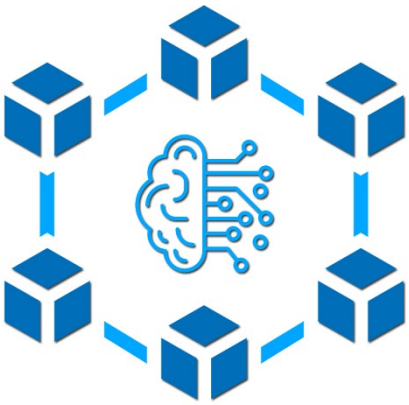

An Artificial Intelligent Aided Unified Network for Secure Beyond 5G Long Term Evolution [GA: 101096456]

Deliverable D6.6

Italian in-lab testbed dataset 1

Programme: HORIZON-JU-SNS-2022-STREAM-A-01-06

Start Date: 01 January 2023

Duration: 36 Months

Co-funded by the European Union

6GSNS

NANCY project has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101096456.

Document Control Page

Deliverable Name	Italian in-lab testbed dataset 1
Deliverable Number	D6.6
Deliverable Type	Type DATA — data sets, microdata, etc
Work Package	WP6
Associated Task	Task 6.7 Italian in-lab testbed
Dissemination Level	Public
Due Date	31 December 2023 (M12)
Completion Date	29 December 2023
Submission Date	30 December 2023
Deliverable Lead Partner	ITL
Deliverable Author(s)	Clavenna Antonella (ITL)
Version	1.0

Document History

Version	Date	Change History	Author(s)	Organisation
0.1	06 November 2023	Initial version	Antonella Clavenna	ITL
0.2	17 November 2023	ToC aligned with the ToC of corresponding D6.5	Antonella Clavenna	ITL
0.3	5 December 2023	First draft for internal review by CRAT and VOS	Antonella Clavenna	ITL
0.4	15 December 2023	Comments by CRAT addressed	Antonella Clavenna	ITL
0.5	18 December 2023	Dataset 1 available for uploading on “IEEE dataport” and “Zenodo”	Marco Beccari	ITL
0.6	19 December 2023	Comments by VOS addressed – version ready for Quality Manager revision	Antonella Clavenna	ITL
0.10	29 December 2023	Comments by QRM and PC addressed	Antonella Clavenna	ITL

Internal Review History

Name	Organisation	Date
Alvise Rigo	VOS	19 December 2023
Emanuele De Santis	CRAT	11 December 2023

Quality Manager Revision

Name	Organisation	Date
Anna Triantafyllou, Dimitrios Pliatsios	UOWM	28 December 2023

Table of Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Explanation
AVC	Advanced Video Coding
B5G	Beyond Fifth-Generation
CPU	Central Processing Unit
CSV	Comma Separated Values
CoMP	Coordinated Multi Point
DX.Y	Deliverable X.Y
DL	Downlink
5G	Fifth-Generation
FDD	Frequency Division Duplex
FR	Frequency Range
GPU	Graphics Processing Unit
HD	High Definition
HLS	HTTP Live Streaming
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
ML	Machine Learning
mmWave	Millimeter Wave
MEC	Multi-access Edge Computing
NR	New Radio
PCAP	Packet Capture
RTMP	Real-Time Messaging Protocol
RTSP	Real-Time Streaming Protocol
RTP	Real-time Transport Protocol
RAN	Radio Access Network
SD	Standard Definition
TDD	Time Division Duplex
UE	User Equipment
UL	Uplink
USB	Universal Serial Bus
VTU	Video streaming and Transcoding Unit
WebRTC	Web Real-Time Communication
WP	Work Package

Executive Summary

Machine learning (ML) techniques are seen to have great potential to be employed to solve telecommunication networks' optimization problems, which range from the design of hardware elements to network self-optimization. D6.6 provides input data for the development of the Blockchain Radio Access Network (RAN) and attack models for the NANCY framework, to model training and model inference functions. The data collected plays the role of ML algorithm-specific data preparation. The dataset contains time series, collected by transmitting video content through the Italtel VTU application, which can convert audio and video streams from one format to another, at multiple encodings schemes, changing resolution, bitrate, and video parameters. The datasets collected are related to the observation of some of the resources involved in the Usage Scenario 1: "Fronthaul network of fixed topology – Direct Connectivity & Coordinated Multi Point (CoMP) connectivity". The variety of collected features ranges from radio front-end metrics to physical server operating system and network function metrics. Moreover, D6.6 describes the testbed deployed and the procedures to support the creation of different types of network traffic starting from Italtel's VTU video streaming application.

