ICT-777137

5G-RANGE

5G-RANGE: Remote Area Access Network for the 5th Generation

Research and Innovation Action
H2020-EUB-2017 – EU-BRAZIL Joint Call

D5.2 Final Version of the Network-level Architecture and Procedures

Due date of deliverable: 31st December 2019
Actual submission date: 10th February 2020

Start date of project: 1 November 2017 Duration: 30 months
Project website: http://5g-range.eu
Lead contractor for this deliverable: UC3M / UnB
Version 5 date 30st June 2020
Confidentiality status: Public

5G-Range receives funding from the European Union Horizon 2020 Programme (H2020/2017-2019) under grant agreement n° 777137 and from the Ministry of Science, Technology and Innovation of Brazil through *Rede Nacional de Ensino e Pesquisa* (RNP) under the 4th EU-BR Coordinated Call Information and Communication Technologies.

Abstract

This document includes the first version of the network level architecture of the 5G-RANGE project. Up to the date of this deliverable, an architectural deliverable has already been produced in the project in work package 2 (public deliverable 2.2) and although it has already introduced some ideas about the network level architecture, it has been mainly focused on the physical and MAC layers. This document will also comment the work done in work package 2 and will complete the network layer work including the activity developed in work package 5 to provide the first version of the network architecture. This architecture will be completed in deliverable D5.2 providing the final version of the network level architecture.

Target audience

This document is particularly focused on the 3GPP standard for 5G access networks and requires a background in related technologies to be able to fully understand the contribution.

Impressum

Full project title:	5G-RANGE: Remote Area Access Network for the 5th Generation
Document title:	D5.2 Final Version of the Network-level Architecture and Procedures
Editor:	Priscila Solis (UnB)
Work Package No. and Title:	WP5, Network Layer
Project Co-ordinator:	Marcelo Bagnulo, UC3M (EU), Priscila Solis, UnB (BR)
Technical Manager:	Luciano Mendes, Inatel (BR)

Executive Summary

This document provides the final version of the network level architecture for the 5G-RANGE project. As described in the previous deliverable, D5.1, the proposed architecture is aligned with the System Architecture for the 5G System as defined in the 3GPP specification TS 23.501 version 16.3.0 (System Architecture for the 5G System, December 2019) and the 3GPP companion specification TS 23.502 version 16.3.0 (Procedures for the 5G System, December 2019). The architecture includes as its main pillars the description of the interaction with the 5G core network, the interaction with the IMS core and finally the interaction with the management and orchestration Network Function Virtualization (NFV) infrastructure that would assist the whole platform.

In this deliverable, D5.2 – "Final Version of the Network-level Architecture and Procedures", the architecture is updated with the following 3GPP technical specifications: "System architecture for the 5G system; stage 2," TS 23.501, version 16.3.0 (December 2019), "Procedures for the 5G system; stage 2,", TS 23.502, version 16.3.0 (December 2019) and "IP Multimedia Subsystem (IMS); Stage 2", TS 23.228, version 16.3.0 (September 2019). Then, this deliverable shows how the network layer of 5G-RANGE is designed to provide end user terminals with secure end-to-end Internet Protocol (IP) network connectivity, considering and complementing the solutions adopted at the lower levels.

The architecture described in this document supports the different use cases defined in deliverable D2.1, Applications and Requirements Report, for the project (agribusiness and smart farming for remote areas, voice and data connectivity over long distances for remote areas, wireless backhaul and local high-quality connections and remote healthcare for remote areas) and obviously the different network services associated to these use cases as described in D2.1: Mobile Broadband (MBB), Machine Type Communications (MTC) and Voice over IP (VoIP)

The document also provides an extension for the 5G-RANGE architecture based on the cost-effective deployment of programmable Unmanned Aerial Vehicles (UAV) that will enable a bigger coverage of the 5G-RANGE access network for certain areas. It will be fully integrated with the rest of the architecture since the 5G-RANGE Network Function Virtualization orchestrator will also be responsible for this extension. Also, the main protocols that will be executed by the different network entities are detailed (some of them will be later on selected and demonstrated in the Proof of Concept provided in WP6), including the development of a specific proposal for resource management and traffic classes.

Considering the interaction of the 5G-Range with the management and orchestration Network Function Virtualization (NFV) infrastructure, the document also provides a study of control traffic Voice over IP (VoIP), that may be useful to consider future extensions and applications of 5G-RANGE.
This document also describes the resource management mechanisms incorporated in the architecture to enable the support of flows with different quality of service characteristics, as required by the different 5G-RANGE use cases.

List of Authors

Iván Vidal (UC3M)
Francisco Valera (UC3M)
Marcelo Bagnulo (UC3M)
Marcos Caetano (UnB)
Priscila Solis (UnB)
Cristoffer Leite (UnB)

Table of contents

List of figures

Definitions and abbreviations

3GPP	3rd Generation Partnership Project
5GC	5G Core Network
5GPPP	5G Public-Private Partnership
ACK	Acknowledgement
AMF	Access and Mobility Management Function
AMQP	Advanced Message Queuing Protocol
BW	Bandwidth
CSCF	Call Session Control Functions
CSS	Chirp Spread Spectrum
DDS	Data Distribution Service
DHCP	Dynamic Host Configuration Protocol
EPC	Evolved Packet Core Nework
ETSI	European Telecommunications Standards Institute
FDMA	Frequency Division Multiple Access
GPS	Global Positioning Systems
HGW	Home Gateway
HTTP	Hypertext Transfer Protocol
I-CSCF	Interrogating-CSCF
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
IMS	IP Multimedia Subsystem
IoT	Internet of Things
IP	Internet Protocol
IPSec	IP Security Protocol
LAN	Local Area Network
LEDBAT	Less than best effort/scavenger traffic
LoRA	Long Range
LPWA	Lower-Power Wide-Area
LTE	Long Term Evolution
MAC	Medium Access Control
MANET	Mobile Ad Hoc Network
MANO	Management and Orchestration
MBB	Mobile Broadband
N3IWF	Non-3GPP Inter-Working Function
NAS	Non-Access-Stratum
NFV	Network Functions Virtualization
NFVI	NFV Infrastructure
NFVO	NFV Orchestrator
NR	New Radio
NS	Network Service
NSA	Non-Standalone
PFC	Policy Control Function
PoC	Proof of Concept
QoS	Quality of Service
QUIC	Quick UDP Internet Connections
RAN	Radio Access Network
RCVWND	Receive Window

RTP	Real Time Protocol
RTT	Round Trip Time
SA	Stand Alone
S-CSCF	Serving-CSCF
SIP	Session Initiation Protocol
SMF	Session Management Function
SUAV	Small Unmanned Aerial Vehicles
TCP	Transmission Control Protocol
TDMA	Time Division Multiple Access
TLS	Transport Layer Security
TS	Technical Specification
TSG	Technical Specification Groups
TSG AS	Technical Specification Groups Service and System Aspects
TSG CT	Technical Specification Groups Core Terminals
TSG RAN	Technical Specification Groups Radio Access Network
UDP	User Datagram Protocol
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
UNB	Ultra Narrow Band
UPF	User Plane Function
V2X	Vehicle-to-everything
VIM	Virtual Infrastructure Manager
VNF	Virtual Network Function
VNFM	VNF Manager
VoIP	Voice over IP
VoLTE	Voice over LTE

1 Introduction

1.1 3GPP specification roadmap

In March of 2017, the 3GPP Technical Specifications Groups (TSG) started the Release 15 of the 3GPP 5G specifications with the purpose of delivering the first full set of 5G standards: the New Radio (NR) Access Technology. The 3GPP TSGs involved in this Release 15 are the TSG CT (core network and terminals), the TSG RAN (Radio Access Network, responsible for 5G New Radio), and the TSG SA (Service and System Aspects). This work was split into three different phases [1]:

- **Phase 1**, Early Release 15 which is focused on the non-standalone New Radio (NSA NR) and that was considered as the first migration step of adding NR base stations (called gNB) to an LTE-Advanced system of LTE base stations (eNB) and an evolved packet core network (EPC). In this option there is no 5G core network (5GC). This is the architecture option 3 (see Figure 1).
- **Phase 2**, Regular Release 15 focused on the standalone NR architecture which would be a network of NR base stations (gNB) connected to the 5G core network (5GC) without any LTE involvement. This is the architecture option 2.
- **Phase 3**, final Late Release 15 focused on architecture option 4 (adding an LTE base station to an SA NR network where the control plane is handled via the NR base station) and architecture option 7 (adding an LTE base station to an SA NR network where the control plane is handled via the LTE base station) plus NR-NR Dual Connectivity.

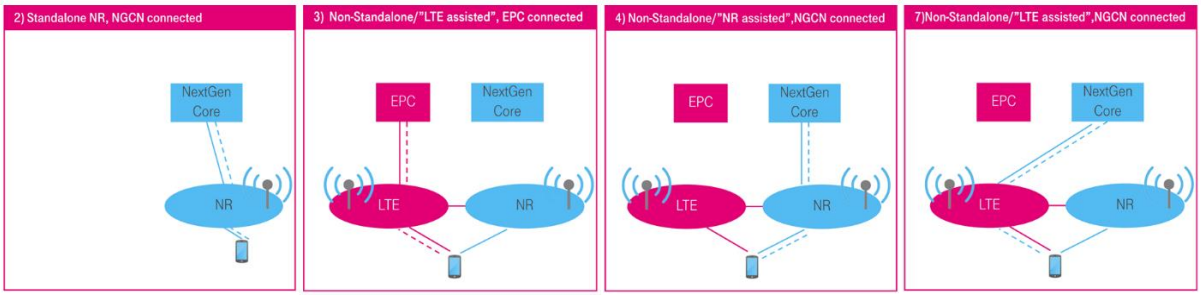

Figure 1. Main architecture options covered by Release 15 [2]

After the release of the 5G New Radio specification for non-standalone (NSA) operation in December 2017 (phase 1), the 3GPP approved in June 2018 the standalone (SA) Release 15 of the 5G specification (phase 2) and by March 2019 completed phase 3 (Late Release) including different companion documents (check [3] and [4]).

During 2019 the 3GPP has been working on the study phase for Release 16 which is the completion of 5G Phase 2, bringing in massive Machine Type Communication (MTC), ultra-reliable and low latency features in the following areas: Multimedia Priority Service, Vehicle-to-everything (V2X) application layer services, 5G satellite access, Local Area Network (LAN) support in 5G, wireless and wireline convergence for 5G, terminal positioning and location, communications in vertical domains and network automation and novel radio techniques, codecs and streaming services, Local Area Network interworking, network slicing and the Internet of Things (IoT).

As a result of this study phase, different Technical Reports have been developed (see [3]). Additionally, different Technical Specifications have been updated (some of them of particular relevance for this document due to their impact in the network architecture, [5], [6] and [7]).

The final 3GPP Release 16 will be completed by June 2020 (previously Release 16 functional specification freeze will happen by March 2020). The whole time schedule can be seen in Figure 2.

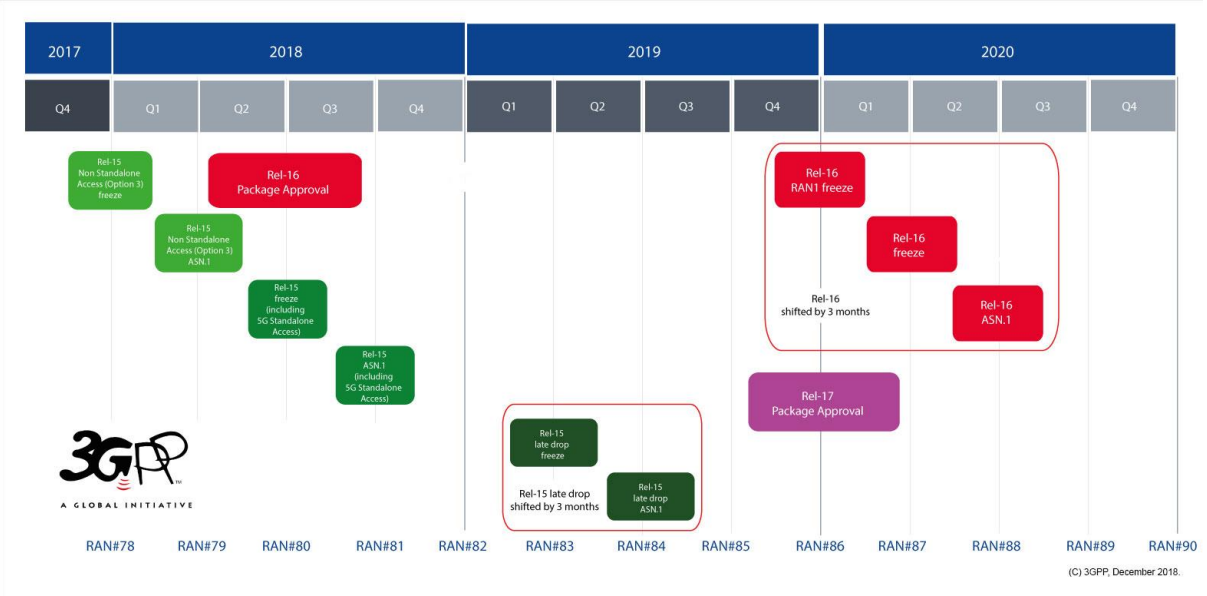

Figure 2. Release 15 & 16 NR milestones [4]

Finally the work in Release 17 has already started and different studies have already been published including new requirements from the verticals, new verticals (like satellite), fixed Broadband Access, requirements for Mission Critical Services [3]. In addition, there has been a notorious interest in 3GPP and a lot of activity in the working groups to ensure that the 5G system will meet the connectivity needs of Unmanned Aerial Systems (UAS), which are of particular interest for this Work Package 5. These documents for the moment are only studies and works in progress ([8] and [9]) but it is important for the project to see that UAS have been considered in 3GPP. A specific portal has been enabled in 3GPP to support this activity [10].

1.2 5G-RANGE Network level architecture

This document provides the reference network level architecture for the 5G-RANGE project. This architecture is aligned with the System Architecture for the 5G System as defined in the 3GPP specification TS 23.501 version 16.3.0 (System Architecture for the 5G System; Stage 2, December 2019, [5]) and the 3GPP companion specification TS 23.502 version 16.3.0 (Procedures for the 5G System; Stage 2, December 2019, [6]) which corresponds to the Regular Release 16.

This deliverable complements deliverable D2.2 (Architecture, system and interface definitions of a 5G for Remote Area Network) [16] and updates D5.1 0 providing the architecture definition for a 5G core network with a untrusted non-3GPP access network (following the specification TS 23.501).

1.3 Deliverable structure

The structure of the document can be summarised as follows:

- **Section 1** contains the introduction, objective and structure of this deliverable;
- **Section 2** presents the main objectives of the architecture and the design requirements;
- **Section 3** provides the definition of the 5G-RANGE network level architectures, that has been defined around three main axes: the integration with operator networks, the support of operator specific services and the virtualization of network level functions
- **Section 4** includes the definition of the transport and network signaling procedures, including the control plane and user plane protocol stacks but also the specific protocols defined at the end user device.
- **Section 5** describes the extensions that have been defined to the 5G-RANGE network level architecture;
- **Section 6** provides the conclusions and finishes the deliverable;

2 Design objective and use cases

2.1 Objective

The main goal of this 5G-RANGE project consists in designing and implementing a Remote Area Access Network for 5G that can be used to provide reliable and cost effective Internet access in low populated areas.
In this context, while 5G-RANGE physical and Medium Access Control (MAC) layers will increase the communication range of 5G networks (current 5G state-of-the-art has mainly focused on improving the data rate or increasing the number of connections and decrease latency), the network layer will be responsible for the end-to-end transport of information of the whole project.
In particular, **the focus of this layer is set on the establishment of IP (Internet Protocol) network connectivity, mobility management, network security and interoperation with the public IP network to support access to Internet services**.
The network-level architecture explained in this document will provide the general overview of the connectivity requirements (in terms of architectural blocks and protocols) needed by a potential rural operator (see Figure 3) in order to be able to enable its radio access network for global connectivity through telecommunication operators.
To maintain the 5G related Key Performance Indicators of the access network (see [16] and [16]) it is mandatory to use a 5G core network that is capable of supporting these parameters. The architecture that is explained in this document is therefore assuming a direct connection of the access network to the operator network through a 5G core (option 2 in Figure 1).

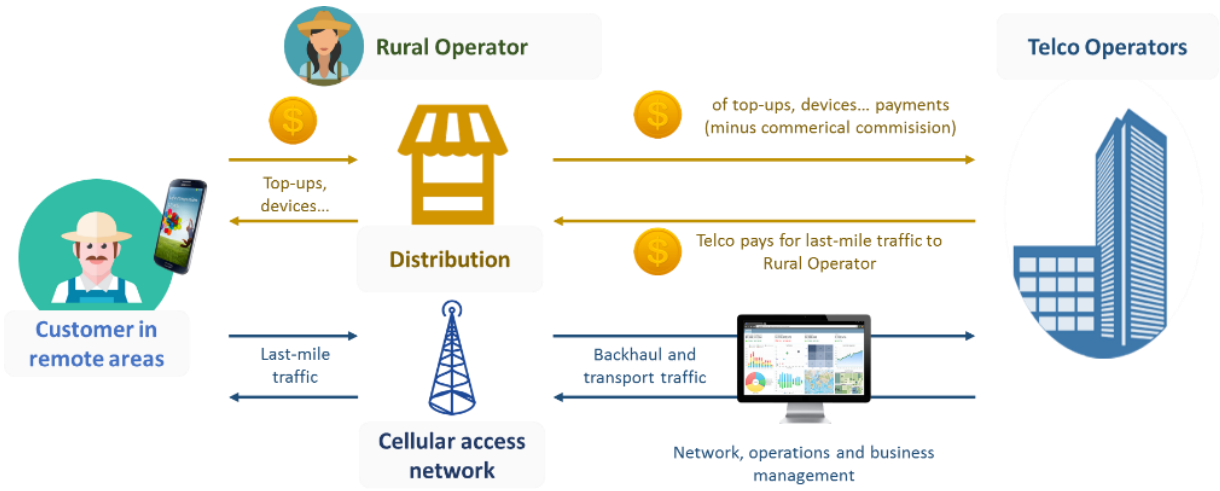

Figure 3. Business model as initially foreseen in 5G-RANGE project proposal

2.2 Use cases

2.2.1 5G-RANGE core use cases

Deliverable 2.1 [16] has introduced different examples and fields of application for the 5G-RANGE Radio Access Network and in particular presents four core uses cases that are mainly considered by the project:

- Agribusiness and Smart Farming for Remote Areas
- Voice and Data connectivity over long distances for Remote Areas
- Wireless Backhaul and Local High-Quality Connections
- Remote Health Care (e-Health) for remote areas

For the sake of completeness a summary of these use cases is included here (for further details and requirements of the use cases, please refer to Deliverable 2.1 [16]).

