure VNFs in service chain (VNF)**	Tstart = Request for instantiate logged in the system Tend = VNF fully instantiated, configuration complete, VNF 'alive' and functional
2.04	**Instantiate & Configure MEC Application (MEC App)**	Tstart = Request for instantiate logged in the system Tend = MEC App fully instantiated, configuration complete, 'alive' and functional
2.05	**Instantiate & Configure other applications**	Tstart = Request for instantiate logged in the system Tend = Other app fully instantiated, configuration complete, 'alive' and functional

Figure 14. Service Deployment Time input form (adapted from [28])

For each subnet, the deployment time could be further split into subtimes. As a matter of fact, NSS deployment time can be defined as the time needed to activate a Network Service that comprises multiple VNFs in a service chain [29]. For the collection of the different timings we will resort to listings similar to the ones already used successfully by the 5GPPP Architecture WG to gather input for the KPI verification [28]. We provide a snippet of such table in Figure 14, in which different subitems are present.

3.3.8 Security

In general, the definition of security KPIs is a difficult task, while the design of evaluation methods for validation of specific security factors is cumbersome, because both the KPI definition and the evaluation methodology highly depends on the objectives of the use cases and the already available capabilities (security functionalities) of the pilot sites. In addition, the use of a single metric for the evaluation of the overall network security is impossible. In 5G-TOURS in order to progress with the task of security evaluation, a set of security metrics were identified, which are related with the 5G-TOURS security targets. The security metrics are presented below. Then, it is the responsibility of each pilot site to select the subset of the security metrics that better reflect: a) the capabilities of the sites in terms of security functions already present and security software already deployed; b) the requirements of the pilot tests cases in a per use case manner.

The set of related security metrics identified are presented below:

1) Incident identification time: How much time (mean time) does it take to identify an incident?
2) Incident response resolution time: How much time (mean time) does it take to respond and resolve an incident?
3) False/Positive percentage identification: Effectiveness to identify false positive and false negative events.
4) Cyberattacks/Threats prevented: How many threats have been successfully blocked from the security mechanisms in place?
5) Data center availability (uptime): For how long has the datacenter been operating without interruption?
6) Number of systems with known vulnerabilities: Can you effectively identify through a security vulnerability management system if there are unpatched systems?
7) Days to patch: How much time does it take to patch vulnerabilities within your network?
8) Number of SSL certificates configured incorrectly: Monitoring the security requirements for each certificate, as well as ensuring that they are properly configured on servers and if they have expired.
9) Number of cybersecurity incidents reported: Are users reporting cybersecurity issues to your team?
10) Intrusion attempts: How many times have bad actors tried to breach your networks?
11) Time to identify devices on the internal network: How much time does it take to identify rogue/new devices within the network?
12) Monitoring of privileged accounts on network devices: How much time does it take to detect the amount of authorized privileged accounts vs rogue privileged accounts
13) Monitoring of authorized changes in network security mechanisms: How much time does it take to detect unauthorized equipment connected to the network.
14) On-Time Removal of Unauthorized Third Party Connectivity: Measures the percentage of Unauthorized Third Party Connections that were removed on time within the Measurement Period.

3.3.9 Location accuracy

5G TOURS will deliver a wide number of use cases with different requirements of position accuracy in different type of locations (indoor, outdoor or mixed) as reported in deliverables D4.1[9], D5.1[10] and D6.1[11].

Table 13: UC positioning accuracy requirements

Node	UC #	Description	Location	Required accuracy [m]	Available position markers
Turin	1	Augmented tourism experience	Indoor/outdoor	0.5	Estimote Proximity Beacons[6]
	2	Telepresence	Indoor	0.1	Estimote Proximity Beacons