1. Introduction

The "Italian in-lab testbed dataset 1" is a collection of labeled Packet Capture (PCAP) files to assist the development of the B-RAN modelling, created by leveraging a video streaming application for the generation of the data flow, in a limited laboratory environment. In this environment, the network conditions are measured and controlled to produce a set of data relevant to the Usage Scenario 1: "Fronthaul network of fixed topology – Direct Connectivity & Coordinated Multi Point (CoMP) connectivity".

PCAP files contain the raw data of network packets, including the headers and payloads of each packet. These files are generated by packet capture tools such as Wireshark, tcpdump, or other network monitoring software.

1.1. Purpose of the Deliverable

The primary objective of Deliverable D6.6 "Italian in-lab testbed dataset 1" is to generate the first dataset to provide adequate experimental data for the development of the B-RAN and attacks models for the NANCY framework. It contains the first dataset of collected network information, conducted in the Italian in-lab testbed. The data collected provide the required experimental data for B-RAN modelling, performance measurement, and theoretical framework validation, to prove the technical maturity and stability of the NANCY's B-RAN concept in the Usage Scenario 1: "Fronthaul network of fixed topology – Direct Connectivity & Coordinated Multi Point (CoMP)".

Data collection enables exploiting prior knowledge of the system, based on mathematical models, as the initialization point from which AI methods start interacting with the environment for further system optimization. The experimental data collected will be used by AI methods to analyse, e.g., the topology, channel assignment, and other parameters of the current wireless network and to define the model of the initial network status, considering the specific characteristics relevant for the Usage Scenario 1.

1.2. Relation to other Deliverables

D6.6, primarily interconnects with D6.4 – "In-lab testbeds definition", from which it takes the requirements and the activities' plan for the Italian testbed; furthermore, it receives input from D2.1 – "NANCY Requirements Analysis" and D3.1 – "NANCY Architecture Design. As a final step, the outcomes of D6.6 – "Italian in-lab testbed dataset 1" will subsequently contribute to Work Package 2 "Usage Scenario and B-RAN Modelling, Network Requirements, and Research Framework".