2.2.1.1 Agribusiness and Smart Farming for Remote Areas

This use case scenario addresses the agribusiness vertical market with the objective to provide reliable connectivity and networking for underserved and remote rural areas. It intends to enable smart farming and broadband Internet access in a sustainable and cost-effective way. Moreover, it deals with real-time services (not necessarily low latency) such as data collection and analysis, crop monitoring, production traceability, remote maintenance and diagnosis, cattle counting, etc.

This use case could be viewed as a set of aggregated mobile sensors monitored by a single user equipment module that would work as an access gateway. Agriculture vehicles like harvesters, tractors, and trucks would be the target for these modules. In fact, much of this equipment, provided by specialized manufactures like Case IH and John Deer, already have built-in embedded sensors for real-time data collection. However, they usually lack wireless connectivity to transmit this information to a data centre for real-time analysis and monitoring. Additionally, such gateways could also support video surveillance in cases where it is necessary.

Obviously, this specific mobile scenario has no limitations regarding battery usage since power can be supplied directly by the vehicles. However, the same cannot be said about the stationary scenario, where sensors are deployed directly into the field. Because of the need to transmit data over long distances, improved battery capacity and/or external energy harvesting techniques, like solar panels, may be required. The agribusiness scenario addresses this issue by using gateways that can be easily installed in the vehicles. These gateways can then provide the required connectivity for a large set of sensors that will be typically embedded in the vehicles themselves.
In addition, other services like crops monitoring or automation for flows (for example irrigation systems) located in very large areas, isolated and difficult to access can be considered.

2.2.1.2 Voice and Data connectivity over long distances for Remote Areas

This use case is focused on providing access to typical Internet applications in very large areas (underserved / rural far remote) with extreme coverage requirements and low density of users. The users may be humans and machines (e.g. wilderness, farms, areas where only highways are

located, etc.) and will have limited availability of Internet broadband access where traditional network deployments are not economically feasible.

This use case could be located not only in Brazil's countryside, but also in wealthier, developed countries in Europe, where connectivity continues to vary between various members of the European Union, with many rural regions still stuck in the slow digital line. The Nordics and Northwest Europe enjoy a high level of access to high-speed Internet while some countries in Southeast Europe lags in the slow lane.
Some types of services that will be evaluated in this use case are enhanced web browsing, email, VoIP, multimedia on the web, audio/video conference, file sharing and interactive video on demand.

2.2.1.3 Wireless Backhaul and Local High-Quality Connections

This use case focuses on the usage of TV broadcast network infrastructure (towers and frequency channels) for wireless backhaul implementation for rural area network. Such regions that currently have a relatively good TV-coverage, the use of that infrastructure for backhaul implementation could be a cost-effective solution. By utilizing low frequencies (VHF and UHF), large multi-antenna systems, beamforming and high TV-towers, enough long wireless backhaul link distances and required capacity may be achieved.

The proposed wireless backhaul approach could be useful for diverse types of scenarios for remote rural locations, which do not have existing (fixed) Internet and cellular network connection or which have connections with only very limited data rates. Here we propose that wireless backhaul could be useful especially for local rural places like tourist venues, schools, industrial or farming premises, remote settlements (villages) and industrial areas which require high-quality connections e.g., mines located in far remote areas, in countries with large areas in which the low population density make a cost-effective solution the 5G-RANGE technology (North European countries in artic Lapland has mean of 2 people for every square kilometre).

The assumption is that a 5G-RANGE base station that is installed to TV tower, can provide line-of-sight (LOS) for a 50 km link using VHF or UHF band (unoccupied channels based on spectrum sensing reports) to the local small cell BS which is located at the rural location. This link works in a transparent way to connect the small cells with the network core. Mobile users are connected to small cell BS in this local rural area. A further assumption is that the population density is low in this area, but the user at small cell BS coverage (< 500 m) must be supported by a high throughput (100 Mbps) connection.

It should be noticed that, in this use case, the 5G-RANGE user would be each one of the small cell base station. It has been estimated that the small cell density will be much lower (small cells/km2) than the normal user density in underserved areas (2 users/km2) due to the difference in the concept, in which the 5G-RANGE UE is a small cell base station. For this reason, it will be possible to give high user throughput.

2.2.1.4 Remote Health Care (e-Health) for remote areas

Health care is an important social service that is usually underserved in remote and rural areas. People living in places far away from urban centers do not have access to specialist physicians. Those who need specific treatments usually are required to travel to an urban area for consults, exams and other interventions. 5G networks will play an important role in changing this

scenario for remote areas. The advent of e-health, based on the integration of IoT and mobile communications for health care, will benefit those living in areas were medical assistance is deficient. E-health will allow patient to be remotely monitored, exams can be collected locally and the results can be remotely analysed by a specialist on-the-fly and only those needing critical care are transferred to the closest medical facility. And, even in this case, the vital information of the patient can be transferred online from the ambulance to the medical centre, allowing the physicians to be better prepared when the patient arrives

Several studies published by 5GPPP[1] show that the average medical expenses in Europe is 10% of the GPB and these expenses are growing over time, since the population is aging. 5GPPP also stated that e-health could have saved up to 99 billion euros in EU in 2017, if the technology were fully developed and deployed. The economic importance of e-health has been discussed by other authors. Manyika has predicted that the impact of e-health in the world economy will vary between US$ 1,1 e US$ 2,5 trillion in 2025[2] and IHS 5G report[3] also highlighted the importance of e-health for the global economy.
5G-RANGE project covers e-health as one of the main use cases, supporting these services in remote areas. This new ecosystem could be divided into three main parts:

- First-Aid Care: During the call to emergency services, a video conference will help the emergency center to provide first care instructions to ensure the patient's safety during the wait for the ambulance as well as to provide more information to ambulance staff, so they can be more effective.
- Ambulance Attendance: Due to the potentially long time that it takes to arrive at the hospital, it is important to have the hospital team updated, providing real-time information and video image, in order to get them ready to the patient's arrival.
- Hospital Care: It is the final patient treatment, which should not dependent on 5G innovative technologies. However, during the first-aid, rescue process, or even for routine health monitoring, physicians can monitor the situation, provide instructions, etc.

The main requirement for e-health in remote areas are:

- throughput of at least 1 Mbps per sensor or equipment: allows latency-sensitive data, collected from patient by health monitors, to be transmitted to the medical center in real-time;
- voice connectivity: allows the paramedics to be in contact with the medical center anywhere;
- at least 120 km/h speed: allows connectivity of the ambulance with the medical center even while travelling in highways.

E-health application in remote areas do not need to have a dedicated communication infrastructure. Since 5G-RANGE is flexible to cover different use cases simultaneously, the same infrastructure can be shared among different service providers. For instance, a rural micro operator can deploy the network to offer Internet access and IoT services in remote areas and the resources to cover the e-health applications can be allocated under demand. The rural micro-operator can receive revenues from the public authorities or health insurance companies when its network is used by e-health applications. Private subscribers can also contract e-health QoS from the micro-operator, in case they want to continuously use these services. One example for

[1] 5GPPP (2015b, September). 5G and e-Health, 5G-Infrastructure-Association. 5GPPP. Available in www.5g-ppp.eu

[2] Manyika, J. et. al. (2013). Disruptive technologies advances that will transform life, business, and the global economy. McKinsey Global Institute, 1-162.

[3] IHS, "IHS Economics & IHS Technology Report: 5G Economy - How 5G Technology will Contribute to the Global Economy".

this situation is the elderly monitoring service, where the health, position and activities of advanced aged people can be constantly monitored remotely.

2.2.2 Use cases and the network layer architecture

For these use cases the most relevant KPIs and requirements are presented mainly oriented to support the provisioning of MBB for distances above 50 km. The project challenge is to achieve 100 Mbps in the downlink for stationary reception. The different requirements are defined for the physical and MAC layer that are the main components of the 5G-RANGE RAN (in Table 1 it can be seen that the required services are common, Voice, MBB, MTC).

Table 1. Core use cases summary

	Use Case Name	Use Case	Vertical Market	Service	Scenario
Core Use Cases	Voice and Data Connectivity	Basic data speeds and voice services for very large areas	Telecom service providers	Voice, MBB	Remote and Underserved
	Smart Farm	Data collection and analysis, crop monitoring, production traceability, remote maintenance and diagnosis, cattle counting, etc.	Agribusiness	Voice, MBB, MTC	Underserved
	Wireless Backhaul	Usage of TV broadcast network infrastructure for wireless backhaul implementation	Telecom service providers	MBB	Underserved
	Remote Health Care	Health/medical assistance and monitoring	Health	Voice, MBB	Remote and Underserved

For the network layer, apart from assuming that it will be honoring the different end to end quantitative requirements (e.g. voice E2E max latency, medium latency, throughput) a mandatory functional requirement is stablished:

- Req-F.m.12: 5G-RANGE system should be based on 3GPP features (LTE Release 14 and NR Release 15) for topics that are not the scope of the project, for example, the upper layers (above MAC) of the radio protocol stack.

In order to accomplish this requirement, the Network-level architecture will be based on current 3GPP Regular Release 16. Throughout the execution of 5G-RANGE project we have been considering release 15 in D5.1 and now 16 in this document (release 17 is still in its early stages although the architectural sections are already quite stablished in release 16).
Different parts of the architecture included in this document will be validated in the Proof of Concept (PoC) that will be mainly focused on the first two core use cases. However, the network architecture is generic and has not been particularly designed for these use cases so that it can accommodate different services. The PoC will serve to verify different parts of it. There is currently no full implementation available of the 5G core network that can be used for testing or comparing purposes so the final functions to be verified in the testbed will heavily depend on the evolution of the 5G testing ecosystem.

3 Definition of the network-level architecture

Figure 4 outlines the architectural design of the network-level architecture of 5G-RANGE. The architecture has been defined according to the design objectives indicated in the previous section, with the main goal to provide end-to-end IP network connectivity to User Equipment (**UE**) of 5G-RANGE.

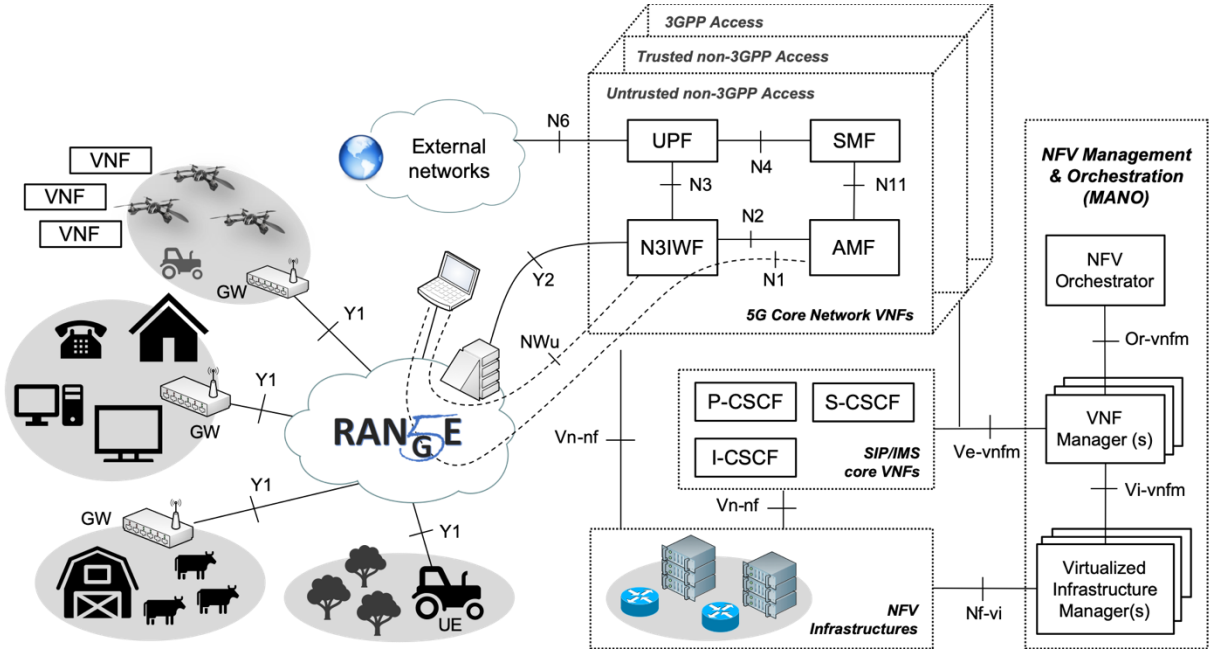

Figure 4. Overview of the network-level architecture of 5G-RANGE

The network layer has a notorious significance beyond the radio access network. It considers and complements the physical and the cognitive MAC layers with essential features to support an appropriate end-to-end operation, including:

- Obtaining IP network connectivity by end devices;
- managing the terminal mobility;
- interoperating with public Internet services, such as the email or the world-wide web, as well as with operator-specific services, like IP telephony; and,
- supporting security services to protect the access of end-user terminals to the network.

On the other hand, given the application scope of 5G-RANGE, which aims at supporting Internet and 5G communication services over remote geographic areas, the design of the network-layer considers UE of heterogeneous nature and capacities. These do not only encompass traditional end-user devices that are of common use to access the Internet and other operator-specific services, such as laptops, TV sets, or mobile smartphones, but also, other devices that enable use cases of relevance in rural areas, for instance Small Unmanned Aerial Vehicles (SUAVs) or harvesters to support agribusiness and smart farming. As it is described later in this document, these devices are also considered in the context of the 5G-RANGE network layer as enabling platforms to transport computing and networking resources, allowing to complement the access network infrastructure available on remote areas. In the network-level architecture of 5G-RANGE, end-user devices can directly be connected to the 5G-RANGE access network, or they can get network access connectivity through a Gateway (**GW).**

This component provides the control and user-plane functions of a 3GPP UE, and acts as a data-traffic relay between end-user devices and the access network.

Last, but not least, the network-layer architecture of 5G-RANGE is strongly based on a set of supporting technologies that have proven to be fundamental enablers in the development of the 5th generation of mobile networks, or 5G. In the following subsections, we place the focus on each of these technologies, describing their applicability in the context of 5G-RANGE.

3.1 Integration with operator networks

One aspect that is fundamental to provide secure end-to-end IP network connectivity to UE, is to support the integration of the 5G-RANGE access network, i.e. the wireless network the UE attaches to, with the core network of a telecommunication operator. This will allow 5G-RANGE users to access Internet services (e.g., Email, Web browsing, stored/live video streaming, etc.), as well as to communicate with other users in external data networks. In this respect, Release 16 of 3GPPP specifications defines the architecture of the core network for a 5G system, to be implemented by telecommunication operators [1]. Figure 5 shows the architectural reference model of the 5G system defined by 3GPP, including its main building blocks: the 5G core network, the 5G (radio) access network, and the UE.

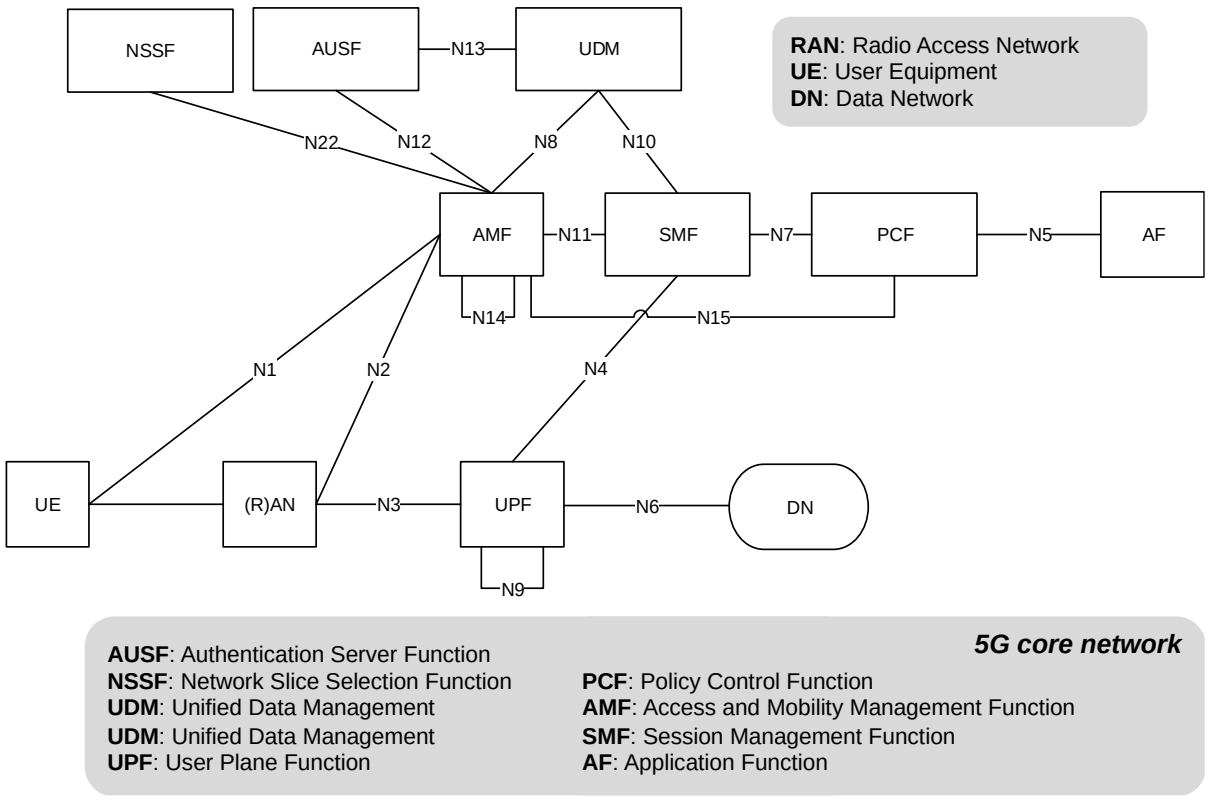

Figure 5. 5G system architecture defined by 3GPP [1]

Among the main network functions of the 5G core network, which are relevant in the context of this document, we may cite the following ones (a comprehensive description of all the network functions of the 3GPP 5G system can be found in [1]):

- The Access and Mobility Management Function (AMF): it provides functions related with management of registration, connection reachability and mobility, as well as access authentication and authorization, among others.
- The Session Management Function (SMF): this network function is in charge of session establishment, modification and release; it also supports the allocation of IP addresses to UE, as well as their configuration in the user terminals via DHCP (Dynamic Host Configuration Protocol); additionally, it provides functions related with the configuration of routing policies in the user plane and supports charging with the collection of relevant information.
- The User Plane Function (UPF): it provides the UE with the point of interconnection with external networks in the user plane, being in charge of the inspection, routing and forwarding of data packets towards these networks; the UPF supports the enforcement of policy rules (e.g., gating or redirection of data traffic); it also provides QoS (Quality of Service) functionalities in the user plane, such as guaranteeing established limitations on data rates, as well as transport-level packet marking and packet buffering functions.