[6] Estimote Proximity Beacons: https://estimote.com/products/

	3	Robot-assisted museum guide	Indoor	0.1	Pozyx beacons and creator system[7]
	4	High quality video services distribution	Indoor/outdoor	N/A	
	5	Remote and distributed video production	Indoor/outdoor	N/A	
Rennes	6	Remote health monitoring and emergency notification	Indoor/outdoor	Not clear	
	7	Teleguidance for diagnosis and intervention support	Indoor/outdoor	1	
	8	Wireless operating room	Indoor	N/A	
	9	Optimal ambulance routing	Outdoor	Not clear	
Athens	10	Smart airport parking management	Outdoor	0.5-1	Parking sensors[8]
	11	Video-enhanced ground-based moving vehicles	Outdoor	1-5	
	12	Emergency airport evacuation	Indoor	0.3-1	
	13	Excursion on an AR/VR-enhanced bus	Indoor/Outdoor	1-4	

3GPP set for release 16 [3] a target for commercial use cases for RAT dependent positioning solutions a goal of:

- Horizontal positioning error < 3m for 80% of UEs in indoor deployment scenarios
- Vertical positioning error < 3m for 80% of UEs in indoor deployment scenarios
- Horizontal positioning error < 10m for 80% of UEs in outdoor deployments scenarios
- Vertical positioning error < 3m for 80% of UEs in outdoor deployment scenarios
- Positioning service latency <1s

Improvement of positioning accuracy are on discussion for release 17 [4] to reach targets of:

- Horizontal positioning error <0.3m (absolute, indoor/outdoor)
- Vertical positioning error <2m (absolute, indoor/outdoor)
- Positioning service latency <10ms

For better positioning accuracy 5G shall support the combination of 3GPP and non 3GPP technologies such as Global Navigation Satellite System (GNSS), Terrestrial Beacon Systems (TBS), sensors, etc. In the case of 5G TOURS use cases, several of them exceed requirement of pure 5G network thus they will require a hybrid approach where positioning provided by the network itself is complemented by measurements coming from different sensors and from their known position of sensors (especially for indoor). New localization methods, for instance based on machine learning techniques, have to be developed to achieve accurate, seamless, and robust localization.

[7] Pozyx accurate positioning: https://www.pozyx.io/#products-and-services-menu

[8] Parking occupancy sensors for UC 10 leverages on a custom board designed by a partner, complete description on [11] sec. 3.1.2.3

As general validation method to assess positioning accuracy is to compare the position provided by 5G network with measurements coming GNSS for outdoor (GPS, Galileo, others)[9] locations and with measurements directly taken during site survey (e.g. with a laser telemeter) of several spots in the use case trial scene and reported on maps for indoor sites; number of spots have to be decided during site survey and depends from number and location of radio emitters and other sensors used for hybrid positioning.

Table 14 reports, for each use case, the suggested position accuracy evaluation method; adequate number of measurements must be taken over time to reduce statistical error of the measurement.

Table 14: Position accuracy validation method per UC

Node	UC #	Description	Location	Required accuracy [m]	Validation method
Turin	1	Augmented tourism experience	Indoor/outdoor	0.5	Indoor: known position spots Outdoor: GNSS Galileo HAS
	2	Telepresence	Indoor	0.1	known position spots
	3	Robot-assisted museum guide	Indoor	0.1	known position spots
	4	High quality video services distribution	Indoor/outdoor	N/A	N/A
	5	Remote and distributed video production	Indoor/outdoor	N/A	N/A
Rennes	6	Remote health monitoring and emergency notification	Indoor/outdoor	Not clear	Any GNSS
	7	Teleguidance for diagnosis and intervention support	Indoor/outdoor	1	Any GNSS
	8	Wireless operating room	Indoor	N/A	N/A
	9	Optimal ambulance routing	Outdoor	Not clear	Any GNSS
Athens	10	Smart airport parking management	Outdoor	0.5-1	GNSS Galileo HAS
	11	Video-enhanced ground-based moving vehicles	Outdoor	1-5	GNSS Galileo HAS
	12	Emergency airport evacuation	Indoor	0.3-1	known position spots
	13	Excursion on an AR/VR-enhanced bus	Indoor/Outdoor	1-4	Indoor: known position spots Outdoor: GNSS Galileo HAS

[9] GPS has horizontal position accuracy better of 2 m and vertical better than 4m for 95% of time [5]; Galileo Open Service has similar performances while, when available, High Accuracy Service (HAS) shall have accuracy precision <0.2m. All Galileo services are expected to be available from end of 2020 when all satellites (24) will be positioned and working