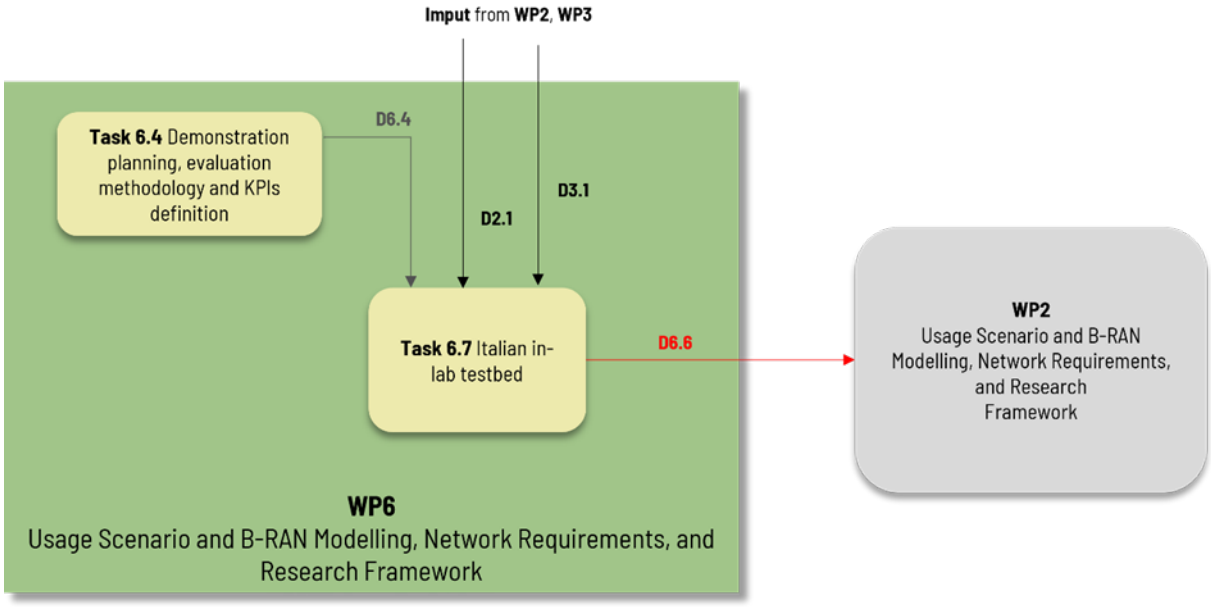

Figure 1: Relation to other deliverables

1.3. Structure of the Deliverable

The deliverable is organized as follows:

- **Section 1 - Introduction:** This section provides a general introduction to the deliverable.
- **Section 2 – Testbed description:** This section presents the overall description, in terms of hardware and software components, of the Italian in-lab testbed, which serves as a baseline reference setup for the data generation procedures.
- **Section 3 – Dataset description:** This section covers the details of the information collected from the network that forms the ”Italian in-lab testbed dataset 1”.
- **Section 4 - Conclusion:** This section concludes the deliverable.

2. Testbed Description

This section presents the Italian in-lab testbed for Usage Scenario 1: “Fronthaul network of fixed topology – Direct Connectivity & Coordinated Multi Point (CoMP)” for B-RAN modelling, performance measurement, and theoretical framework validation. The implementation of the testbed, in an indoor environment, located in Italtel’s premises in Milano, provides regulated and measured experimental conditions, enabling data collection and analysis. The section details the testbed design principles and its implementation. Figure 2 depicts the main functional blocks and overall architecture, while Figure 3 shows the physical testbed. The configuration of the testbed for collecting the first set of data is “Direct connectivity”. D6.8 “Italian in-lab testbed dataset 2” will address CoMP configuration.