In addition, the architectural specification of the 5G system by 3GPP makes a distinction between 3GPP access and non-3GPP access networks. In particular, Release 16 of 3GPP specifications considers two categories of wireless non-3GPP access networks from the point of view of the public land mobile network operator: untrusted and trusted. In addition, it describes two network functions to support the connectivity of an untrusted and a trusted non-3GPP access network to the 5G core network, i.e., the Non-3GPP Inter-Working Function (**N3IWF**) and the Trusted Non-3GPP Gateway Function (**TNGF**), respectively. Both network functions support the N2 and N3 reference points shown in Figure 5, to support control plane and user plane procedures with the 5G core network.

The N3IWF supports the connection of untrusted standalone non-3GPP access networks to the 5G core network. The reference architecture for untrusted non-3GPP accesses is shown in Figure 6. In this figure, the N3IWF represents the point of contact of the UE in the 5G core network, acting as an intermediate entity that receives the control and data-plane information exchanged with the UE.

The N3IWF performs, among others, the following functions: it participates in the authentication and authorization of a UE to access the 5G core network from an untrusted non-3GGP access; it supports the establishment of Internet Key Exchange (IKE) and IP Security Protocol (IPsec) security associations (see [5]) to protect the exchange of control and data traffic; it forwards control-plane information between the UE and the AMF, as well as user-plane information between the UE and the UPF; it processes information received from the SMF related to the user-plane sessions and QoS; and, it provides mobility support within non-3GPP access networks and QoS enforcement functions.

On the other hand, the TNGF enables a 3GPP UE to get connectivity to a 5G core network via a trusted standalone non-3GPP access. Figure 7 outlines the reference architecture for trusted non-3GPP accesses. In this architecture, a Trusted Non-3GPP Access Point (TNAP) represents the point of attachment of the UE to the non-3GPP access network (as an example, in IEEE 802 .11, this would be a wireless LAN access point). The TNGF provides similar functions to the N3IWF, i.e., it supports authentication procedures during the registration of the UE, and relays control and user-plane information between the UE and the 5G core network. In a trusted non-3GPP access, security relies on the specific mechanisms provided at the data-link level between the UE and the TNAP. Therefore, encryption between the UE and the TNGF is not necessary, and IPsec may only be used for integrity protection. In any case, the utilization of IKE and IPsec

technologies over a trusted non-3GPP access simplifies the implementation of a 3GPP UE reducing the heterogeneity of control and user plane mechanisms, as the UE may use similar procedures to connect to a 5G core network either from trusted or untrusted accesses.

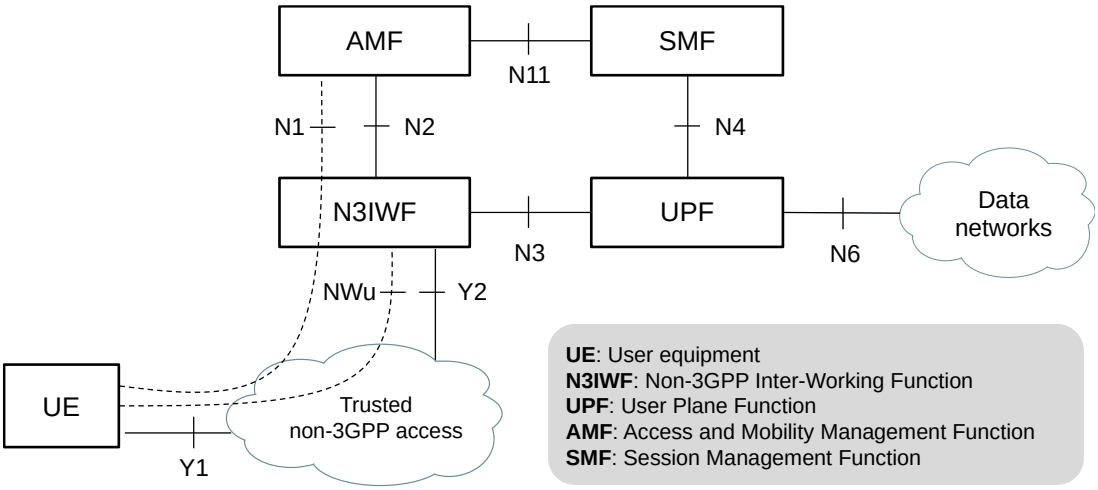

Figure 6. Reference architecture for non-3GPP accesses

At the time of writing this deliverable, with the standardization work in an early stage, it is still unclear whether the 5G-RANGE access network will receive the consideration of a 3GPP access network. Consequently, the architectural design of the 5G-RANGE network-level, as described in this document, contemplates the three options illustrated in Figure 5, Figure 6, and Figure 7, to support the connectivity of the access network with the 5G core network.

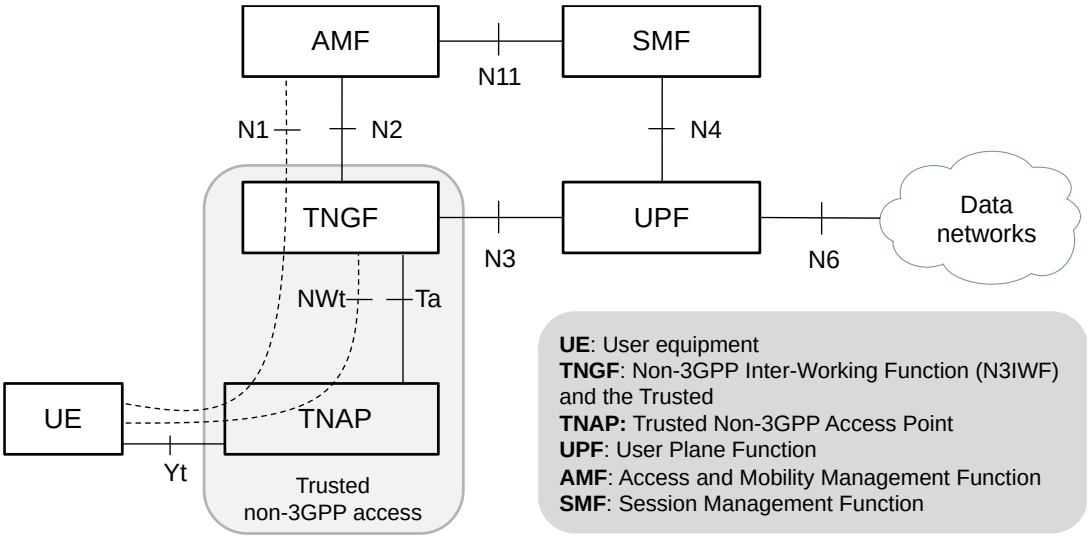

Figure 7. Reference architecture for trusted non-3GPP accesses

The 5G system introduces the concept of a Service-Based Architecture (SBA) as an alternative communication system between the core Network Function (NF). In this system, functions expose their outputs as micro-services or using a Representational State Transfer (REST) Application Programming Interface (API) and other functions can consume these services with a request-respond or a subscribe-notify scheme. The use of VMs as the foundation of VNFs is

required to allow flexible and reconfigurable conditions, however, virtualised environments may suffer of performance degradation over time. Finally, it is important to consider that latency and reliability are essential requirements that could be heavily impacted by VNF degradation on virtualised conditions.

3.2 Support of operator specific services

The IP Multimedia Subsystem (IMS) [11] was developed by 3GPP during the standardization process of the Universal Mobile Telecommunications System (UMTS). It was originally designed as a key element to enable the delivery of value-added multimedia services, including those that were traditionally provided through circuit-switched networks, over packet-switched operator networks.

In this respect, the IMS provides a set of key functionalities to support the establishment, modification and termination of multimedia sessions over packet-switched operator networks, also covering fundamental aspects related with QoS provision, charging functionalities, security, roaming, and interworking with Internet and circuit-switched network services.
Nowadays, IMS is commonly deployed by operators to support IP telephony services over their fixed access networks (e.g., over optical accesses), and is also utilized in telephony service deployments over 4G/LTE networks (a concept referred to as VoLTE, or Voice over LTE). Several standard developing organizations like European Telecommunications Standards Institute (ETSI)[4] and ITU-T[5], have studied its applicability to support other types of multimedia services, like IP television and video-on-demand. However, the standard specifications produced these organization regarding these services have obtained limited adoption by the telecommunications market, and other standard Internet solutions, such as HTTP (Hypertext Transfer Protocol) streaming for video-on-demand, as known as Dynamic Adaptive Streaming over HTTP protocol (DASH or MPEG-DASH), have obtained a higher impact.

In any case, given its current relevance to support IP telephony services in operator networks, the network-layer of 5G-RANGE considers the utilization of a **SIP/IMS core** (see Figure 4). In IMS, the Session Initiation Protocol (SIP) [12], defined by the Internet Engineering Task Force (IETF[6]), is used to provide the signalling functions that are needed to establish audio/video telephone calls and conferences. The Call Session Control Functions (CSCFs) of the IMS core are the network functions that handle the SIP signalling originated and terminated by the users' phones. In particular, the Proxy-CSCF (**P-CSCF**) is the contact point of the user terminal into the IMS. It can contact the PCF (Policy Control Function) of the 5G core network to provide relevant information regarding the multimedia sessions that are being established through the IMS core (e.g., audio/video calls), such that the PCF may authorize the necessary QoS resources for the session.

The Interrogating-CSCF (**I-CSCF**) is the entry point to the operator network for calls addressed to the operator users. Finally, the Serving-CSCF (**S-CSCF**) performs session control functionalities, and authorizes, authenticates and registers the operator users that use the IP telephone service.

[4] The European Telecommunications Standards Institute (ETSI): https://www.etsi.org (last access on Dec. 2018).

[5] The ITU Telecommunication Standardization Sector (ITU-T): https://www.itu.int (last access on Dec. 2018).

[6] The Internet Engineering Task Force (IETF): https://www.ietf.org (last access on December 2018).

3.3 Virtualization of network level functions

The advent of 5G networks imposes new and stringent requirements to guarantee a fair balance between speed, latency and cost of communications, which are expected to be potentially originated by millions of connected devices. For that purpose, the shared use of resources, allowing the dynamic provision of network functions, is foreseen as one of the key enablers of this new generation of networks. In this context, the softwarization of network functions, or Network Functions Virtualization (NFV), presents an innovative solution in the sector of information and communication technologies, aiming at supporting the active and high-performance processing of traffic delivered across 5G networks.
The ETSI is leading the standardization activities on NFV, which encompass the definition of a reference architectural framework [13] for NFV deployments.

Among the main reference blocks of the ETSI NFV reference architecture, the Virtual Network Function (**VNF**) is the element in charge of providing the functionality of a network component (e.g., a residential gateway, an access point, a firewall, etc.) through a software implementation. The definition of the 5G system by 3GPP [1] already embraces the NFV paradigm, considering the possibility of implementing the 5G core network components as virtual functions and deploying them over a cloud environment. As shown in Figure 4, the network-layer architecture of 5G-RANGE is strongly based on NFV technologies, supporting the execution of all its architectural components as VNFs.

A set of interconnected VNFs creates a Network Service (NS). As a simplified but illustrative example, an instance of a N3IWF and an instance of a UPF may be provided as virtualized functions in the form of separate VNFs, and be interconnected to build a Network Service (NS) providing user-plane functionalities to several users.

On the other hand, the NFV Infrastructure (**NFVI**) provides the hardware and software resources that are needed to deploy, execute and manage the VNFs. These include hardware resources such as computing, storage and networking, which are allocated to the VNFs through a virtualization layer abstraction that enables the decoupling of the VNF software from the underlying hardware.

Finally, the Management and Orchestration (**MANO**) framework supports the proper coordination of all the operations carried out in the NFV environment. These operations enable the automated deployment of network services composed by the VNFs over the NFVI. For this purpose, the definition of the MANO framework contains three functional elements: 1) the **Virtual Infrastructure Manager** (VIM), controlling the hardware and software resources of the NFVI; 2) the **VNF Manager** (VNFM), which handles the life cycle of VNFs and NSes; and 3) the **NFV Orchestrator** (NFVO), in charge of building up end-to-end services, and coordinating with the VNFM and VIM functions the available resources and the deployment and configuration of the VNFs.

3.4 Extensions to the network-level

3.4.1 Motivation

The work in WP5 also addresses the design of extensions to the baseline architecture of the network-level presented in Figure 4. These extensions aim at providing a cost-effective approach to complement the network infrastructure resources of the 5G-RANGE access network, e.g., to support network communications beyond the boundaries of the 5G-RANGE radio cells, using computing platforms (on-boarded for instance on Small Unmanned Aerial Vehicles, SUAVs) that could be deployed, or opportunistically used when available, on remote areas.

The combination of 5G communications and SUAVs has recently turned out to be more than just a trendy idea. Separately, they have an unquestionable impact in our society, and an increasing number of new related applications are continuously appearing. However, it has only been recently that the amalgamation of these fields has received attention from the research community, and their natural synergies are beginning to be explored considering aerial vehicles as enablers of 5G communications [14][15]. On the one hand, the development of 5G standards and technologies does not only aim at providing performant communications to end users. Whereas this was a primary driver in the previous generations of mobile networks, 5G considers vertical sectors as key adopters, and aspires to create a novel ecosystem where technical innovations and business models are also driven by vertical-specific use cases. On the other hand, SUAVs are currently gaining prominence as key assets in different vertical markets, such as precision agriculture, disaster or accidents assistance, city management and public safety.

In these contexts, SUAVs have traditionally been considered as appropriate platforms to generate, process and/or transport relevant information (e.g., video and other sensed data). However, with the recent advancements on the miniaturization of electronic devices, SUAVs can onboard lightweight hardware platforms (e.g., battery-powered single board computers [20]), providing compute, storage, and network resources. These platforms open new and exciting opportunities to support a cost-effective and flexible provision of vertical applications. As an example, a smart farming provider could use a number of SUAVs to rapidly build an aerial network infrastructure over an extension of farmland. These SUAVs could support the collection, generation, processing, and dissemination of relevant information to offer smart farming operations, e.g., surveying and evaluating the status of a crop field, or detecting objects of interest to perform closer inspection or aid tasks such as precision spraying. In a different environment, an emergency service provider could rapidly deploy a fleet of SUAVs to support voice and data communications to a team of firefighters, working on a fire extinction activity along a large forest area, where existing infrastructures may be insufficient or even unavailable.

In this future view, swarms combining UAVs of different sizes and characteristics might be positioned over delimited geographic areas, and provide a baseline resource infrastructure, capable of flexibly executing different network functions and services, which could then be offered on-demand by a service provider over that areas. As an example, a telecommunication operator could deploy several SUAVs to temporarily provide Internet access over a specific geographic location, for instance because the access network infrastructure of the telecommunication operator in that area is unavailable (e.g., as a consequence of a natural disaster) or insufficient (e.g., in a crowded event). In these cases, SUAV platforms could enable a cost-effective deployment of network functionalities (e.g., wireless access points, routing

functions, IP telephony servers, etc.) over a target geographic area. This way, SUAVs could be used to complement the available network infrastructure over specific areas, which could favour the provision of the network services expected for 5G.

For instance the 5G-RANGE radio cell as presented in the general architecture may not be covering the desired area (because it can be out of it limits or because being inside the covered area there may not be a proper line of sight visibility and then an under-covered shadow area has been created). In these areas and for particular events (search and rescue operation for persons of even for cattle, a certain field inspection, medical urgent requirement in a certain area like a camping or a concert, disaster management like fires or flooding, etc.), the proposed infrastructure extension could be deployed. The main objective is not just to have an ad hoc infrastructure providing wireless access. The main objective is to extend the advantages of the 5G-RANGE created infrastructure (including the RAN, but also the core network, resource orchestrator, etc.) to extend access to all the available services. Other state-of-the-art possibilities like satellite alternatives or fixed infrastructure deployment may not be cost effective or may penalize the network architecture in terms of performance (e.g. additional latency).

It is important to emphasize that the practical realization of this view presents a first and fundamental challenge, i.e., the lack of flexibility exposed by existing SUAV products. These products are commonly equipped for specific purpose missions with a fixed set of software and hardware appliances and base their operation on proprietary mechanisms and protocols. It is only when a programmable logic is embedded into the aerial vehicles that the flexibility of multi-mission SUAVs can be freely expressed [17][19]. On the one hand, a programmable logic might allow the provision and configuration of different functions and services over adaptable SUAV units (e.g., to act as an aerial relay for voice and data communications). On the other hand, such a logic could enable the flexible interconnection of SUAVs to build programmable aerial networks, which could be rapidly deployed over delimited geographic areas and be fully integrated with existing 5G deployments like 5G-RANGE one.

The NFV-SUAV combination has delivered different benefits as it can be seen in for instance in [20], where authors propose a full video-surveillance system for big poorly internet-covered areas using UAVs, whose mobility can be used to distribute the aircraft along an specific geographical region to obtain video footage, using NFV as the platform to transmit the video signal through a network of several VNFs running inside the aircraft. This behaviour can be considered rather complex to perform through pure virtualization, due to the execution times needed for performing all aircraft actions, which might not be fast enough to provide the service, especially since the amount of resources that can be onboarded are constrained by their space and weight to allow the aircraft to stay in the air as much as possible. Instead, authors propose using paravirtualization, where Virtual Machines (VMs) share the hardware directly with their host, allowing its Operative System (OS) to be a platform for directly (and exclusively) operating with the VNFM, saving valuable execution time and allowing a better management of the resources on the aircraft. Other outstanding challenge in this area is related with the migration of VNFs between the NFV units of a network composed of several SUAVs in order to perform a seamless transition of VNFs between UAVs to minimize service cuts and/or reduce the amount of time a service remains offline. This behaviour can only be performed if all the associated network services, routing and operational control migrate quickly. An example of this can be seen in [21], where authors propose an NFV-based solution that takes into account the high-mobility requirements of these networks. The work in [22]

presents an architecture of function softwarization in an irrigation system based on a network of sensors and UAVs.

Based on these observations, a prior work in [13] explores the utilization of SUAVs as 5G points of presence, with the capacity to on-board computing, storage and networking resources that support a cost-effective deployment of network infrastructure over delimited geographic areas. This concept was further elaborated in [19] and [20], which present the design of an NFV system capable of supporting the agile configuration and deployment of moderately complex network services over a cloud platform offered by a swarm of resource-constrained SUAV equipment. In that work, the first steps towards validating the practical feasibility of this approach were given, showcasing the deployment of a functional IP telephony service over a set of NFV-enabled SUAVs. This particular approach was continued in 5G-RANGE in [20], studying the feasibility of utilizing programmable SUAVs to support heterogenous vertical services. To this purpose, a practical approach was followed. A multisite NFV experimentation testbed was designed and built, including an infrastructure of server computers, SUAVs and other resource-constrained devices. This testbed was set up with the main goal of facilitating experimentation with vertical services, with SUAVs allowing for the prototyping and validation of these services in a realistic multi-site environment.