4 Conclusions

This deliverable presents the 5G-TOURS evaluation methodology including a general validation methodology accompanied with a set of evaluation procedures presented in a per KPI manner. The general evaluation methodology extends the 5G EVE model-based testing and validation methodology (D5.2 [2]) and further adopts many aspects of the 5G-MoNArch evaluation methodologies and procedures (D6.3 [8]). More specifically, the 5G-TOURS evaluation methodology will go beyond the state of the art, evaluating not only specific KPIs (as done in this first release of the deliverable), but also the level of satisfaction of user and verticals, through e.g., surveys or questionnaires.

The target of the current deliverable is to hence to provide a set of well-defined evaluation procedures that can be used for the definition of the trials, the execution of the use cases, the collection of the appropriate measurements, the analysis of these measurements and finally the evaluation of the analysis results against the predefined KPI targets. This is an initial, needed step to achieve the overall WP7 goals. Specifically, we

- Defined specific questionnaires to gather the requirements from the verticals
- Identified the KPIs for individual use cases, eventually splitting them into well-defined categories
- Described compelling methodologies for each of them.

This deliverable hence provides a baseline for the subsequent steps in the definition and implementation of the evaluation methodology, which will be captured by the following releases of the deliverables of this work package. Namely, we will

- Further specify the presented evaluation procedures to the special characteristics of each use case in order to final conclude to well-defined evaluation procedures in a per KPI and per Use case manner
- Include the Quality of Experience (QoE) as one of the main metrics in the overall evaluation methodology. The first part of this work started with the network KPIs as QoE could also be derived from them, in addition to approaches that purely rely on Mean Opinion Score (MOS) approaches. This will both help to provide information about the fulfilment of service level KPIs and also give additional inputs to the network architectural work, as QoE triggers can be used to perform re-orchestration and re-configuration of the network.
- Specific interactions with other work packages: as the initial evaluation methodology is defined, specific cross work package interactions are envisioned
    - With WP2 on the continuous refinement of the targeted KPIs, also from the QoE point of view
    - With WP3 on the specific network capabilities required by the evaluation methodology. Many of the evaluation metrics rely on standard approaches that have to be fulfilled by the network vendors, while others such as the ones related to coverage, will require tight collaborations with network providers of each test site. Finally, another envisioned includes the interaction with T5.3 and will provide the final evaluation methodology fine-tuned to the network capabilities available in the pilot sites (e.g. MEC, SDN, MANO).
    - With WP 4, 5 and 6 for the actual fulfilment of the envisioned KPI, especially for the efficient mapping of the QoE with application level parameters
    - With WP8 for the tecno-economic analysis of the achieved KPIs.

All the above will be reported in Deliverable D7.2 on M22

Acknowledgment

This project has received funding from the EU H2020 research and innovation programme under Grant Agreement No. 856950.

References

[1] 5G EVE Deliverable D5.1 “Disjoint testing and validation tools” – ICT-17-2018 5G EVE Project – 5/2019 https://doi.org/10.5281/zenodo.3625619

[2] 5G EVE Deliverable D5.2 “Model-based testing framework document” – ICT-17-2018 5G EVE Project –12/2019 https://doi.org/10.5281/zenodo.3628341

[3] 3GPP TR 38.855 V16.0.0 (2019-03) Study on NR positioning support (Release 16)

[4] 3GPP TS 22.261 V17.1.0 (2019-12) Service requirements for the 5G system Stage 1 (Release 17)

[5] Global Positioning System (GPS) Standard Positioning Service (SPS) Performance Analysis Report Report #96 January 31, 2017

[6] European GNSS (GALILEO) Initial Services – Open Service – Quarterly performance report April-June 2019. Link: https://www.gsc-europa.eu/sites/default/files/sites/all/files/Galileo-IS-OS-Quarterly-Performance_Report-Q2-2019.pdf

[7] Availability definition: https://searchnetworking.techtarget.com/definition/availability