2.1. Testbed Topology

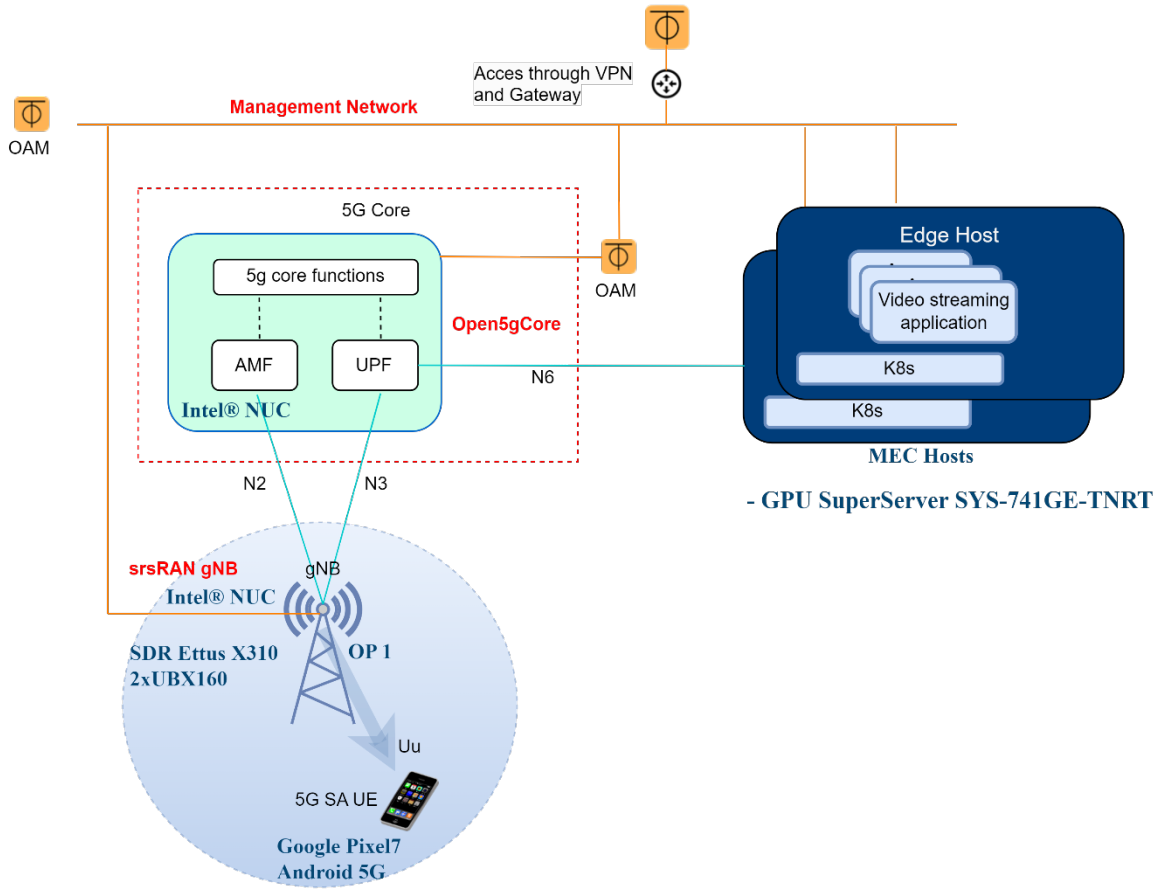

Figure 2: Italian In-lab Testbed topology and configuration

2.2. Hardware Components

As depicted in Figure 3, for the purpose of collecting the first set of data “Data set 1”, the Italian in-lab testbed consists of:

- a **Multi-access Edge Computing (MEC) Host**, an **Edge server integrating a GPU (NVIDIA P4),** as HW accelerator to increase performance (portable)
- a **Graphics Processing Unit (GPU) SuperServer SYS-741GE-TNRT** + 2 **GPU NVIDIA A40 (Supermicro®**)

- **Ettus X310** 2xUBX160 board (USRP X310 High-Performance Software Defined Radio - Ettus Research);
- **Fifth-generation (5G) terminal with eSIM**
- an **open 5G RAN system**, using open-source SW (5g-srsRAN Project - Open source 4G/5G software suites developed by Software Radio Systems, running on an **Intel® NUC**, and Open5GCore running on an intel-based server)