This approach may be particularly useful in some of the use cases under consideration in the project [16], particularly regarding *Voice and Data Connectivity over Long Distances* and *Smart Farming for Remote Areas* (or *eHealth*, although this last use case has not explicitly been considered by 5G-RANGE proof of concept activities in WP6). In this respect, the 5G-RANGE network layer considers the utilization of SUAVs (carrying the required payloads like as high-resolution daylight video cameras, thermal imaging cameras, single board computers, Global Positioning Systems (GPS), and a myriad of other low-cost sensor and electronic devices), as well as other facilities and/or vehicles that may be available on the ground (e.g., harvesters, tractors, sprayers, or even mobile ground stations that may be specifically deployed to aid data communications during specific events).

The network layer of Figure 4 relies on inexpensive, general-purpose hardware and software platforms. Taking advantage of the commented baseline work, these components could be transported by SUAVs to specific geographic locations as needed. In addition, the network layer would opportunistically use other computing, storage and networking resources that might be available in a target area (e.g., single board computers on-boarded on harvesters or tractors). All these devices would be interconnected to build an ad-hoc aerial/ground NFV infrastructure, which could be leveraged by the 5G-RANGE MANO system to deploy value-added NFV services over the target area, complementing the communication services offered by the 5G-RANGE access network.

In the following, we describe how this approach could be of interest to support three use cases of relevance in 5G-RANGE: *Voice and Data Connectivity over Long Distances*, *Smart Farming for Remote Areas* and *eHealth*. The *Wireless Backhaul and Local High-Quality Connections* is out of the scope of these extensions by its own definition, and since it is a backhauling use case while these extensions are for beyond the edge use cases (although it is obviously considered by the general network layer architecture).

The 5G-RANGE proof of concept showing this proposed extension has been selected (early 2020) by ETSI OSM as one of its reference PoCs (see [21]) and have recently been awarded by ETSI as the Best Proof of Concept with OSM during release eight cycle.

3.4.2 Voice and data connectivity over long distances

A first use case where the utilization of virtualization technologies and aerial/ground vehicles can be helpful is in the provision of voice and data connectivity over long distances. For example, in case that network connectivity needs to be provided to users participating in a festive event in a remote area, beyond the limits of a 5GRANGE radio cell; or in an emergency, to support efficient communication services for an emergency response team, as in a fire extinction activity in a rural area.

An example of the first case is illustrated in Figure 8, where several SUAVs are deployed to provide voice and data services to users beyond the coverage of the 5G-RANGE access network. These SUAVs, along with other ground vehicles (e.g., tractors, harvesters, etc.) that might be available within that area, provide a programmable network infrastructure under the control of the 5G-RANGE MANO platform, which can deploy a number of VNFs over the target area, whose composition enables the provision of the intended voice and data services.

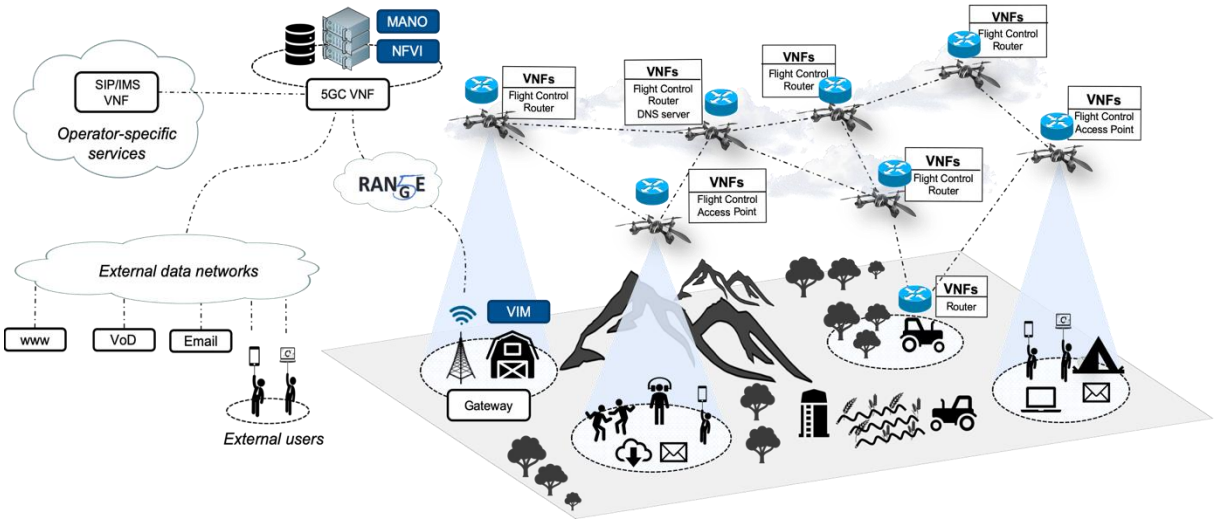

Figure 8. Voice and data connectivity over long distances.

In the considered example, a subset of the SUAVs would deploy a Wi-Fi access point, to provide network access connectivity to ground users (e.g., people participating in a pilgrimage or celebration). Other SUAVs can be positioned at specific locations (flying or perched on land, for instance in specific-purpose ground structures) and offer routing and forwarding functionalities for data traffic, creating an aerial network infrastructure that supports data communications over the geographic. Terrestrial vehicles could opportunistically be used to complement the network infrastructure, e.g., providing routing functions.

A flight control function can be instantiated at each aerial vehicle, and be configured with information on fixed trajectories in the form of waypoints, to be autonomously followed by the vehicle to get to a stable GPS position. A 5G- gateway equipment could be used at a ground

facility, providing an entry point to the 5G-RANGE access network, supporting data communications between the remote area and external data networks through the 5G core network. This would be provided in the form of one or several VNFs in the operator domain (i.e., the 5GC VNF in Figure 8). This way, the end-users would be able to access Internet services from the remote area, such as web browsing, email, or video-on-demand (VoD). In addition, an IMS core deployed as a VNF would allow establishing IP telephony calls among the remote users, and well as with other Internet users.

The most important KPI in this example is the service creation time cycle, i.e., the time to deploy the network service conformed by the aforementioned VNFs. This time is to be reduced to an average of 90 minutes according to the 5G-PPP vision [23]. As a first step towards verifying the practical feasibility of utilizing virtualization technologies and SUAVs/ground vehicles in laboratory conditions, we have designed and implemented a specific experiment aligned with the scenario of Figure 8. In this experiment, an IP telephony service was deployed over a set of SUAVs and cloud platforms, including a baseline implementation of a 5G core network and a gateway, being its operation successfully verified. The details and results of this experiment are provided in deliverable D5.3 [42].

3.4.3 Smart farming for remote areas.

A second use case where the proposed extensions to the network layer can be of interest is in smart farming for remote areas. In a smart farming scenario, a number of aerial and ground vehicles could be used to efficiently build an ad-hoc network infrastructure over an extension of farmland. This infrastructure could then be used to support the generation and dissemination of relevant information, e.g., daylight and thermal images, real-time video, data from sensors deployed on the ground and/or on-boarded on the aerial/ground vehicles, etc. This information might serve several uses, such as the early identification of diseases, or the estimation of production volumes.

Figure 9 shows an example where a service provider uses a MANO platform, with the 5G-RANGE network layer extensions, to deploy a smart farming service on a remote area. The service would be deployed as a composition of VNFs over a set of SUAVs and other ground devices that might be available on the farmland. Some VNFs running at the SUAVs would support the collection of relevant information from the hardware/software payloads of the aerial vehicles (i.e., video, images from daylight/thermal cameras, and other sensed data).

This information would then be delivered towards a ground facility, possibly through other SUAVs and/or ground devices providing routing functions. A gateway at the ground facility would facilitate the distribution of the collected information towards the service provider domain, through the 5G-RANGE access network and the 5G core network (both components are under the control of the MANO platform of the operator). Finally, the received information would be processed, stored, and exposed to authorized users through specific functions at the service provider domain. These functions would be supported as VNFs in an NFV infrastructure owned by the service provider.

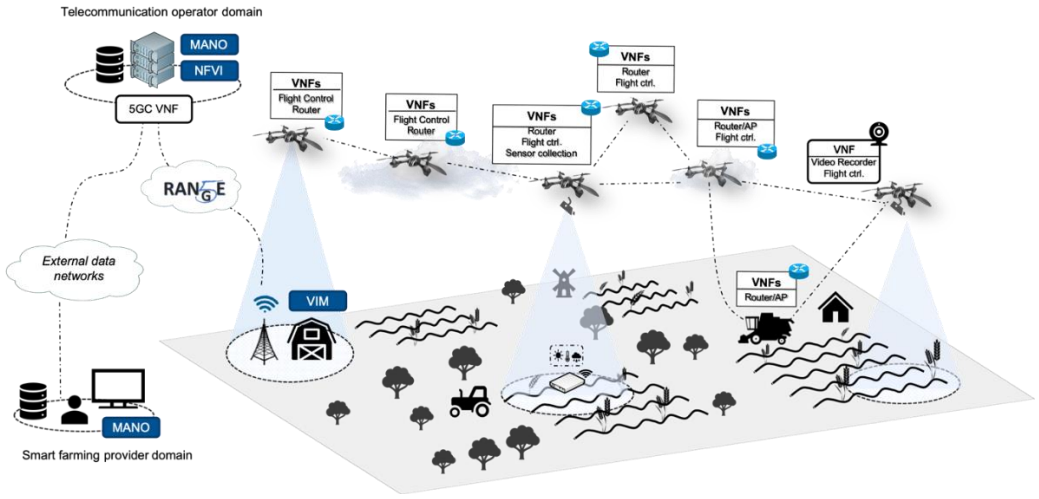

Figure 9. Smart Farming for remote areas

3.4.4 Health care for remote areas.

A third use case where the proposed extensions to the network layer can be interesting is in eHealth applications beyond the edge of the radio cell boundaries defined by the 5G-RANGE base station.

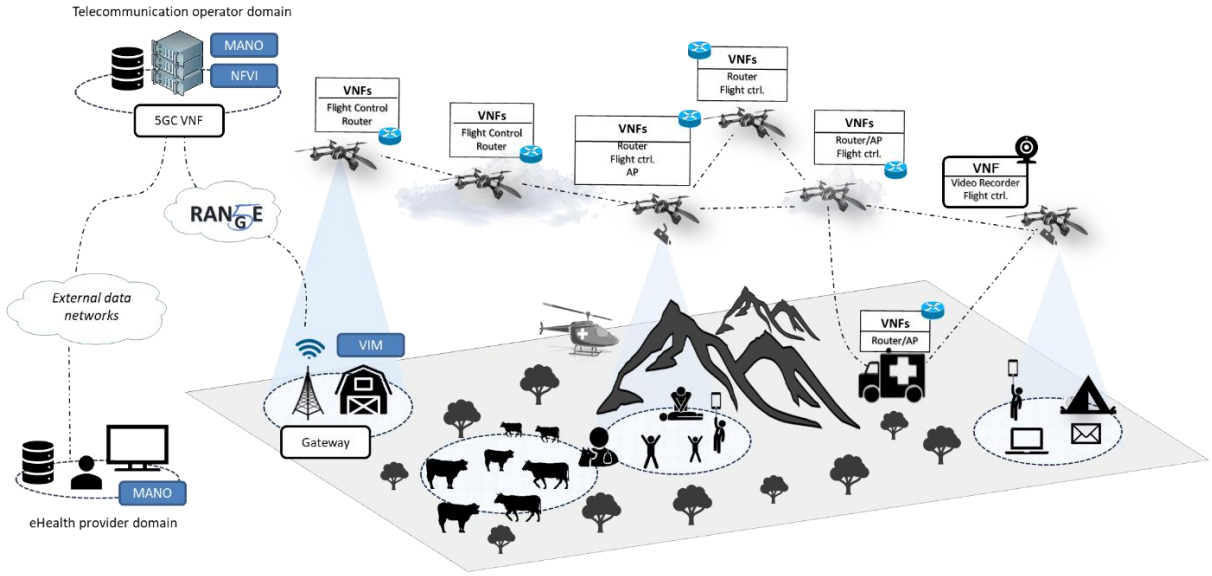

Figure 10. Health care for remote areas

Like in the previous examples (sections 3.4.1 an 3.4.2) a number of aerial and ground vehicles may build an ad-hoc network infrastructure wherever an accident has occurred or just in case this eHealth service is wanted to be temporarily provided (remote camping, particular events, even for cattle veterinary purposes, remote mountains, etc.).

Figure 10 shows an example of the deployment of this service with different SUAVs connected in a network together with different users, or ambulances or medical helicopters taking advantage of the 5G-RANGE deployed radio cell to allow communications to external

networks. Through this infrastructure relevant information could be disseminated like real-time video, GPS position, health relevant information, etc.

As it has been described, the communications across the closest 5G-RANGE RAN deployment would also enable the proper orchestration of the different VNFs that are required on the SUAVs by the MANO platform installed at the eHealth operator (from simple wireless access points, to SIP servers to enable VoIP communications, to video applications, or medical apps, medical equipment delivery services, etc.).

4 Definition of signalling, transport and network procedures

After describing the main building blocks of the network-level architecture of 5G-RANGE, this section identifies the main protocols that can be used to provide secure end-to-end IP network connectivity to 5G-RANGE UE. As it was previously commented, the network-level operates over the cognitive MAC and physical levels, complementing their functionality with fundamental features like IP address allocation and configuration, interoperation with Internet and operator-specific services, mobility management and security. It is important to highlight that these features can only be provided with the coordinated operation of diverse protocols that operate at different levels of the TCP/IP protocol stack, not only at the network level itself, but also at the application and transport levels). Regarding mobility management, since the UPF acts as the anchor point for the UE sessions, we use the mobility that the 5G core provides and its important to note that this is functionality that can also be supported at other layers below and over the IP layer.

Considering the aforementioned considerations, in the following sections we identify the main application, transport, and network-level protocols that, being defined under the umbrella of recognized standards developing organizations, are or particular relevance in the context of the UE of 5G-RANGE.

4.1 Control plane protocols in the 5G system

According to the design of the network-layer architecture, presented in the previous section, the 5G-RANGE UE would get access to external networks and Internet services through the 5G core network of a telecommunication operator. Consequently, the 5G-RANGE UE must implement the control-plane procedures that are necessary to access a 5G core network and establish user-plane sessions.

The protocols that are required at the UE to support these control-plane procedures have been identified by 3GPP in [5], and their operation is described in detail in [6]. For convenience, these protocols are overviewed in Figure 11. In this figure, the Non-Access-Stratum (NAS) protocol [24] supports the communication between the UE and the AMF, for registration and connection management purposes; it also enables the communication with the SMF to manage user plane sessions, via de AMF.

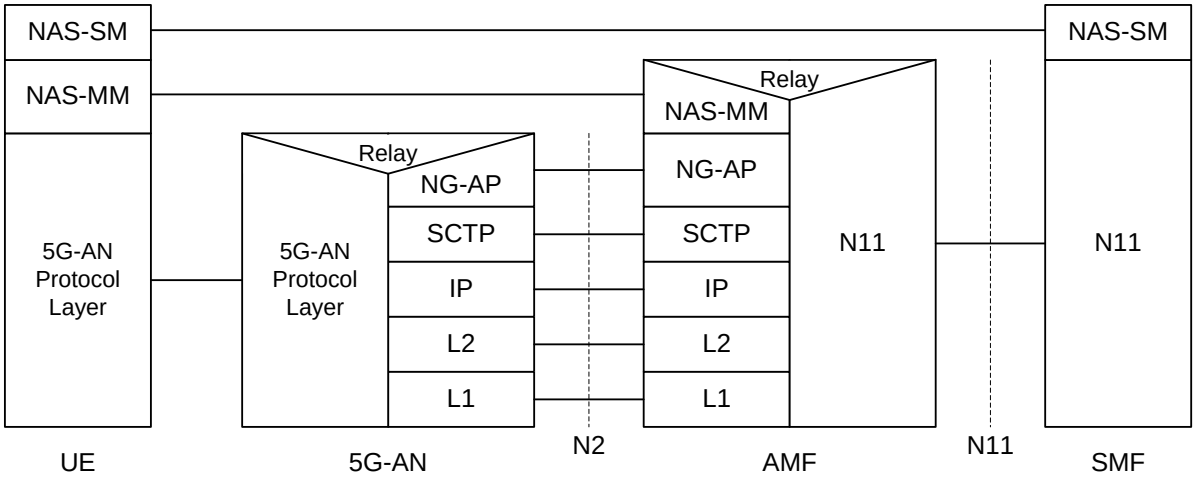

Figure 11. Control plane between the UE and the 5G core network [5]

In the case that the 5G-RANGE access network is considered as an untrusted non-3GPP access, the UE connects to the 5G core network via the N3IWF. In this case, the NAS protocol is also used, and the exchange of control information through the access network is protected using IKE and IPsec. Figure 12 shows the protocol stack that supports control plane operations between the UE and the AMF in this case. Additional protocols not shown is this figure are still necessary to configure the protected IPsec communication between the UE and the N3IWF. Comprehensive information regarding control-plane communications in a 5G system over untrusted non-3GPP accesses is provided in [5] and [6].

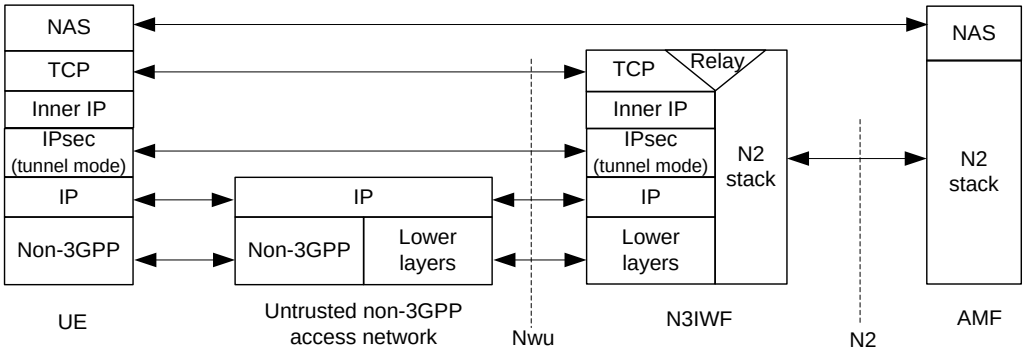

Figure 12. Control plane for untrusted non-3GPP access via IPsec [5]

The abovementioned protocol stack valid for trusted non-3GPP accesses, with the utilization of a Trusted Non-3GPP Access Point (TNAP) and a Trusted Non-3GPP Gateway Function (TNGF), as shown in Figure 13. As already commented, security is based on the specific data-link level mechanisms supported by the UE and the TNAP (this is a trusted entity). Therefore, the IPsec security association established between the UE and the TNGF does not use encryption, it only applies integrity protection.

