Metro-Haul

METRO High bandwidth, 5G Application-aware optical network, with edge storage, compute and low latency

Grant No. 761727

Deliverable D2.1
Definition of Use Cases, Service Requirements and KPIs

Abstract

This deliverable describes the work done in T2.1 and is based on selected Service Use Cases, which are representative of the three service types as defined by 3GPP. Their analysis has resulted in the derivation of the main KPIs that the Metro-Haul architecture needs to satisfy. Each Use Case has been decomposed into their components and requirements, which are then mapped onto the end-to-end network infrastructure.

The holistic view of all the possible services running over a metro network as envisioned by Metro-Haul has allowed to describe the main characteristics such a network must have, which will enable the services as envisioned by the 5G-PPP. This document provides detailed analysis of the service KPIs and subsequent requirements have provided key objectives, definition, and apportionment of the service KPIs for the Metro-Haul network vision.

Document properties

Status and Version:		
Date of issue:	12.02.2018	
Distribution:	Public	
Author(s):	**Name**	**Partner**
	Marco Quagliotti (Editor)	TIM
	Albert Rafel (Co-Editor)	BT
	Oscar Gonzales De Dios	TID
	Víctor López	TID
	Rafael López da Silva	TID
	José Alberto Hernández	UC3M
	Manuel Urueña	UC3M
	David Larrabeiti	UC3M
	Ignacio Martín	UC3M
	Neelakandan Manihatty Bojan	UNIVBRIS
	Ftoeini Ntavou	UNIVBRIS
	Emilio Hugues Salas	UNIVBRIS
	Luis Velasco	UPC
	Marc Ruiz	UPC
	Gabriel Junyent	UPC
	Jaume Comellas	UPC
	Miquel Garrich	UPCT
	Francisco Javier Moreno	UPCT
	Pablo Pavón	UPCT
	Jorge E. López de Vergara Méndez	Naudit HPC
	Mike Parker	LEX
	Mayur Channegowda	ZEETTA NETWORKS

	Chigo Okonkwo	TU/e
	Nicola Calabretta	TU/e
	Nicklas Eiselt	ADVA DE
	Achim Autenrieth	ADVA DE
	Bogdan Mihai-Andrus	ADVA DE
	Omran Ayoub	POLIMI
	Massimo Tornatore	POLIMI
	Guido Maier	POLIMI
	Francesco Musumeci	POLIMI
Checked by:	Daniel King (OLD DOG CONSULT) and João Pedro (CORIANT PT)	

Revision History

Revision	Date	Responsible	Comment
0.0	30.11.2017	Editors	Initial version
1.0	31.01.2018	Editors	First complete revision provided by Daniel King
2.0	06.02.2018	Editors	Second complete revision provided by João Pedro
3.0	12.02.2018	Editors	Final document
3.1	15.02.2018	Editors	Editorial bugs pointed out by Johannes have been fixed

Executive Summary

The Metro-Haul architecture must be capable of meeting future service requirements, encompassing the design of all-optical metro nodes (including full compute and storage capabilities), which interface effectively with both 5G access and multi-terabit per second elastic core networks.

This work aims to deliver a preliminary set of requirements and KPIs for the Metro-Haul architecture. It provides inputs to the project to define solutions and strategies compliant with the fulfilment of the requirements coming from three different and complementary lines of analysis.

The first line of analysis adopts a general "high level" perspective, it considers the trends emerging in recent technology and market outlook surveys to infer the fallouts on metro-Haul architecture. The second line of analysis assumes, as input, the services demanded by customers belonging to a wide and relevant set of verticals. A number of Use Cases are defined and characterized in high detail, and their impact on Metro-Haul have been quantitatively evaluated. The third line of analysis reviews the impact of 5G Radio Network Access on Metro-Haul architecture. This thread is orthogonal as it assumes 5G as the main access technology and analyses the possible 5G transport architectures and how they can map onto the Metro-Haul architecture.

Putting together the results of these three approaches, we are able to provide the preliminary requisites for the Metro-Haul network including edge nodes and optical transport for both data and control planes. The requirements also include the networking infrastructure, as well as Data Centre resources (computing and storage). We are assuming that the network nodes are going to be implemented according to the CORD model.

After a short introduction on Metro-Haul architecture, this summary presents the results achieved in the three analyses separately, after it provides a combined view on Metro-Haul requirements.

The Metro-Haul Architecture

The Metro-Haul architecture (Figure ES-1) describes the network elements necessary to interconnect a number of access network nodes (AMEN) with the core network nodes (MCEN) via a high capacity dynamic and flexible optical network. Transmission distances range between 50km and 200km. Access network is out of scope of Metro-Haul but 5G RAN and fixed fibre networks (e.g., GPON) are the access solutions presumed widely deployed at the time Metro-Haul will be introduced.