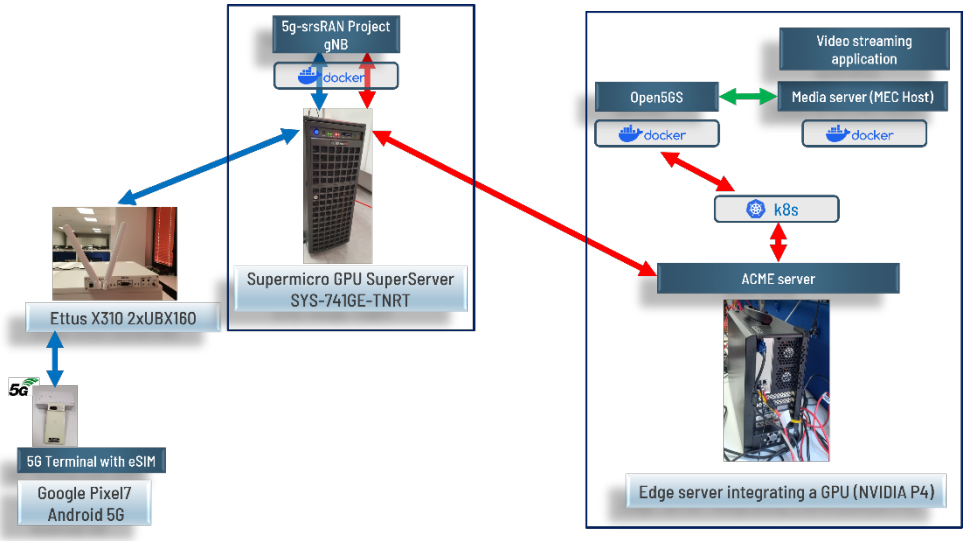

Figure 3: Italian In-lab Testbed equipment

The testbed hosts 1 user equipment (UE), attached to a virtualized gNB instance running on a shared computing platform. The UE consists of a radio head and a set of dedicated computing resources provided by a Supermicro®. Such resources host the complete radio protocol stack and processes from heterogeneous mobile applications. Both UE and gNB use a USRP x310 board as a radio head, and the srsRAN software to implement the radio protocol stack. The gNB USRP board is attached to the computing pool via a Universal Serial Bus (USB) 3.0 connector, while the srsRAN gNB runs as containerized software instances using Docker. The UE is connected to the gNB, emulating the traffic volume generated over a cell. The testbed supports a maximum of 2 UEs and 2 gNBs.

2.3. Software Components

To generate the downstream video stream from the Edge node, “VTU video streaming and transcoding application” developed by Italtel was selected.

To watch the video on the UE, the VLC media player application for the Android™ platform was selected.

To generate the upstream video stream from the UE, “Simple HTTP server” Android™ app was selected, an application that will allow running a local lightweight hypertext transfer protocol (HTTP) server with static content. It is possible to select any folder and it will become available on the local network via the HTTP protocol.

The recorded video selected for both scenarios is “Big Buck Bunny” (link for downloading: "BigBuckBunny.mp4").

To capture network information and to generate the related data files in the Italian in-lab testbed, the packet sniffers used are, respectively, PCAPdroid on the UE, srsRAN on the gNB (ref. to srsran documentation) and tcpdump on the edge host.

The srsRAN Project gNB can output PCAPs at the following layers:

- MAC
- NGAP
- GTP-U
- E1AP
- F1AP
- E2AP

3. Dataset Description

In this section, there is a description of the organization and information of the datasets collected in the Italian in-lab testbed. Different types of files are available: ".pcap", ".csv", ".json" and ".log".

The datasets are available at IEEE DataPort:

- **Permalink:** http://ieee-dataport.org/documents/nancy-sns-ju-project-italian-lab-testbed-dataset-1-d66
- **DOI Link:** https://dx.doi.org/10.21227/y6ah-j205
- **Short Link:** http://ieee-dataport.org/12093

The dataset consists of approximately 36 GB of raw network traffic (in the PCAP file format) and extracted flow-based data (having a set of different features in PCAP, CSV, JSON, and log format) captured in separate files.

Packet data is recorded primarily in files with the ".pcap" file extension and can be used to find performance problems and cyberattacks on the network.

Each file captured is associated with a 10-minute video streaming of the "Big Buck Bunny" video. This video was transmitted on two different bands, N3 and N78, with different resolutions, 480p, 720p, and 1080p; both in uplink (UL) and in downlink (DL); the type of protocol monitored is "HTTP protocol"; in case of N78 band, data related to the resource usage were also captured, for a total of 123 data files.

These data are relevant to assess the scenarios associated with the Italian indoor testbed before and without considering the impact of NANCY architecture and components.

3.1. Scenarios

In the Italian in-lab testbed, a MEC-assisted 5G network scenario with a video streaming application for generating traffic is provided. The recorded video selected for both scenarios is "Big Buck Bunny" (link for downloading: "BigBuckBunny.mp4").

The streaming was made taking into account:

- H.264 and VP8 encodings;
- different bit rate (3Mbps, 6Mbps, 10 Mbps);
- different resolutions (480p, 720p, 1080p);
- different bands (N3 and N78).

Finally, two different scenarios were set up, related to downstream and upstream video flows, as depicted respectively in the two figures below:

Downstream

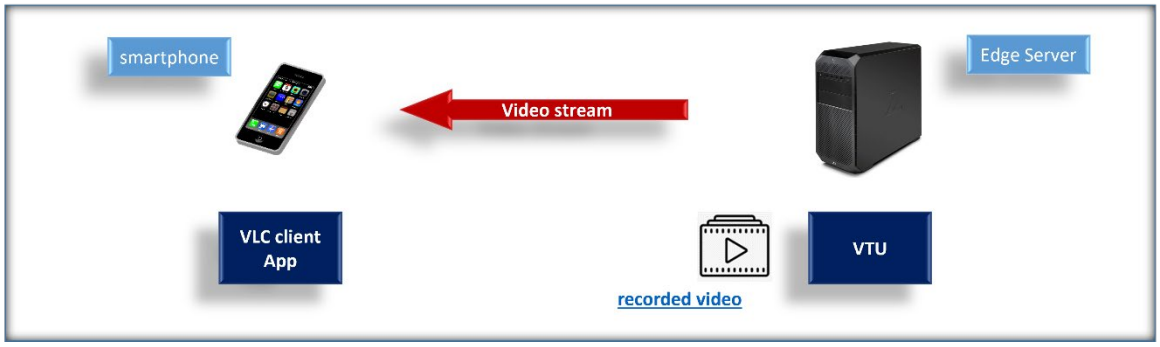

Figure 4: Video streaming – downstream scenario

Upstream

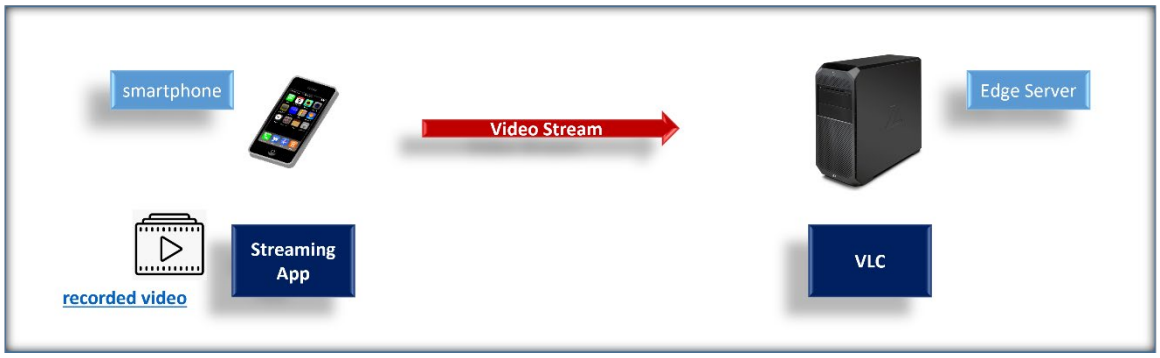

Figure 5: Video streaming – upstream scenario

For generating the downstream video stream, the VTU (video streaming and transcoding) application is used. The VTU is a video streaming and transcoding application developed by Italtel; it can convert audio and video streams from one format to another changing resolution, bitrate, and video parameters. The source stream can originate from a file within the local storage system, or maybe a packetized network stream. The requested transcoding service can be monodirectional, as in video streaming, or bi-directional, like in videoconferencing. The transcoding capabilities of the VTU are provided by Libav (https://github.com/libav/libav). Libav is an open-source library, which can handle a wide variety of audio and video coding standards. For the most computationally intensive video encoding tasks, the VTU can rely on GPU resources.

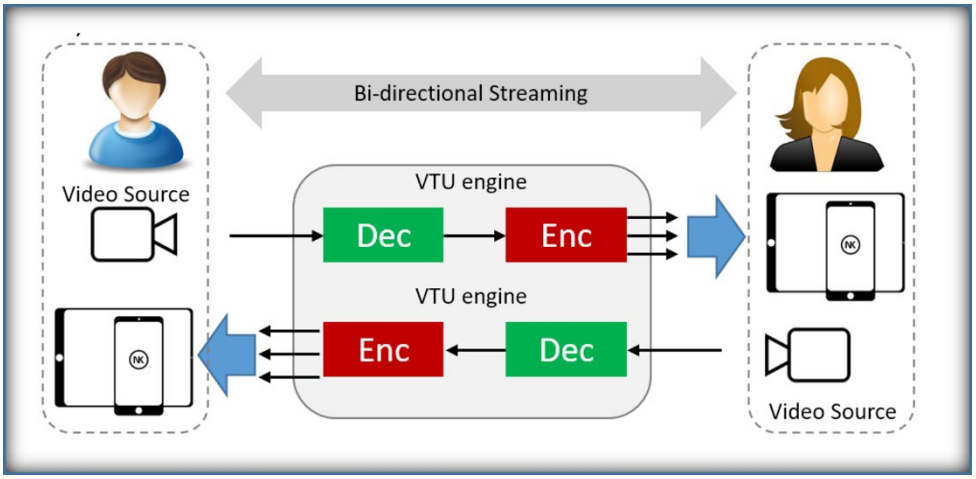

Figure 6: Italtel VTU - Video Streaming and Transcoding - application

The main features of VTU are the following:

- Real-time video streaming management from different sources
- Video processing: transcoding, transrating, transizing
- Capability to share video among sources/users/applications
- Protocol support: Real-Time Messaging Protocol (RTMP), Real-Time Streaming Protocol (RTSP), HTTP, HTTP Live Streaming (HLS), Real-time Transport Protocol (RTP), and Websocket.

For generating the upstream video stream, the "Simple HTTP Server Android Application" was selected, an application that allows running a lightweight local HTTP server with static content. Considering the reference testbed scenario depicted below, the data were collected at three different points of the network, namely the UE, gNB, and edge host (i.e., point A, point B, and point E respectively in Figure 8 below).

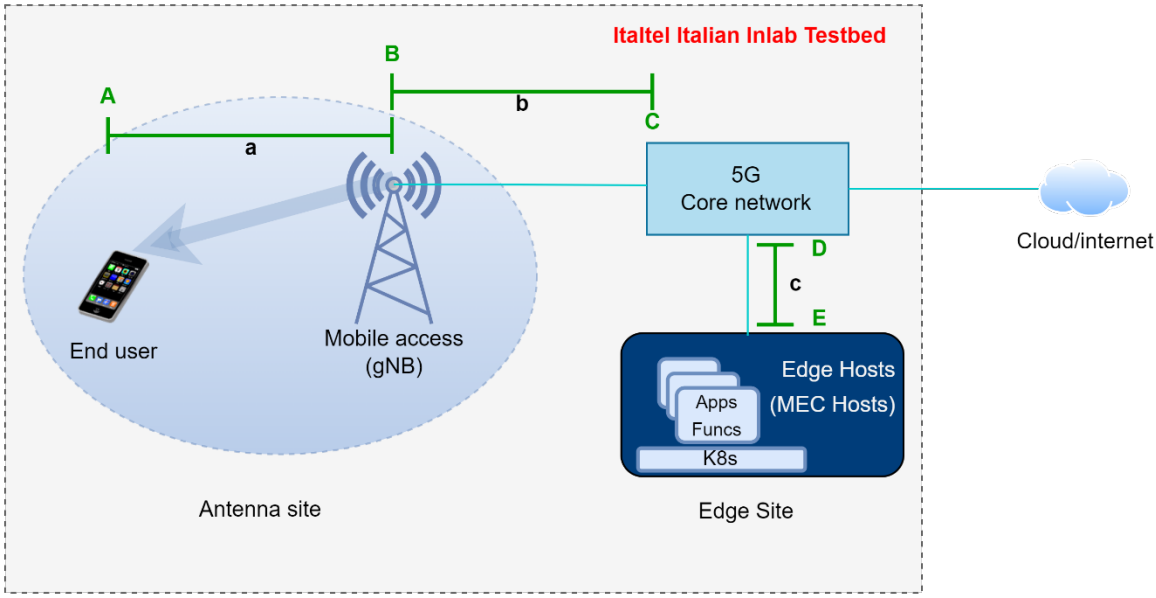

Figure 7: ITL Italian In-Lab Testbed for data set 1 generation

Considering the downstream scenario, the packet sniffer software selected for traffic measurement and data set generation on the UE (ref. Figure 8 - point A) is PCAPdroid; it was installed in the terminal through Google Play. Traffic was measured on one stationary terminal (Google Pixel7).

The packet sniffer software selected for traffic measurement and data set generation on the gNB (ref. Figure 8 - point B) is srsRAN while the one selected for the edge host (ref. Figure 8 - point E) is tcpdump.