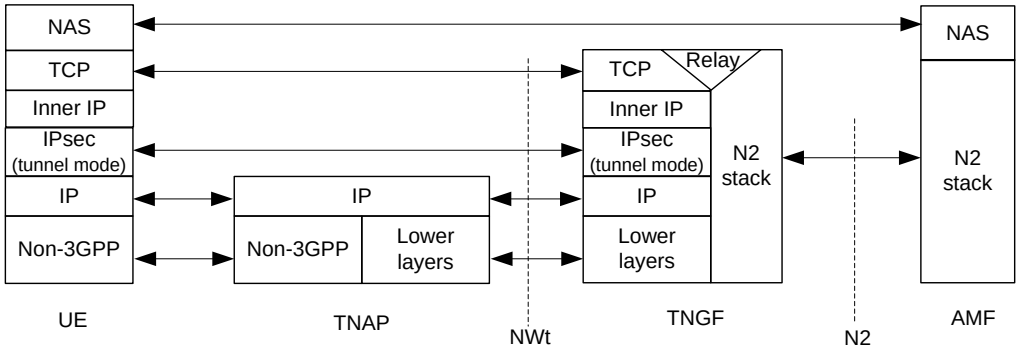

Figure 13. Control plane for trusted non-3GPP access via IPsec

4.2 User plane protocols in the 5G system

Regarding the exchange of data with external networks, in order to support communications with other Internet user and services (e.g., video-on-demand, or web browsing), the first point of contact of the UE at the IP layer in the 5G core network is provided by a UPF. This network function provides the necessary functionalities for the routing and forwarding of data traffic with external networks and offers an anchor point to support the mobility of the UE over the access network.

Figure 14 shows the protocol stack that is needed to support the exchange of data traffic in the user plane over a 5G system. As it can be observed in the figure, if the 5G-RANGE access network is considered a 3GPP access, the protocol stack of the UE above the MAC cognitive and the physical layers of the 5G-RANGE access network (i.e., network, transport and application levels) would not require additional protocols to support the delivery of user traffic through the 5G core network.

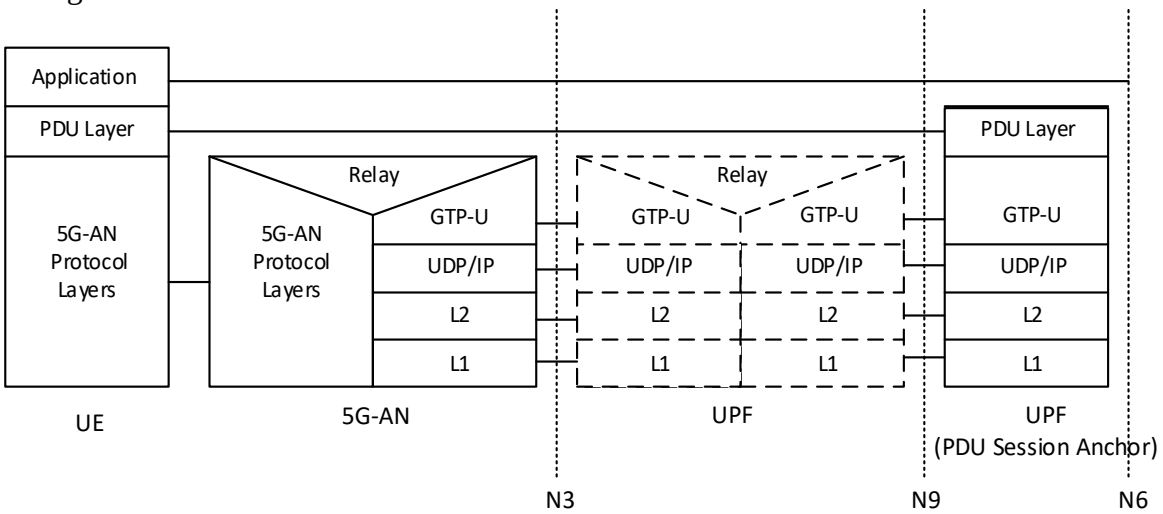

Figure 14. User plane for 3GPP access [5]

Figure 15 shows the protocol stack in the user plane of a 5G system for untrusted non 3GPP accesses. As in the case of control plane information, if the 5G-RANGE access network is considered an untrusted non-3GPP access, the exchange of user plane information between the UE and the 5G core network is delivered via the N3IWF, being protected with the utilization of IPsec technologies.

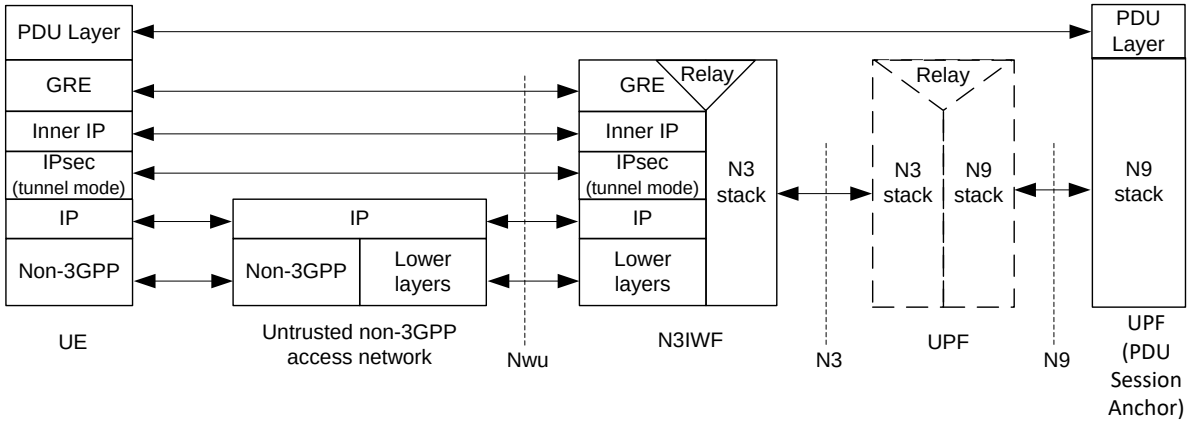

Figure 15. User plane for untrusted non-3GPP access [5]

The user plane protocol stack for trusted non-3GPP accesses are similar to the one shown in Figure 15, replacing the "*Untrusted non-3GPP access network*" and the N3IWF components, by the TNAP and the TNGF, respectively. The IPsec associations established between the UE and the TNGF for user plane communications do not need encryption.

4.3 Protocols at the end user device

Regardless of the protocols that are required at the MAC cognitive and physical layers, as well as those that are needed to support end-to-end communications through a 5G core network, the network applications running at the UE (e.g., a web browser, a video-on-demand client, or an IP telephony application, to name a few of them) may utilize different internet protocols at the network, transport, and application levels.

Figure 16 identifies a non-exhaustive list of the Internet protocols that may be required by the UE of 5G-RANGE, to support the different use cases defined by the project. The figure considers the case where the UE connects to the 5G-RANGE access network via a gateway, which may support UE connectivity using, for instance, a wireless LAN or an Ethernet network, although the same internet protocols could be used by a standalone UE that is directly attached to the access network.

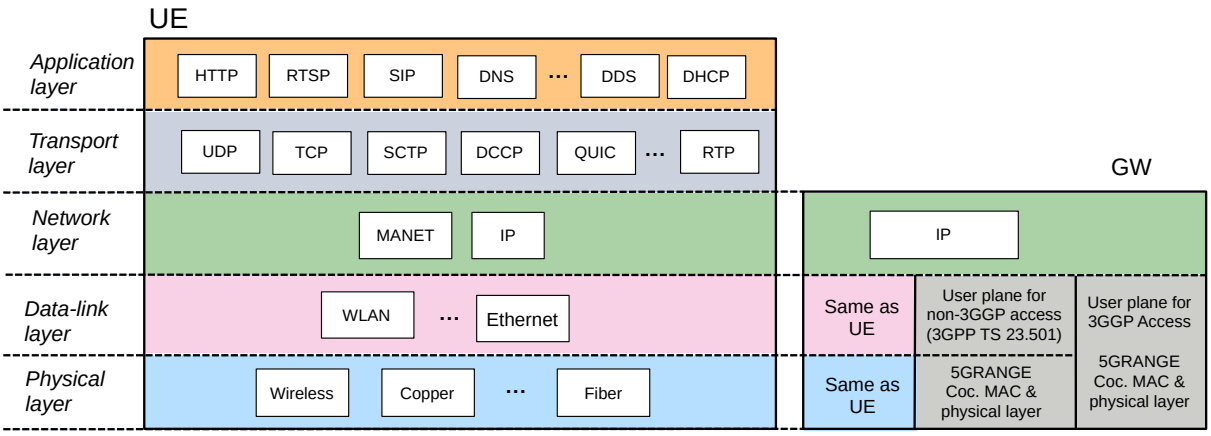

Figure 16. Options for the protocol stack at the end-user device

Without pretending to provide an exhaustive list, we can mention the following protocols that may be relevant in the use cases identified for 5G-RANGE:

- Network layer: IP version 4; IP version 6; or Mobile Ad Hoc Network (MANET) protocols, such as [25] and [26].
- Transport layer: UDP (User Datagram Protocol) [27], to support a datagram oriented transport; TCP (Transmission Control Protocol) [28], to enable a reliable transport service with flow and congestion control functionalities; RTP Real Time Protocol) [29], to facilitate the end-to-end transport of real-time information; TLS (Transport Layer Security) [30] and QUIC (Quick UDP Internet Connections) [31], to enable transport-level security; etc.
- Application layer: HTTP [32], to access web browsing and video streaming services; SIP [12], to support the signalling mechanisms of IP telephone services; the middleware framework of the Data Distribution Service (DDS) [33], e.g., to support efficient mechanisms to distribute information in IoT scenarios; etc.

4.3.1 Control Data Monitoring in the 5G Core

In the scenarios presented by Figure 8 and Figure 9, there are different types of flows routed through the topology, such as: VNFs control data, SUAVs flight control data, and user data traffic. The VNFs control data is managed by MANO and executed in the SUAVs. It is used to manage the NFV infrastructure, defining, for instance, which service should be deployed in a specific SUAV. SUAVs flight control data is related to the control functionality of the SUAVs and its mobility as well. The user data traffic is the user data routed from/to users' devices through the VNF topology to/from external data networks. Different kind of communication technology could be explored to investigate its impact on the previously described infrastructure. The use an out-of-band communication channel to transmit the VNFs control data to manage the VNF topology appears as an interesting option since Lower-Power Wide Area (LPWA) networks could be used in the SUAV/VNF scenario as alternatives technologies to be evaluated in the context of the 5G-Range project.

As described in D5.1, in the SUAV/VNF scenario, the use of LoRA[34] is being considered as an out-of-band communication channel to transmit VNFs control data to manage the VNF topology. The use of a Lower-Power Wide Area (LPWA) appears as an interesting option to expand the coverage area of the SUAV/VNF scenario and to complement the solution. Considering the different LPWA options, as described previously in D5.1, LoRA has relevant features for this purpose, since it has important energy economy features and works with data rates between 0.3 kbps and 50 kbps, depending on the channel bandwidth while allowing ranges of 2-5 km in crowded areas [35] .

To better evaluate the behaviour of the control data traffic in the VNF topology and the viability of using LoRA (or other similar technologies), we developed in WP5 a monitoring set of functions integrated into the 5G architecture. The set is formed by two functions, that intend to evaluate the traffic in the virtualised infrastructure. The functions are the following: an **Observation Function (ObF)** that monitors traffic at each compute node and forwards gathered data to a centralised **Performance Evaluation Function (PEvF)**, responsible for monitoring all virtual resources and generating a consolidated report. Figure 17 show how these two new functions can be integrated to the 5GC architecture by using SBI (Service Based Interface) communication as an access point. Figure 18 shows the prototyped implementation of both functions, to illustrate their integration in the NFVI using a simple configuration, with a local ObF that listens to the main vSwitch at all compute nodes, monitoring transferred data

and performing packet inspection to filter which packets are 5G SBA common interactions. To better evaluate and reduce resource competition with other 5G SBA components, at this stage of the project, the ObF was developed using a bare-metal implementation.

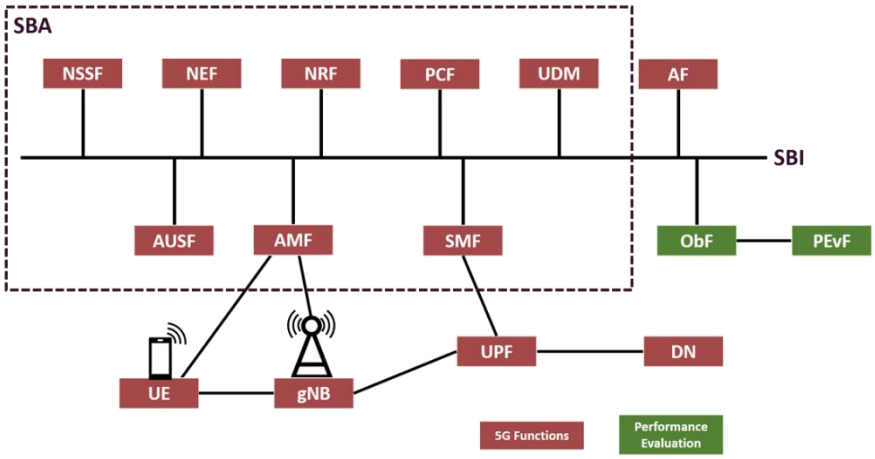

Figure 17. Integration of the Proposed Monitoring Functions to 5GC

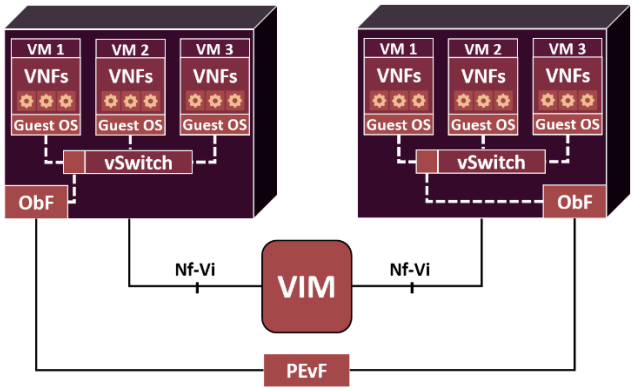

Figure 18. Prototype of Implementation of ObF and PEvF

The PEvF function analyses and evaluates how packets flow between network functions. PEvF continuously monitors VIM and VNFM interfaces and aggregates multiple metrics gathered by the distributed ObFs. From the VIM and the VNMF, PEvF receives information to create both the virtualised infrastructure map and the slicing map. Then, from ObF, the PEvF receives information collected from several compute nodes. Finally, PEvF consolidates all gathered information to create events and notifications about virtualisation health.

The PEvF function works in two modes: (a) a packet monitoring mode, where all vSwitch sniffing will be analysed considering delay and will provide performance information about the overall architecture and specific VN functions; (b) a native SBA mode, to analyse micro-services, filtering the groups by NS and also by network slices to provide information about specific network subdivisions and to identify their response time. The PEvF does not provide micro-services itself and only consumes from the ObF.

ETSI GS NFV-TST 008 V2.4.1[39] describes a group of metrics to measure the QoS perceived by the consumer of any NFV regarding both physical resources and their virtual equivalents. Table 2 shows these metrics classified by resource and indicating the source in our implementation. Each metric uses an unique code. For compute resources, Processor Usage (C.1) represents the total time of instruction execution at the compute node compared with the whole monitoring interval, while Processor Utilisation (C.2) represents the ratio between C.1 and the monitoring interval. For network resources, Packet Count (N.1) represents the number of packets successfully transferred over an interface, Octet Count (N.2) represents the total amount of bytes transferred over an interface considering the sum of all successfully delivered packets, Dropped Packet Count (N.3) represents the number of unsuccessful packet transmissions which occurred because of lack of resource, and Errored Packet Count (N.4) accounts for two groups of corrupted packets in relation to a unique interface. The first group is related to corrupted packets received with wrong sizes or integrity problems; while the second covers all failed attempts to send any packet. For memory resources, Buffered Memory (M.1) represents the total space used for raw disk block storage at a given time; Cached Memory (M.2) is the sum of all memory used as cache; Free Memory (M.3) accounts for the unused memory; Memory Slab (M.4) is the total amount memory used as cache by the kernel for data structure and object storage; and Total Memory (M.5) is the sum of usable memory; and Used Memory (M.6) is a generated metric that can be calculated using previous memory-related metrics.

Since the focus of this evaluation in WP5 is control data, we did not use Compute and Memory metrics in our analysis. Considering Table 1, only the group of network metrics were used in our implementation.

Resource	Code	Metric	Units	Source
Compute	C.1	Processor Usage	Seconds (s)	VIM
	C.2	Processor Utilization	%	VIM
Network	N.1	Packet Count	Number (#)	Obf
	N.2	Octect Count	Number (#)	Obf
	N.3	Dropped Packet Count	Number (#)	Obf
	N.4	Error Packet Count	Number (#)	Obf
Memory	M.1	Buffered Memory	MB	VIM
	M.2	Cached Memory	MB	VIM
	M.3	Free Memory	MB	VIM
	M.4	Memory Slab	MB	VIM
	M.5	Total Memory	MB	VIM
	M.6	Memory Used	MB	VIM

Table 2. Metrics to measure NFV

4.3.1.1 Implementation and Evaluation

The first evaluations include an analysis of the main functionalities given by the NFV deployment for NFs, as well as the effectiveness of the implemented virtualisation enablers. We evaluated how the selected software implementation covers all the requirements for a complete NFV scenario considering the 5G SBA specification. Then, a set of tests were performed in the deployed infrastructure to evaluate the implementation of a VoIP application.

The implementation of the virtualised architecture was done using ETSI NFV guidelines [39]. OSM Release FIVE acts as the NFVO and the VNFM, while OpenStack (Ocata) provides VIM

functionalities. For the VNFM, OSM provides native VNF instantiation, service initialisation and runtime management of virtualised services, achieving a complete lifecycle management. The NFVO provides software resource management. Openstack allows the implementation of compute, storage (both through Nova component) and network (through Neutron component) resources over the NFVI for the whole environment. Both OSM and Openstack provide northbound interfaces as required by ETSI NFV guidelines. For advanced functionalities, Openstack natively supports Service Function Chaining and Network Slicing, and OSM started to support Network Slicing since version 5.

In the tests, we replicated a 5G SBA scenario with decentralised VNF communication over a resource-constrained infrastructure. Figure 19 shows the implementation of a decentralised NF developed with two main components. The first component is a server acting as a CU (Centralised Unit) , executing MANO for available resources, serving as a NFV MANO. The second component is the NFVI, formed by different compute nodes acting as DUs(Distributed Units) and making virtual resources available for the VNF deployment.