[8] 5G MoNArch Deliverable D6.3 “Final report on architectural verification and validation” https://5g-monarch.eu/wp-content/uploads/2019/07/5G-MoNArch_761445_D6.3_Verification_and_validation_final_report_v1.0.pdf

[9] 5G TOURS Deliverable D4.1 “Robotic, Media and Smart Cities solutions for Touristic Cities” http://5gtours.eu/documents/deliverables/D4.1.pdf

[10] 5G TOURS Deliverable D5.1 “5G-enabled solutions for safe cities” http://5gtours.eu/documents/deliverables/D5.1.pdf

[11] 5G TOURS Deliverable D6.1 “5G-enabled solutions for mobility-efficient cities” http://5gtours.eu/documents/deliverables/D6.1.pdf

[12] V. Beal, “webopedia,” 2019. [Online]. Available: https://www.webopedia.com/TERM/L/latency.html

[13] D. H. Wolaver, Phase-Locked Loop Circuit Design, Prentice Hall, 1991.

[14] D. C. Salyers, A. Striegel and C. Poellabauer, “Wireless Reliability: Rethinking 802.11 Packet Loss,” University of Notre Dame, Notre Dame, 2008.

[15] K. Hedayat and R. Krzanowski, “A Two-Way Active Measurement Protocol (TWAMP),” 2008. [Online]. Available: https://tools.ietf.org/html/rfc5357.

[16] M. Lauridsen, L. C. Gimenez, I. Rodriguez, T. B. Sorensen and P. Mogensen, "From LTE to 5G for Connected Mobility," in IEEE Communications Magazine, vol. 55, no. 3, pp. 156-162, March 2017. doi: 10.1109/MCOM.2017.1600778CM

[17] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," in IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3098-3130, Fourthquarter 2018. doi: 10.1109/COMST.2018.2841349

[18] C. A. Garcia-Perez and P. Merino, "Enabling Low Latency Services on LTE Networks," 2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS*W), Augsburg, 2016, pp. 248-255. doi: 10.1109/FAS-W.2016.59

[19] 5G-NORMA project. Website: https://5gnorma.5g-ppp.eu/

[20] METIS-II project. Website:https://metis-ii.5g-ppp.eu/

[21] 5G EVE project. Website: https://www.5g-eve.eu

[22] 5G MoNArch project. Website: https://5g-monarch.eu/

[23] 5G EVE Deliverable D1.4 “KPI collection Framework – ICT-17-2018 5G EVE Project – 6/2019

[24] Ericsson Mobility Report, November 2019 https://doi.org/10.5281/zenodo.3625637

[25] 5G-MoNArch D6.1 Documentation of Requirements and KPIs http://5g-monarch.eu/wp-content/uploads/2017/10/5G-MoNArch_761445_D6.1_Documentation_of_Requirements_and_KPIs_and_Definition_of_Suitable_Evaluation_Criteria_v1.0.pdf

[26] ETSI GS NFV-REL 003, “Network Functions Virtualisation (NFV); Reliability; Report on Models and Features for End-to-End Reliability”, V1.1.1, Apr. 2016.

[27] 5G TOURS Deliverable D3.1 “Baseline architecture and deployment objectives” http://5gtours.eu/documents/deliverables/D3.1.pdf

[28] 5G PPP Architecture Working Group, View on 5G Architecture, Version 3.0, June 2019

[29] ETSI GS NFV-TST 001 V1.1.1 Network Functions Virtualisation (NFV), Pre-deployment Testing, Report on Validation of NFV Environments and Services

[30] 3GPP TS 28.552 V16.3.0, “Technical Specification Group Services and System Aspects, Management and orchestration, 5G performance measurements (Release 16)”, September 2019
[31] 3GPP TS 28.554 V16.2.0, “Technical Specification Group Services and System Aspects, Management and orchestration, 5G end to end Key Performance Indicators (KPI) (Release 16)”, September 2019
[32] ITU-R M.2410, “Minimum requirements related to technical performance for IMT-2020 radio interface(s)”, November 2017
[33] ITU-R M.2412, Guidelines for evaluation of radio interface technologies for IMT-2020”, October 2017
[34] RFC2679, “A One-way Delay Metric for IPPM”, September 1999