The collected dataset is representative resource-intensive video traffic that has the greatest impact on 5G/B5G network planning and provisioning. The video streaming dataset includes data directly measured while watching the video on mobile devices and data directly measured while generating downstream video stream traversing the gNB (i.e., downstream scenario), and vice versa (i.e., upstream scenario). In each experiment, we fixed the location of the UE and the gNB.

3.2. Captured Data

A (network) flow (flow-ID) is identified by the 5-tuple: Source IP, Destination IP, Source Port, Destination Port, and Protocol. Each file captured is associated with a 10-minute video streaming of the “Big Buck Bunny” video. This video was transmitted on two different bands, N3 and N78, with different resolutions, 480p, 720p, 1080p; both in UL and in DL; the type of protocol monitored is “HTTP protocol”; in the case of N78 band, data related to the resource usage were also captured using *vmstat*, for a total of 123 data files.

All the collected data are relevant to assess the scenarios associated with the Italian indoor testbed before and without considering the impact of NANCY architecture and components.

3.3. Dataset Folder Structure

The data collected are organized in folders. Starting from the upper level, the dataset consists of the following folders:

- the folder “Movies_Sample” contains the “Big Buck Bunny” video, used for the test, at different resolutions
- the folder “Pixel7_android_v13” contains the datasets collected

The folder “Pixel7_android_v13” is organized into two main subfolders, “Band_N3_10Mhz” and “Band_N78_10Mhz”, respectively associated with the data captured using bands N3 and N78.

Similarly, each “Band_Nx_10Mhz” folder is organized in as many subfolders as the resolution types (i.e. 480p, 720p,1080p).
Each one of these latter similarly contains a subfolder “HTTP_protocol” organized in two subfolders: “download” and “upload”. The “download” folder contains all the files related to the data captured during the “downstream” tests, while the “upload” folder contains all the files related to the data captured during the “upstream” tests.

Finally, each file present in these folders and named:

<capture point>_<resolution>_<video stream direction>.<type>

follows the naming convention described in the tables below.

Table 1: “capture point” naming convention

capture point	Description	Notes
gnb_rlc_metrics	Data collected at the gNB, related to the Radio Link Control (RLC) layer	Refer to srsran documentation
gnb_vmstat	Data collected at the gNB using vmstat and related to the resource usage	
gnb_trace	Contains the config and the trace of the gNB	Refer to srsran documentation
gnb_ngap	NG application protocol (NGAP) 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
gnb_mac	MAC protocol 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
gnb_gtpu	GTP-U 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
gnb_e1ap	E1AP 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
gnb_f1ap	F1AP 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
gnb_bitrate	Bitrate at the gNB	
gnb_e2ap	E2AP 5G traffic data collected at gNB through srsRAN	Refer to srsran documentation
EdgeServer	Data collected at the edge server	
EdgeClient	Data received by the edge client (generated by UE)	The Data coming out from the UE cannot be captured by PCAPdroid on the UE

Table 2: “resolution” naming convention

Resolution	Description	Notes
480p	Standard definition (SD) resolution	
720p	High Definition (HD) resolution	
1080p	Full HD resolution	

Table 3: “video stream direction” naming convention

Video Stream Direction	Description	Notes
UL	Uploading direction – from UE to Edge host	
DL	Downloading direction – from Edge host to UE	

Table 4: “type” naming convention

Type of File	Description	Notes
pcap	PCAP - Packet capture files - are a common format for storing packet captures.	
csv	Comma-separated Value (CSV) file allows data to be saved in a tabular format.	
json	JavaScript Object Notation (JSON) is an open standard file format for sharing data that uses human-readable text to store and transmit data.	
log	Log files are a historical record of everything and anything that happens within a system.	

4. Conclusion

D6.6 provides a first set of collected parameters (Dataset 1) relevant to assessing the scenarios associated with the Italian indoor testbed before and without considering the impact of NANCY architecture and components.

The data were collected, exploiting the Italian in-lab testbed where a MEC-assisted 5G network scenario with a video streaming application for generating traffic was set up.

Bibliography

[1] L. F. Sikos, «Packet analysis for network forensics: A comprehensive survey,» *Forensic Science International: Digital Investigation,* vol. 32, March 2020.