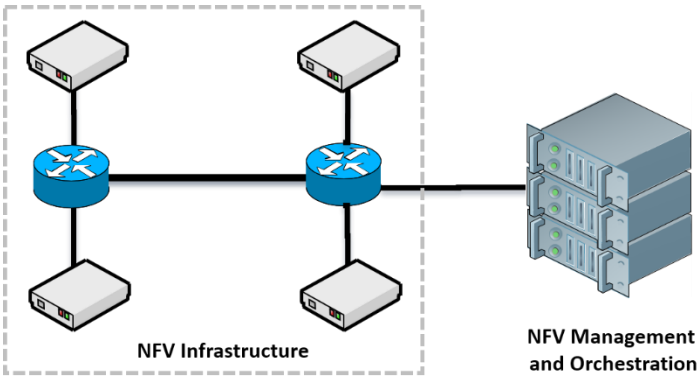

Figure 19. Infrastructure Implementation and relevant Components

The equipment used for the server is a Mini-ITX i5 3Ghz with 8GB and 1TB SDD, operating with Ubuntu Server 18.04 for Hyper-V support. A VM runs the OpenStack Ocata controller along with the OSM Release FIVE. The VM runs as a KVM guest and has three network interfaces. The first one delivers VIM communication between the OpenStack controller and the NFVI. The second interface is for the VNF lifecycle management by the OSM. The last interface gives connectivity and allows remote configuration of the server.

The compute nodes are a group of different RPi versions (3B, 3B+, 4B and Zero W) to allow a low-cost and power-constrained test scenario. All RPis operate with Raspbian Release from 06-2019. Openstack Nova and Neutron components were executed on bare-metal configuration. All RPis also have three interfaces. The first one delivers VIM communication between the local Nova and Neutron components and the centralised OpenStack controller for physical resource management. The second interface is exclusively for VNF lifecycle management by the OSM. The last interface acts as an access point for NS users, providing connectivity to the deployed VNFs through virtualised interfaces.

Figure 20 illustrates the implementation of a VoIP call and its components using different VNFs, describing the implementation of the physical infrastructure and virtualised resources. The centralised server locally runs a SIP Core using a Kamailio server VNF with native RTP

support. The first compute node has two VNFs: a router and a DNS. The second and the third compute nodes run one AP VNF each, connecting with the first compute node for service integration. Both APs were RPi 3B+ and the Router was a RPi 3B. All VNFs ran using VMs with one vCPU, 128MB of memory and 4GB of storage.

Figure 21 shows all virtual links and VNFs in the implementation. The NS only needs to know the configuration of the virtualised network since the NFV environment abstracts the details about the physical topology.

To simulate a VoIP application, we developed a specific traffic generator that will be used in several tests and simulations during the project. The generator sends a packet with a voice payload of 20 Bytes and a sample period of 20 ms. The time between packets is 0.02 seconds, using a normal distribution and a standard deviation of 0.0038 seconds. With such configuration, the workload follows the specifications of Codec G.729, with sample size of 10 bytes (80 bits) and operating in intervals of 10 ms, allowing a bit rate of 8 Kbps, which is the expected data rate for this type of codec.

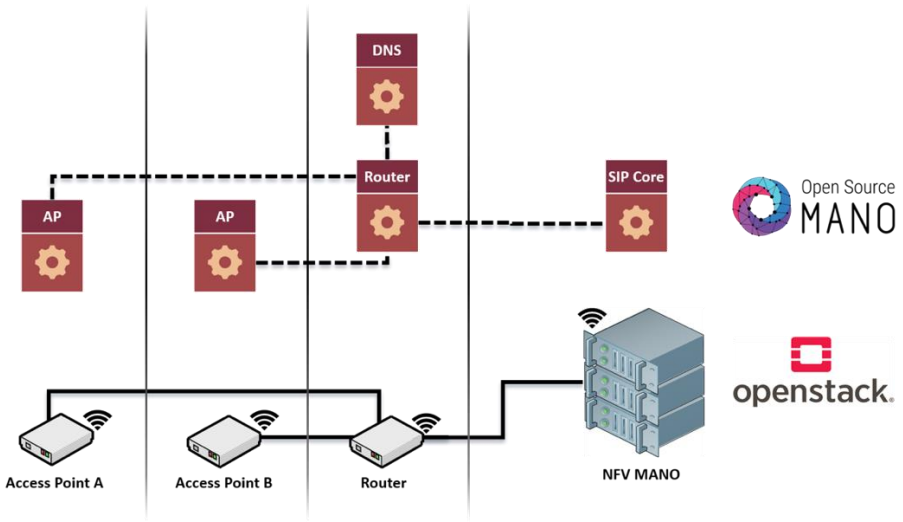

Figure 20. VoIP Infrastructure Implementation

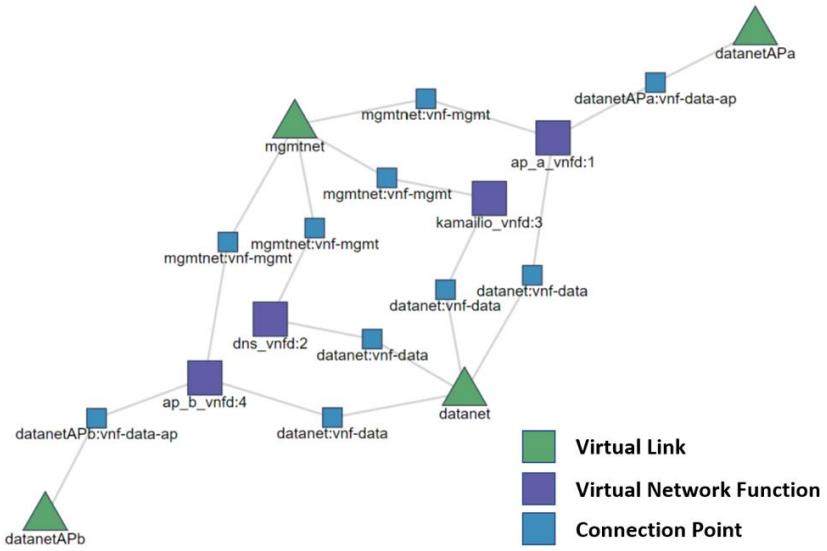

Figure 21. Virtual Components for the VoIP and their Virtual Links

Studies point that VoIP calls have an average duration of 181 seconds with a standard deviation of 2.3 seconds. Also, each base station serves an average of 750 devices, or 250 per sector taking into account LTE 3-sector coverage base stations. Because of infrastructure limitations, which made not possible the execution of concurrent calls, our analysis uses a sample of 250 calls with only one call at a given time. The test results consolidate more than 12 hours of gathered data.

To monitor the traffic in the deployed scenario, all Neutron vSwitches operated in promiscuous mode, allowing the attachment of one ObF on each compute node as previously described in Figure 18. This configuration enabled the monitoring of VNF communication over the infrastructure. On the centralised server, PEvF was executed as an autonomous performance evaluation function, providing data analysis and using the distributed ObFs as monitors. The centralised PEvF also executed a local ObF to observe the infrastructure communication.

The source code of PEvF was developed using Python 3.7. The ObF implementation is based on `Pcapy`, which allows an interface with `libpcap` to enable capturing packets over the network. Through the tests, the distributed ObFs gathered information directly from the vSwitches to avoid reading data not related to the virtualised environment. With this configuration, external communication mediums such as Wi-Fi did not interfere. Another ObF running alongside PEvF monitored the infrastructure traffic.
During the evaluation, the ObFs operated in SBA mode to cover the whole system and provide a better visualisation of how the physical infrastructure reacts to the virtualised functions when running common tasks. The monitoring provided three perspectives of the system:

- The first perspective gathered both SIP Core and DNS VNFs management data sent to the NFV MANO Kamailio SIP server.
- The second perspective monitored the actual data transferred between the APs by sniffing traffic sent and received by them through RTP, to better visualise how data plane communication.
- The third perspective monitored all data sent and received between the Router RPi and the server; this last test helps to evaluate if an actual service running at the virtualised layer could

impact the infrastructure, only needing a comparison with data collected by the second perspective to develop this observation. The ObFs monitored all different perspectives for 4 minutes, gathering data over the same circumstances to avoid the influence of unknown conditions.

Figure 22 shows the results of the first perspective. The DNS server only produced a single packet of 1758 bits (around 220 bytes) at 5 seconds, which served the SIP Core request for name resolution. SIP produced only a few packets along the connection, peaking at 9587 bits (around 1.20 kilobytes) at 26 seconds and sending only five messages through the whole monitoring process. The results of the first perspective show that both DNS and SIP are performing appropriately, and such small transmission packets cannot confirm any details about the performance of virtualised functions over the provided infrastructure, but further analysis and comparisons with the two other perspectives can give a better view of the implementation results.

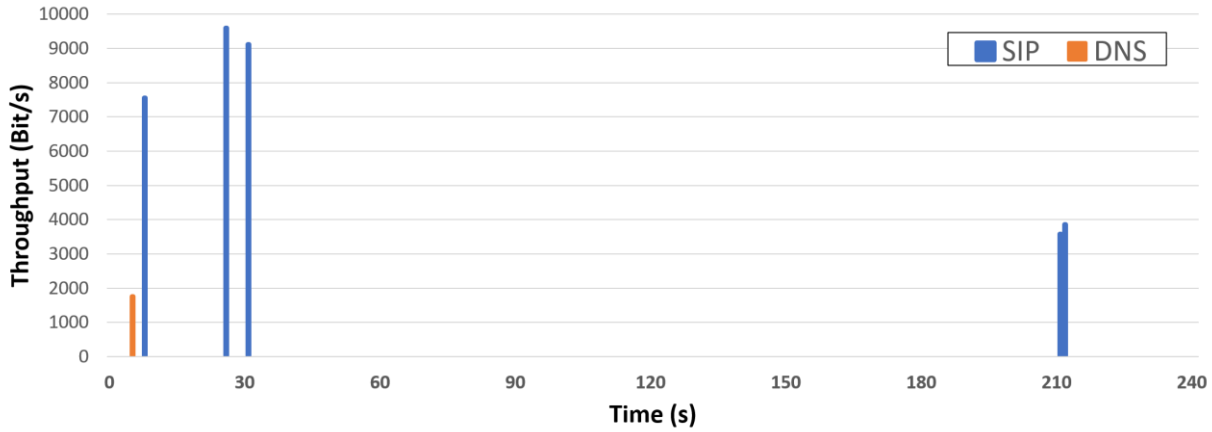

Figure 22. SIP Core and DNS VNFs management data sent to NFV MANO

Figure 23 shows the results of the second perspective, with details about how user data fluctuated over the connection. Between the start of the call at 34 seconds and around 171 seconds, both RX and TX kept a stable throughput between 14558 bits (1.82 kilobytes) and 15358 bits (1.92 kilobytes), with RX keeping this pace until the end of the call at 215 seconds. TX, on the other hand, heavily fluctuated its throughput between 174 seconds and 195 seconds, but it also ended the data flow around the 215 seconds mark with stable throughput.

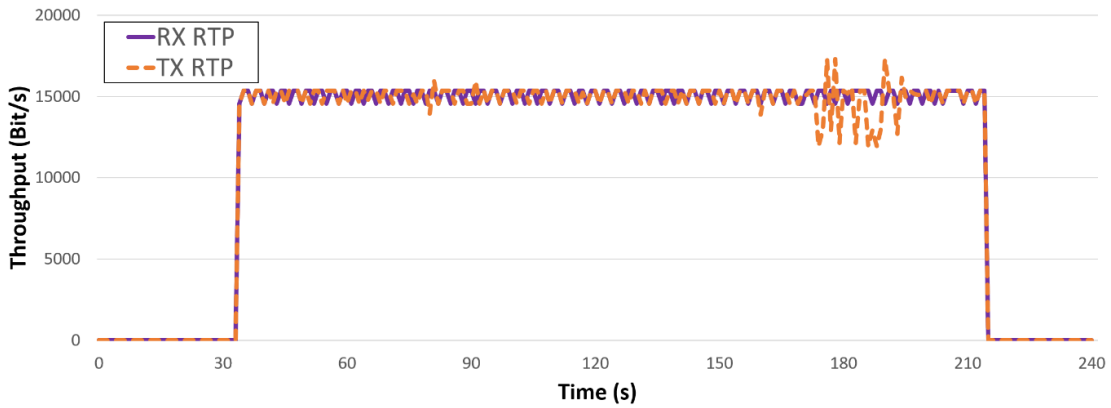

Figure 23. VOIP Data RX and TX

Figure 24 shows the results for the third perspective, which monitors NFVI control data flow between the RPi and the server. Openstack uses an AMQP (Advanced Message Queuing Protocol) message queue to communicate with all compute nodes available, implemented with RabbitMQ on the infrastructure. Analysed data from the test results show a little increase in the total throughput of both channels every 60 seconds, which follows the default temporisation defined at OpenStack for health check of virtualised infrastructure. Data sent by the RPi had an average of 7269.46 bits/s and peaked with 151021 bits at the 121 seconds mark, while data sent by the server reached an average of 457382 bits and peaked with 103220 bits at the same 121 seconds mark. The first health check, near the 57 seconds mark, happened a little after SIP established the call and was the smallest for both channels. The second health check happened around the 119 seconds mark and had a substantial increase on both channels regarding the first health check period. The last health check happened around the 208 seconds mark, right in the middle of the high throughput fluctuation perceived by TX. Comparing with the previous health check period, this last one had a small decrease on data sent by the RPi while maintaining almost the same data rate sent by the server. During almost the whole connection, the RPi kept a higher throughput than the server.

Comparing the results from Figure 23 and Figure 24, appears that the NFVI health check messages start to increase when compute nodes are performing virtualisation functions, but there is also a reasonably stable ratio of control traffic even during high fluctuations of data throughput at the virtualised layer.

Figure 25 shows the results for the VoIP call but only considering 180 seconds. User data represents the total VoIP data sent and received from one of the APs. This data is the information sent and received by the VNFs which perform the functional tasks of the virtualised environment and produce the traffic monitored by the distributed ObFs. Another ObF was deployed at the central server to monitor both the locally deployed VNFs and infrastructure communication. As described before, OpenStack runs a defaulthealth check of virtualised infrastructure every 60 seconds. This traffic variation is shown clearly in Figure 25.

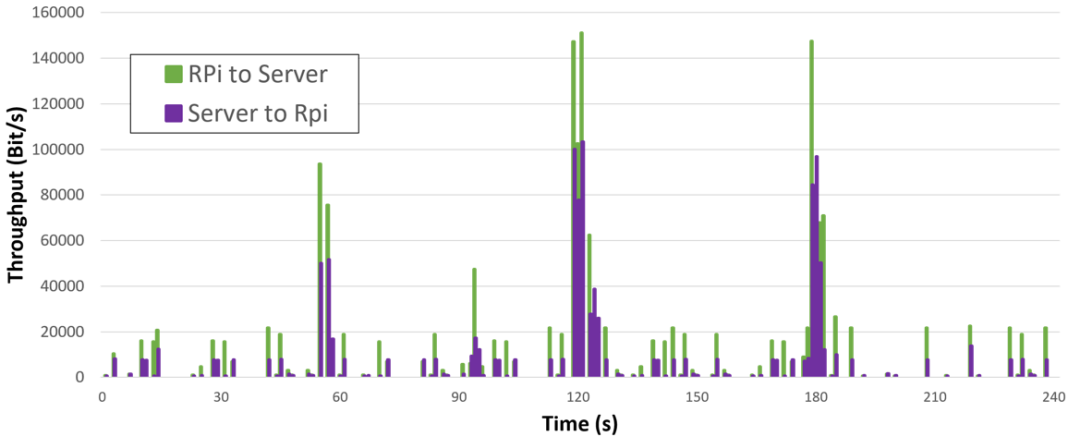

Figure 24. NFVI Control Data Flow during the Tests

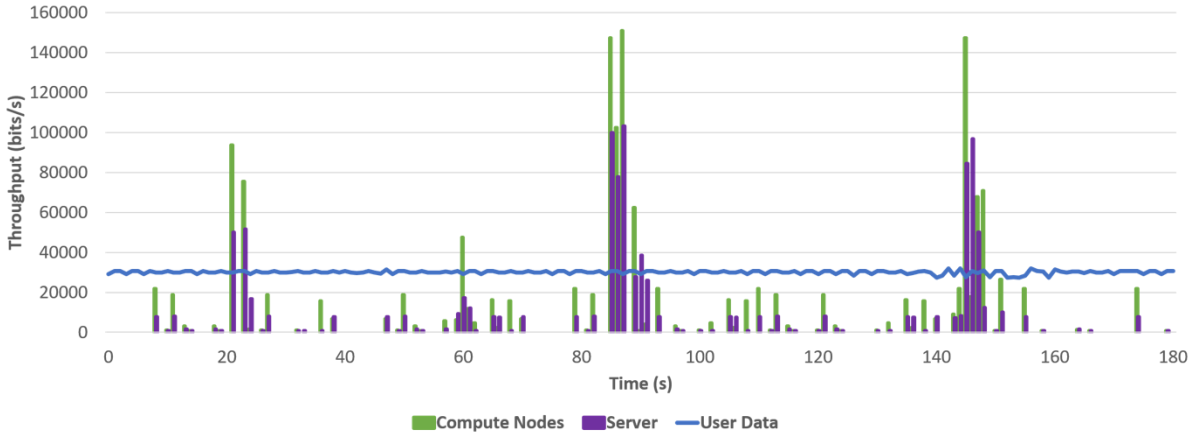

Figure 25. Throughput Comparison During a VoIP Call

By monitoring the synthetic VoIP workload generated and the infrastructure behaviour, the deployed ObFs gathered 12 hours of traffic. The average throughput of actual data was 30.08 kbps, with a peak of 32.85 kbps. This throughput complies with expected bandwidth for Codec G.729 when using RTP if considering the two active channels (TX and RX).

Figure 26 shows an example of the number of packets transferred at each moment per call. As shown by the red line, an average call delivered around 50.33 packets per second, with a total of 9060 packets (N.1) for a 180 seconds long call. At the same situation, it produces an average of just 72 dropped packets (N.3), accounting for only 0.795% of the total number of packets. These observations did not take into account the failures generated by the Wi-Fi connection.

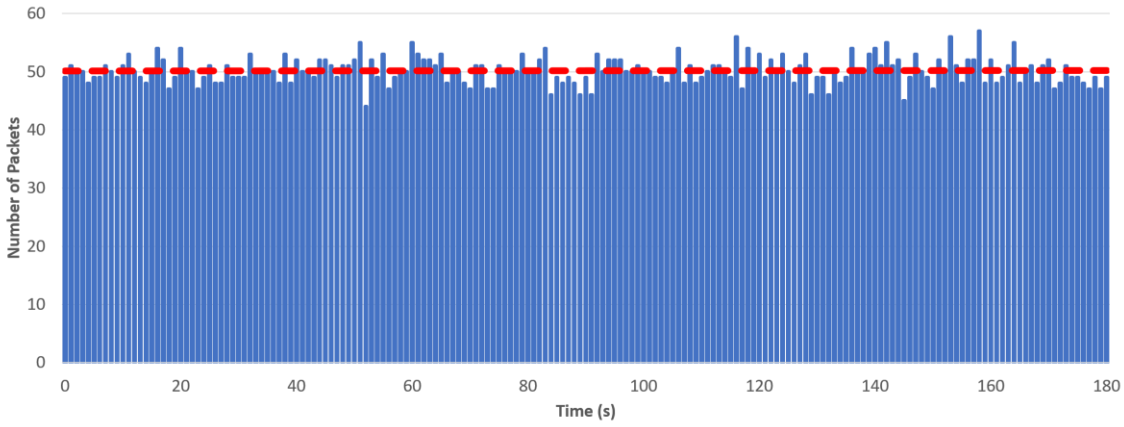

Figure 26. Number of Packets Transferred During a VoIP Call.

During an average call of 180 seconds, the actual data corresponded to around 680.73 KB (N.2), while the infrastructure produced 313.42 KB, which is equivalent to 31.52% (N.6) of total information transmitted within the environment. This situation shows a large amount of additional data produced to perform a VoIP call. On real-world scenarios, around 30% of additional traffic generated would seriously compromise any usefulness of the NFV. To further investigate this case, it is relevant to consider the OpenStack health checks because they significantly increase the amount of information transferred within the infrastructure. As shown before, during health checks, the infrastructure throughput dramatically increases. A health

check lasts for an average period of 3 seconds and increases the amount of data transferred by the infrastructure by more than 600% if considering a 3 seconds' window.

When separating the communication between normal periods and health check periods, the infrastructure corresponds for an average of 14.90% of the total traffic on the first case and 80.81% on the later as shown by Figure 27. In such periods, the throughput of the virtualised environment for the actual data keeps a stable rate. For the test scenario, health check data transmission happens only 5% of the total time, and even with such small windows, it considerably increased the total amount of traffic. For future tests and to any actual deployment with virtualised environments, if the amount of additional data interferes on the data communication channel or overpasses the control channel bandwidth, the waiting window of Openstack health checks could be adjusted in a more intelligent way, for example considering the type of application and bandwidth constraints.

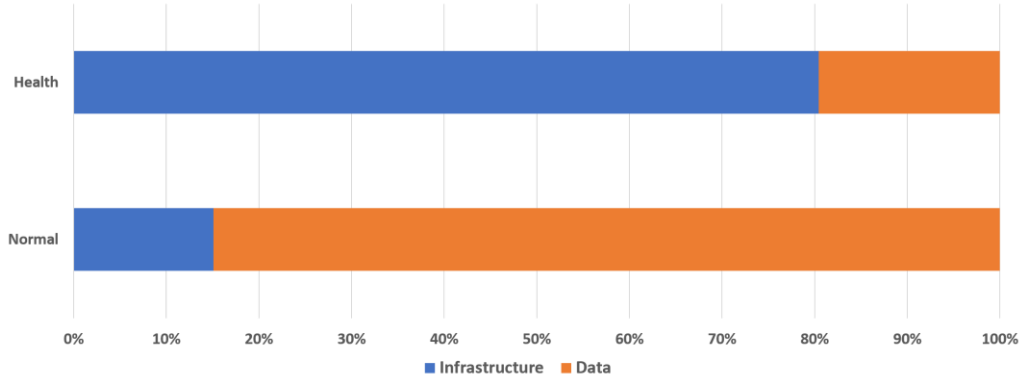

Figure 27. Difference between the amount of data transferred on normal and health check periods.

As shown by the results, VoIP communication reached the expected throughputs. Table 2 presents the final results collected by the ObFs and analysed by the PEvF. The metrics Packet Count (N.1) and Octet Count (N.3) are easily verified by looking at gathered data. Dropped Packet Count (N.3), in this implementation, required a deeper verification to ignore packets dropped between two VNFs because the physical medium could impact this metric. The use of ObF attached to all virtual switches allowed to circumvent this problem by detecting the packets regardless of the transmission medium, because of its direct connection to VNFs. In this experiment, the Errored Packet Count (N.4) metric was affected by Wi-Fi's redundancy checks, which did not allow a direct count in comparison with the NFV. Because of that, in this implementation, we did not analyse this metric. As can be seen, PEvF calculates the Amount of User Data (N.5) and the Amount of Control Data (N.6) by consolidating all received information from the distributed ObFs.

Resource	Code	Metric	Value
Network	N.1	Packet Count	9060
	N.2	Octect Count	680730
	N.3	Dropped Packet Count	72
	N.4	Error Packet Count	-
	N.5	Amount of User Data	68.48 %
	N.6	Amount of Control Data	31.52%

Table 3. NFV Performance Metrics for a VoIP Call of 180 Seconds

The above results allowed the analysis of control data in a 5GC NFV environment for a VoIP application. These results will support the analysis of viability for using technologies such as LoRA in the SUAV/VNF scenario. This analysis will be presented in D5.3.

4.3.2 Resource Management and traffic classes

4.3.2.1 Motivation and requirements

The 5G-RANGE technology aims to enable high-speed Internet access in remote areas. Even though 5G-RANGE permits a much higher speed than available technology in long ranges, the resulting bandwidth is still significantly lower compared to the one available in other 5G scenarios. The implication is that it is expected that in general the 5G-RANGE access link will commonly be the bottleneck of the end to end communication.

The available capacity provided by the 5G-RANGE technology is likely to change significantly. Multiple factors can affect the capacity available for a given UE, including obstacles, mobility, availability of the white space spectrum and others. This implies that it is likely that the available capacity will vary during the lifetime of a communication.

In particular, the variability due to the TVWS spectrum availability is a fundamental design choices made in 5G-RANGE in the early phases. As described in Deliverable 2.1 "Application and Requirements Report", 5G-RANGE will opportunistically use unlicensed TVWS. The trade-off resulting from the selected approach were documented in D2.1 which states that: "Although a non-licensed TVWS carrier can add a significant amount of additional spectrum, it can be difficult to guaranty system reliability and even continuity of service by relying only on TVWS. This occurs due to opportunistic access to TVWS band which requires sophisticated mechanisms of sensing and decision, leading to dynamic and fragmented bandwidth."
These aspects were incorporated into the 5G-RANGE design in the following requirements presented in D2.1:

- Req-F.m.3 - 5G-RANGE system shall provide aggregation of one licensed carrier for broadcast and common control information, and at least one non-licensed TVWS carrier for dedicated user traffic

- Req-F.m.5 - Subject to regional regulatory requirements, 5G-RANGE shall support dynamic spectrum allocation for the TVWS component carrier

- Req-F.m.6 - Non-licensed TVWS component carrier shall support non-continuous bandwidth selected from within a specified spectrum window.

- Req-F.m.11 - 5G-RANGE system shall assure high spectrum efficiency and expected QoS even with the uncertainty regarding TVWS availability.

So, one of the challenges of 5G-RANGE is how to use this fragmented capacity resulting from the opportunistic use of unlicensed TVWS. This really depends on the application using the capacity. There are elastic applications that are able to cope with large changes in available capacity while others cannot. In order to properly understand the requirements in the case of 5G-RANGE, we will next consider the different 5G-RANGE use cases presented in D2.1.

We observe that the following 5G-RANGE use cases presented in D2.1 comprise a mix of different types of traffic with different requirements in the same use case. We next describe different 5G-RANGE use cases and identify the inelastic traffic (with more stringent QoS requirements) and elastic traffic.

In the **Agribusiness and Smart Farming for Remote Areas** use case described in section 5.1 of D2.1 the traffic includes a real time traffic component including crop monitoring and remote maintenance and diagnosis (including the automation of flows, irrigation, etc). This traffic has high reliability requirements and the end to end latency is required to be below 50 ms (see Table 5 on section 5.1.2 of D2.1. Also, in the agribusiness use case, there is also a background traffic component with less stringent requirements. This includes cattle counting, production traceability, video snapshots, etc. Then, as stated in section 5.1 of D2.1, while most traffic for this use case will not have mission critical requirements, there is a fraction of the traffic that will have it with the aforementioned requirement on end-to-end latency.

For the **Voice and Data connectivity over long distances for remote areas** use case, the types of traffic are described in section 5.2 of D2.1 as follows: "Some types of services that will be evaluated in this use case are enhanced web browsing, email, VoIP, multimedia on the web, audiographics conference, file sharing and interactive video on demand. The QoS requirements will vary for the different applications. For example, in conventional text and data networking, delay requirements are the least stringent. The response time in these types of applications can increase from 2 to 5 seconds before becoming unacceptable. For interactive applications, the overall round-trip delay needs to be short to give the user an impression of real-time responses. Normally, a maximum value of 0.1 to 0.5 seconds is required to accomplish this goal."
So, similarly to the previous use case, this use case also exhibits a mix of real-time and an background traffic with very different QoS requirements.

So, the challenge for the 5G-RANGE architecture is how to support the different classes of traffic, in particular in a way that surges in background traffic does not negatively affect the performance of real-time traffic. However, it is also important to provide fully utilization of the available capacity, since it is a scarce and valuable resource. So, background traffic should be able to use as much capacity as possible as long as it does not negatively affect the performance of the real-time traffic class.

Summarizing, the the mechanism to manage the available capacity resources in 5G-RANGE must be able to cope with:

- Due to the opportunistic use of TVWS, noise, obstacles and other factors, available capacity may change dramatically.
- Several 5G-RANGE use cases exhibit a mix of traffic with different QoS requirements, including real time traffic with stringent QoS requirements and background traffic with less demanding requirements. Variations in the amount of background traffic should not negatively affect the performance of real-time traffic but also bulk traffic should be able to use as much capacity as possible as long it does not interfere with the real time traffic performance, to achieve high utilization of resources.

It is then critical to develop the mechanisms to manage the capacity of the 5G-RANGE according to these goals. Ideally, the mechanism proposed to manage available capacity in 5G-RANGE should accommodate changes in the available capacity and surges in background traffic while honouring the performance requirements of real-time traffic.

The mechanisms for managing the capacity of the 5G-RANGE access link will then have fulfil the following requirements:

- must be able to manage both the uplink and the down link channels;
- must be able to support a mix of traffic types, including mixes of latency sensitive traffic and other traffic with less stringent requirements (like VoIP and file sharing);
- must support changes in available capacity;
- must provide high utilization of the link capacity;
- must support encrypted traffic (currently, about half of the traffic in the Internet is encrypted[7], so any solution must accommodate this).
- its deployment must only require upgrading the equipment in the 5G-RANGE access side (the adoption of a solution that required upgrading of the other endpoint in the Internet is much more challenging, as they do not have the incentives for its adoption).
- No changes in the wire protocol. In order to maximize deployability, the proposed solutions must not introduce any change in the wire protocol, as middleboxes such as firewalls may block/drop/strip unknown options/extensions/protocols.

4.3.2.2 Proposed architecture

The selected approach is to leverage on the existing inelastic/TCP/scavenger traffic classes and to include additional mechanisms to enable a deployment model solely based on the UE connected to the 5G-RANGE access.

The proposed approach is to define different traffic classes according to the congestion controller they use. Currently in the Internet, there are roughly three types of traffic classes widely used, namely inelastic traffic (traffic that does not respond to congestion signals, typically UDP traffic), best effort traffic (i.e. TCP traffic) and less than best effort/scavenger traffic (LEDBAT).

These approaches rely on having different congestion controllers on the senders, so that different traffic classes respond differently when there is congestion. This translates that when the capacity of the bottleneck is full, the inelastic traffic continues to send at the same rate, TCP traffic reduces its rate and scavenger traffic reduces earlier and more aggressively than TCP traffic. Such approach could be suitable for 5G-RANGE as it allows to have different types of traffic (i.e. VoIP traffic can be inelastic and file transfer can be scavenger preventing file sharing traffic to disturb VoIP traffic).

Also, having scavenger traffic can make a better use of available capacity, since it tends to fill available capacity with additional scavenger traffic than later on can be reduced if the capacity available decreases.
This approach can be suitable for the uplink traffic, as the UE connected to the 5G-RANGE access link can use different congestion controllers for different types of traffic/applications.

[7] https://www.wired.com/2017/01/half-web-now-encrypted-makes-everyone-safer/

The problem is how to deal with downlink traffic. The congestion controller is located in the sender, and this would require changing the other endpoint which is sending traffic, which would break our deployability requirement.

As part of the 5G-RANGE architecture, we propose to define the mechanisms in the 5G-RANGE UE to implement a receiver-based scavenger congestion controller. In order to do this, we need to enable the receiver to control the sending rate of the sender.
In particular, we present rLEDBAT, a receiver-based implementation of a LEDBAT++ congestion control algorithm for TCP. rLEDBAT performs all the congestion window calculations in the receiver.
In the following sections we describe the proposed mechanism. We start by providing background information of the selected congestion control algorihtm LEDBAT++ in the next section and next we dive in the specifics of the design of the proposed rLEDBAT mechanism.

4.3.2.3 Background on LEDBAT++

LEDBAT++ is a Less-than-Best-Effort (LBE) congestion-control algorithm that reacts both to packet loss and to delay variations. LEDBAT++ controls the sending rate through the calculation of a congestion window (cwnd). LEDBAT++ updates the cwnd based on delay variations and packet loss. With respect to packet loss, LEDBAT++ reacts by reducing the cwnd to half of its value when a loss is detected.

With respect to delay variations, LEDBAT++ aims for a pre-defined queueing delay target T (currently set to 60 ms). LEDBAT++ continuously estimates the current queueing delay, (qd). If the current queueing delay (qd) is larger than the target queuing delay T, LEDBAT++ multiplicatively decreases the congestion window. Conversely, if the delay is smaller than the target, LEDBAT++ additively increases the congestion window.

LEDBAT estimates the current queueing delay (qd) by subtracting the base round-trip-time (RTTb) from the current one-way delay (RTTc). The base RTT is calculated as the minimum RTT observed in the last 10 minutes of the lifetime of the communication. The current delay is the RTT currently measured in the communication. Both values are filtered to eliminate noise by taking the minimum of the last n values. The current queueing delay is then calculated as:
qd = RTTc - RTTb

So if the LEDBAT++ congestion window cwnd was last updated at time t0 and its current value is then cwnd0. At time t1 a packet is received, updating the value of the measured queueing delay qd. Then the LEDBAT++ updates cwnd as follows:

if qd<T, then cwnd1=cwnd0+ a*MSS/cwnd0

if qd > T, then cwnd1=cwnd0*b (once per RTT)

with MSS being the Maximum Segment Size of the TCP connection, and a and b being the additive increase and multiplicative decrease parameters respectively. This base algorithm results that while the queueing delay is below the target T, the congestion window increases by a * MSS per RTT, while if the queueing delay is above the target T, the congestion control window is multiplied by b, 0 < b < 1, at most once per RTT.

By selecting the a parameter to be 1/16 (significantly lower than the additive increase of 1 used by standard TCP congestion control), LEDBAT++ bounds the share of capacity that it is able to seize even when it is working in loss base congestion avoidance (as opposed to when it is working on delay based congestion avoidance).

By using multiplicative decrease (instead of the additive decrease used in LEDBAT), LEDBAT++ overcomes the late-comer advantage observed in LEDBAT and achieves inter-LEDBAT++ fairness.

LEDBAT++ performs a slow start increase, similar to the one of standard TCP (but with a smaller multiplicative increase factor) at the beginning of the connection. Also, LEDBAT++ performs periodic slowdowns to obtain more accurate measurements of the base RTT. LEDBAT++ is specified in [30] and it is available in Windows 2019 Server.

4.3.2.4 rLEDBAT mechanism design

rLEDBAT provides the mechanisms to implement an LBE congestion control algorithm at the receiver-end of a TCP connection. The rLEDBAT receiver controls the sender's rate through the Receive Window announced to the sender in the TCP header.

rLEDBAT assumes that the sender is a standard TCP sender. rLEDBAT does not require any rLEDBAT-specific modifications to the TCP sender. The envisioned deployment model for rLEDBAT is that the clients implement rLEDBAT and this enables rLEDBAT in communications with existent standard TCP senders. In particular, the sender implements RFC793 and it also implements the Time Stamp Option as defined in RFC7323. Also, the sender implements some of the standard congestion control mechanisms, such as Cubic (RFC8312) or New Reno (RFC5681).

rLEDBAT does not defines a new congestion control algorithm. The LBE congestion control algorithm executed in the rLEDBAT receiver is defined in [30]. Because rLEDBAT assumes a standard TCP sender, the sender will be using a "best effort" congestion control algorithm (such as Cubic or New Reno). Since rLEDBAT uses the Receive Window to control the sender's rate and the sender calculates the sender's window as the minimum of the Receive window and the congestion window, rLEDBAT will only be effective as long as the congestion control algorithm executed in the receiver yields a smaller window than the one calculated by the sender. This is normally the case when the receiver is using an LBE congestion control algorithm. Irrespectively of which congestion control algorithm is executed in the receiver, an rLEDBAT connection will never be more aggressive than standard TCP since it is always bounded by the congestion control algorithm executed at the sender.

rLEDBAT is essentially composed of three types of mechanisms, namely, those that provide the means to measure the round trip time, mechanisms to detect packet loss and the means to manipulate the Receive Window to control the sender's rate. We describe them next.

Controlling the receive window

rLEDBAT uses the Receive Window (RCV.WND) of TCP to enable the receiver to control the sender's rate. RFC793 defines that the RCV.WND is used to announce the available receive

buffer to the sender for flow control purposes. In order to avoid confusion, we will call fc.WND the value that a standard RFC793 TCP receiver calculates to set in the receive window for flow control purposes. We call rl.WND the window value calculated by rLEDBAT algorithm and we call RCV.WND the value actually included in the Receive Window field of the TCP header. For a standard TCP receiver, RCV.WND == fc.WND.

In the case of rLEDBAT receiver, the rLEDBAT receiver sets the RCV.WND to the minimum of rl.WND and fc.WND, honoring both.

When using rLEDBAT, two congestion controllers are in action in the flow of data from the sender to the receiver, namely, the congestion control algorithm of TCP in the sender side and the LBE congestion control algorithm executed in the receiver and conveyed to the sender through the RCV.WND. In the normal TCP operation, the sender uses the minimum of the congestion window cwnd and the receiver window RCV.WND to calculate the sender's window SND.WND. This is also true for rLEDBAT, as the sender is a regular TCP sender. This guarantees that the rLEDBAT flow will never transmit more aggressively than a TCP flow, as the sender's congestion window limits the sending rate. Moreover, because LEDBAT++ is designed to react earlier and more aggressively to congestion than regular TCP congestion control, the rl.WND contained in the RCV.WND field of TCP will be in general smaller than the congestion window calculated by the TCP sender, implying that the rLEDBAT congestion control algorithm will be effectively controlling the sender's window.
In summary, the sender's window is: SND.WND = min(cwnd, rl.WND,fc.WND)

Avoiding window shrinking

The LEDBAT++ algorithm executed in a rLEDBAT receiver increases or decreases the rl.WND according to congestion signals (variations on the estimations of the queueing delay and packet loss). If the new value for rl.WND is smaller than the current one then directly announcing it in the RCV.WND may result in shrinking the window, i.e., moving the right window edge to the left. Shrinking the window is discouraged as per RFC793, as it may cause unnecessary packet loss and performance penalty, so the rLEDBAT receiver avoids shrinking the receive window.

In order to avoid window shrinking, upon the reception of a data packet, the announced window can be reduced in the number of bytes contained in the packet at most. This may fall short to honor the new calculated value of the rl.WND. So, in order to reduce the window as dictated by the rLEDBAT algorithm, the receiver SHOULD progressively reduce the advertised RCV.WND, always honoring that the reduction is less or equal than the received bytes, until the target window determined by the rLEDBAT algorithm is reached. This implies that it may take up to one RTT for the rLEDBAT receiver to drain enough in-flight bytes to completely close its receive window without shrinking it. This is more than sufficient to honor the window output from the LEDBAT/LEDBAT++ algorithms since they only allows to perform at most one multiplicative decrease per RTT.

Window Scale Option

The Window Scale (WS) option defined in RFC7323 is a mean to increase the maximum window size permitted by the Receive Window. The use of the WS option implies that the

changes in the window are expressed in the units resulting of the WS option used in the TCP connection. This means that the rLEDBAT client will have to accumulate the increases resulting from the different received packets, and only convey a change in the window when the accumulated sum of increases is equal or higher than one unit used to express the receive window according to the WS option in place for the TCP connection.

Changes in the receive window that are smaller than 1 MSS are unlikely to have any immediate impact on the sender's rate, as usual TCP segmentation practice results in sending full segments (i.e., segments of size equal to the MSS). So, accumulating changes in the receive window until completing a full MSS in the sender or in the receiver makes little difference.

Current WS option specification defines that allowed values for the WS option are between 0 and 14. Assuming a MSS around 1500 bytes, WS option values between 0 and 11 result in the receive window being expressed in units that are about 1 MSS or smaller. So, WS option values between 0 and 11 have no impact in rLEDBAT.

WS option values higher than 11 can affect the dynamics of rLEDBAT, since control may become too coarse (e.g., with WS of 14, a change in one unit of the receive window implies a change of 10 MSS in the effective window).
For the above reasons, the rLEDBAT client sets WS option values lower than 12. Additional experimentation is required to explore the impact of larger WS values in rLEDBAT dynamics.

Note that the recommendation for rLEDBAT to set the WS option value to lower values does not precludes the communication with servers that set the WS option values to larger values, since the WS option value used is set independently for each direction of the TCP connection.

Measuring the Round Trip Time

LEDBAT++ measures base and current RTT to estimate the queueing delay. In the next sections we describe how rLEDBAT mechanisms enable the receiver to measure the RTT.

The original LEDBAT algorithm uses the one-way delay to estimate the queuing delay. We have encountered a number of issues when attempting to measure the one-way delay in TCP, which resulted in deferring the recommendation of the use of one-way delay to estimate the queuing delay in rLEDBAT for the future, when additional research is done in this space.
LEDBAT++ uses the round trip time (RTT) to estimate the queueing delay. In order to estimate the queueing delay using the RTT, the rLEDBAT receiver estimates the base RTT (i.e., the constant components of the RTT) and also measures the current RTT. By subtracting these two values, we obtain the queuing delay to be used by the rLEDBAT controller.
LEDBAT++ discovers the base RTT (RTTb) by taking the minimum value of the measured RTTs over a period of time. The current RTT (RTTc) is estimated using a number of recent samples and applying a filter, such as the minimum (or the mean) of the last k samples. Using the RTT to estimate the queueing delay has a number of shortcomings and difficulties that we discuss next.

The queuing delay measured using the RTT includes also the queueing delay experienced by the return packets in the direction from the rLEDBAT receiver to the sender. This is a fundamental limitation of this approach. The impact of this error is that the rLEDBAT controller

will also react to congestion in the reverse path direction which results in an even more conservative mechanism.
In order to measure the RTT, the rLEDBAT client MUST enable the Time Stamp (TS) option. By matching the TSVal value carried in outgoing packets with the TSecr value observed in incoming packets, it is possible to measure the RTT. This allows the rLEDBAT receiver to measure the RTT even if it is acting as a pure receiver. In a pure receiver there is no data flowing from the rLEDBAT receiver to the sender, making impossible to match data packets with acknowledgements packets to measure the RTT, as it is usually done in TCP for other purposes.

Depending on the frequency of the local clock used to generate the values included in the TS option, several packets may carry the same TSVal value. If that happens, the rLEDBAT receiver will be unable to match the different outgoing packets carrying the same TSVal value with the different incoming packets carrying also the same TSecr value. However, it is not necessary for rLEDBAT to use all packets to estimate the RTT and sampling a subset of in-flight packets per RTT is enough to properly assess the queueing delay. The RTT is then be calculated as the time since the first packet with a given TSVal was sent and the first packet that was received with the same value contained in the TSecr. Other packets with repeated TS value are not used for the RTT calculation.

Several issues must be addressed in order to avoid an artificial increase of the observed RTT. Different issues emerge depending whether the rLEDBAT capable host is sending data packets or pure ACKs to measure the RTT. We next consider the issues separately.

Measuring RTT sending pure ACKs

In this scenario, the rLEDBAT node (node A) sends a pure ACK to the other endpoint of the TCP connection (node B), including the TS option. Upon the reception of the TS Option, host B will copy the value of the TSVal into the TSecr field of the TS option and include that option into the next data packet towards host A. However, there are two reasons why B may not send a packet immediately back to A, artificially increasing the measured RTT. The first reason is when A has no data to send. The second is when A has no available window to put more packets in-flight. We describe next how each of these cases is addressed.

The case where the host B has no data to send when it receives the pure Acknowledgement is expected to be rare in the rLEDBAT use cases. rLEDBAT will be used mostly for background file transfers so the expected common case is that the sender will have data to send throughout the lifetime of the communication. However, if, for example, the file is structured in blocks of data, it may be the case that seldom, the sender will have to wait until the next block is available to proceed with the data transfer and momentarily lack of data to send. To address this situation, the filter used by the congestion control algorithm executed in the receiver discards the larger samples (e.g. a min filter would achieve this) when measuring the RTT using pure ACK packets.
The limitation of available sender's window to send more packets can come either from the TCP congestion window in host B or from the announced receive window from the rLEDBAT in host A. Normally, the receive window will be the one to limit the sender's transmission rate, since the LBE congestion control algorithm used by the rLEDBAT node is designed to be more restrictive on the sender's rate than standard-TCP. If the limiting factor is the congestion window in the sender, it is less relevant if rLEDBAT further reduces the receive window due

to a bloated RTT measurement, since the rLEDBAT is not actively controlling the sender's rate. Nevertheless, the proposed approach to discard larger samples would also address this issue.

To address the case in which the limiting factor is the receive window announced by rLEDBAT, the congestion control algorithm at the receiver SHOULD discard the RTT measurements done using pure ACK packets while reducing the window and avoid including bloated samples in the queueing delay estimation. The rLEDBAT receiver is aware whether a given TSVal value was sent in a pure ACK packet where the window was reduced, and if so, it can discard the corresponding RTT measurement.

Measuring the RTT sending data packets

In the case that the rLEDBAT node is sending data packets and matching them with pure ACKs to measure the RTT, a factor that can artificially increase the RTT measured is the presence of delayed Acknowledgements. According to the TS option generation rules, the value included in the TSecr for a delayed ACK is the one in the TSVal field of the earliest unacknowledged segment. This may artificially increase the measured RTT.

If both endpoints of the connection are sending data packets, Acknowledgments are piggybacked into the data packets and they are not delayed. Delayed ACKs only increase the RTT measurement in the case that the sender has no data to send. Since the expected use case for rLEDBAT is that the sender will be sending background traffic to the rLEDBAT receiver, the cases where delayed ACKs increase the measured RTT are expected to be rare.

Nevertheless, for those measurements done using data packets sent by the rLEDBAT node matching pure ACKs sent from the other endpoint of the connection, they will result in an increased RTT. The additional increase in the measured RTT will range between the transmission delay of on packet and 500 ms. The reason for this is that delayed ACKs are generated every second data packet received and not delayed more than 500 ms. The rLEDBAT receiver discards the RTT measurements done using data packets from the rLEBDAT receiver and matching pure ACKs, especially if it has recent measurements done using other packet combinations. Also, applying a filter that discard larger samples would also address this issue (e.g. a min filter).

Detecting retransmissions and packet losses

The rLEDBAT receiver is capable of detecting retransmitted packets in the following way. We call RCV.HGH the highest sequence number correspondent to a received byte of data (not assuming that all bytes with smaller sequence numbers have been received already, there may be holes) and we call TSV.HGH the TSVal value corresponding to the segment in which that byte was carried. SEG.SEQ stands for the sequence number of a newly received segment and we call TSV.SEQ the TSVal value of the newly received segment.

If SEG.SEQ < RCV.HGH and TSV.SEQ > TSV.HGH then the newly received segment is a retransmission. This is so because the newly received segment was generated later than another already received segment which contained data with a larger sequence number. This means that this segment was lost and was retransmitted.

The proposed mechanism to detect retransmissions at the receiver fails when there are window tail drops. If all packets in the tail of the window are lost, the receiver will not be able to detect a mismatch between the sequence numbers of the packets and the order of the timestamps. In this case, rLEDBAT will not react to losses but the TCP congestion controller at the sender will, most likely, reduce its window to 1MSS and take over the control of the sending rate, until slow start ramps up and catches the current value of the rLEDBAT window.

5 Conclusions

This document described the network-level architecture for the 5G-RANGE project which has been designed following the 3GPP guidelines. It means that the project has closely followed the specifications provided by the 3GPP Release 15 documents, as well as their evolution in Release 16. The document presents and details how to integrate the 5G-RANGE radio access network and also the terminal equipment within the 3GPP 5G overall architecture and service framework to enable Internet access and global connectivity through an operator network.

As a complement to the network-level architecture provided by deliverable D2.2 [16] this deliverable is providing a solution based on operation 2 of the reference model (i.e. no legacy infrastructure) allowing for a direct connection between the end node and the 5G network core. The specification [5] defines the proper way to do this both for 3GPP access networks and for non-3GPP access networks. At the current time of the project, with the standardization work still in an early stage, it is unclear whether the 5G-RANGE access network will receive the consideration of a 3GPP access network. Therefore, this deliverable contemplates the different models defined by 3GPP to support the connectivity of the UE with the 5G core network, considering the 5G-RANGE access network as a 3GPP access, an untrusted non-3GPP access, and a trusted non-3GPP access.

Apart from the architectural design, based on functional entities and reference points, this document also identifies the different protocol stacks required at the different entities (considering as well the user equipment that will be connected to the access network). The document also presents in this section, an analysis of control data in the 5G core and a proposal for resource management and traffic classes.

The implementation and validation of the architectural components of the 5G-RANGE network level are described in the companion deliverable D5.3 [42].

References

[1] 5th Generation (5G). Why do we need 5G. ETSI 2018. Available at: [https://www.etsi.org/technologies-clusters/technologies/mobile/5g] last access January 2020.

[2] RP-161249. Architecture configuration options for NR. Available at: [http://www.3gpp.org/ftp/TSG_RAN/TSG_RAN/TSGR_72/Docs/RP-161249.zip] last access January 2020.

[3] 3GPP Specification Status Report. Available at [https://www.3gpp.org/DynaReport/status-report.htm] last access January 2020

[4] 3GPP Release 15 website. Available at [http://www.3gpp.org/release-15] last access January 2020.

[5] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "System Architecture for the 5G System; Stage 2", 3GPP Technical Specification 23.501, version 16.3.0, December 2019.

[6] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "Procedures for the 5G System; Stage 2", 3GPP Technical Specification 23.502, version 16.3.0, December 2019.

[7] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "Architecture enhancements for non-3GPP accesses", 3GPP Technical Specification 23.402, version 16.0.0, June 2019.

[8] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "Enhancement for Unmanned Aerial Vehicles (UAVs)", 3GPP Technical Specification 22.829, version 17.1.0, September 2019.

[9] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "Unmanned Aerial System (UAS) support in 3GPP", 3GPP Technical Specification 22.125, version 17.1.0, December 2019.

[10] 3rd Generation Partnership Project. UAS – UAV. Available at [https://www.3gpp.org/uas-uav] last access January 2020.

[11] 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; "IP Multimedia Subsystem (IMS); Stage 2", 3GPP Technical Specification 23.228, version 15.3.0, September 2018.

[12] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, Internet Engineering Task Force, June 2002.

[13] J. Suarez, I. Vidal, J. Garcia-Reinoso, F. Valera and A. Azcorra, "Exploring the use of RPAs as 5G points of presence," 2016 European Conference on Networks and Communications (EuCNC), Athens, 2016, pp. 27-31.

[14] Li, Bin, Zesong Fei, and Yan Zhang. "UAV communications for 5G and beyond: Recent advances and future trends," IEEE Internet of Things Journal 6.2 (2018): 2241-2263.

[15] M. Mohammad, et al. "Beyond 5G with UAVs: Foundations of a 3D wireless cellular network," IEEE Transactions on Wireless Communications 18.1 (2018): 357-372.

[16] D2.1 Application and requirements report. 5G-RANGE public deliverable. Available at: [http://5g-range.eu/wp-content/uploads/2018/04/5G-Range_D2.1_Application_Requirement_Report_v1.pdf] last access January 2020.

[17] C. Rametta and G. Schembra, "Designing a Softwarized Network Deployed on a Fleet of Drones for Rural Zone Monitoring," Future Internet, vol. 9, no. 1, p. 8, Mar. 2017.

[18] D2.2 Architecture, system and interface definitions of a 5G for Remote Area network. 5G-RANGE public deliverable. Available at: [http://5g-range.eu/wp-content/uploads/2018/04/5G-Range_D2.2-Architecture_5G_RemoteArea.pdf] last access January 2020.

[19] B. Nogales, V. Sanchez-Aguero, I. Vidal, F. Valera, and J. Garcia-Reinoso. 2018. A NFV system to support configurable and automated multi-UAV service deployments. In Proceedings of the 4th ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications (DroNet'18). ACM, New York, NY, USA, 39-44.

[20] Ivan Vidal, Borja Nogales, Francisco Valera, Luis F. Gonzalez, Victor Sanchez-Agüero, Eduardo Jacob and Cristina Cervelló-Pastor, "A Multi-site NFV Testbed for Experimentation with SUAV-based 5G Vertical Services". IEEE Access, June 2020.

[21] Borja Nogales, Iván Vidal, Víctor Sánchez, Francisco Valera, Luis F. González, Arturo Azcorra. OSM PoC 10 Automated Deployment of an IP Telephony Service on UAVs using OSM. Available at: [https://osm.etsi.org/wikipub/index.php/OSM_PoC_10_Automated_Deployment_of_an_IP_Telephony_Service_on_UAVs_using_OSM] last access June 2020.

[22] B. Nogales, V. Sanchez-Aguero, I. Vidal, F. Valera, Adaptable and Automated Small UAV Deployments via Virtualization, Sensors, 2018, no. 12: 4116.

[23] 5G-PPP Vision and Mission (last access on February 2020): https://5g-ppp.eu/about-us/

[24] 3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3, 3GPP Technical Specification 24.501, version 15.1.0, September 2018.

[25] Perkins, C.E., Belding-Royer, E.M., Das, S., "Ad hoc on-demand distance vector (AODV) routing", Internet Engineering Task Force, RFC 3561 (Experimental), July 2003.

[26] Clausen, T., Dearlove, C., Jacquet, P., and U. Herberg, "The Optimized Link State Routing Protocol Version 2", Internet Engineering Task Force, RFC 7181, April 2014.

[27] J. Postel, "User Datagram Protocol", Internet Engineering Task Force, RFC 768, August 1980.

[28] J. Postel, "Transmission Control Protocol", Internet Engineering Task Force, RFC 793, September 1981.

[29] H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson. "RTP: A Transport Protocol for Real-Time Applications", Internet Engineering Task Force, RFC 3550, July 2003.

[30] T. Dierks, E. Rescorla, The Transport Layer Security (TLS) Protocol Version 1.2, Internet Engineering Task Force, RFC 5246, August 2008.

[31] J. Iyengar and M. Thomson, "QUIC: A UDP-Based Multiplexed and Secure Transport," Internet Engineering Task Force, Internet- Draft, November 2016 (expires June 1, 2017).

[32] M. Belshe, R. Peon, M. Thomson, "Hypertext Transfer Protocol Version 2 (HTTP/2)", Internet Engineering Task Force, RFC 7540, May 2015

[33] Object Management Group (OMG). Data Distribution Service (DDS), version 1.4; OMG: Needham, MA, USA, 2015.

[34] LoRa Alliance (2018), LoRaWAN 1.0.3 specification, Technical Report 1 [Online]. Available at: [https://lora-alliance.org/sites/default/files/2018-07/lorawan1.0.3.pdf] last access January 2020.

[35] U. Raza, P. Kulkarni, and M. Sooriyabandara. “Low Power Wide Area Networks: An Overview”, IEEE Communications Surveys and Tutorials, 19(2):855–873, 2017.

[36] M. Centenaro, L. VangeLista, A. Zanella and M. Zorzi. “Long-Range Communications in Unlicensed Bands: the Rising Stars in the IoT and Smart City Scenarios”. Technical Report 1, 2012.

[37] P. Sethi and S.R. Sarangi. “Internet of Things: Architectures, Protocols, and Applications”. 2017.

[38] B. Reynders, W. Meert and S. Pollin. “Range and Coexistence Analysis of Long Range Unlicensed Communication”. In 23rd International Conference on Telecommunications (ICT), May 2016.

[39] L. Vangelista, A. Zanella, M. Zorzi. “Long-Range IoT Technologies: The Dawn of LoRaTM”. Future Access Enablers for Ubiquitous and Intelligent Infrastructures: First International Conference, pages 51–58, 2015.

[40] P. Balasubramanian et. al., LEDBAT++: Congestion Control for Background Traffic, Internet draft, work in progress, 2019

[41] D5.1 Initial version of the Network-level Architecture and Procedures. 5G-RANGE public deliverable. Available at: [http://5g-range.eu/wp-content/uploads/2018/04/D5.1-Initial-Version-of-Network-Architecture.pdf] last access January 2020.

[42] D5.3 Final report on Network-level mechanisms implementation. 5G-RANGE public deliverable.