To meet the Metro-Haul objectives, both storage and computing resources will be necessary both in the AMEN and the MCEN. These resources may be used to host one or more controllers and/or network monitoring function that will have open interfaces capable of connecting to the various network components and nodes in the optical transport layer.

Control and monitoring functions will be required to manage the Metro-Haul network architecture. The network is assumed adopting a Software Define Networking (SDN) system that controls the assignment of both the computing and the transmission resources to provision the requested services. The SDN controllers interact with the data plane via the monitoring system. The control plane is then completed by the service orchestrator.

Summary of Metro-Haul requirements from a high-level approach

We have performed a high-level analysis to assess the requirements and dimensioning for Metro-Haul nodes and network links. Our high-level approach has estimated trends in data-centre (DC) capacities, end-user bandwidth and data consumption, as well as other technology trends (e.g. 5G RN, SDN/NFV, fog computing) to assess the expected load on the Metro segment of the overall telecoms network. Alongside the significant capacity and performance improvements that 5G networking is expected to cause, in a complementary fashion we can also expect that evolution of DCs will also play a key role in defining the Metro-Haul node and link requirements. From various sources we consider that DC-DC traffic will increase by a factor of x4 from 2015 to 2020 (i.e. 31.9% CAGR) with associated processor metrics (storage, big data volumes) increasing by similar amounts. We assume a similar 32% CAGR for the subsequent 5 years to 2025, so that we therefore expect a 4x4 (i.e. overall x16) increase in traffic load across the metro network over the next decade.

We have defined the Figure ES-1 for the Metro-Haul network architecture, with the dimensioning of the key nodes and links (bandwidth pipes) within the 2020-2025 timeframe as indicated. We see a progression from micro-tiny-small DCs on the left-hand side, as they appear within the fixed and wireless access network, to small-medium-large DCs in the regional networking location (i.e. the main Metro-Haul scope), to the large-hyper-scale DCs appearing in the photonic core. The key parameters are: storage (RAM), bandwidth pipes (Gb/s-Tb/s), and processing power (number of multi-cores) that are expected at each location.

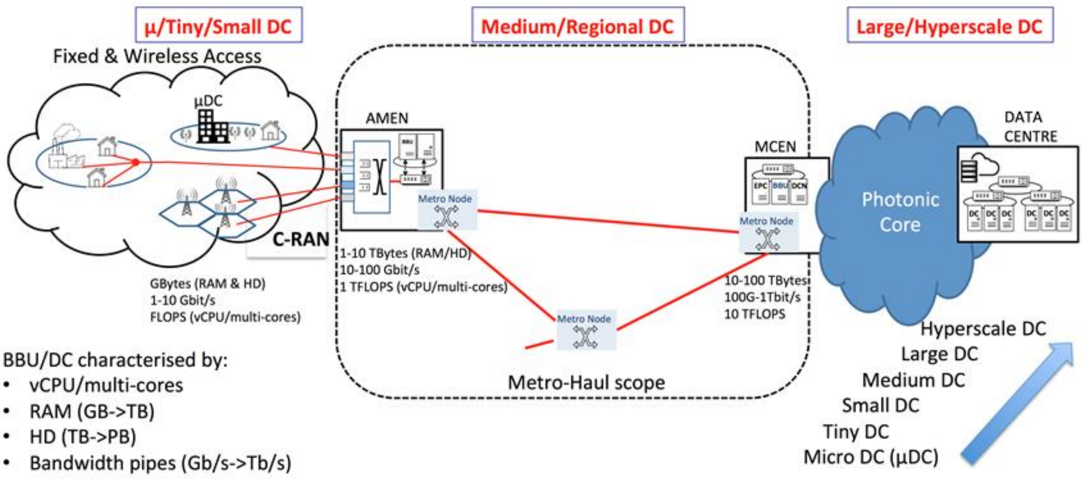


Figure ES-1 – Distribution and dimensioning of DCs across the Access-Metro-Core networks.

5G KPIs suggest end-user bandwidths of 1-10 Gb/s, with latencies of 1-5 ms, along with provision of the other 5G services (mIoT, CriC/URLLC, and eMBB), which places great demands on the CUs of the C-RAN architecture. We have assumed that the compute (storage, processing, data-pipes etc.) dimensioning of a 2017 C-RAN solution will increase by a factor x4 by 2020, and a further factor x4 (i.e. x16 overall) by 2025.

Overall, during the Metro-Haul timeframe we expect the Access-Metro Edge Node (AMEN) to be served by pipes of 10-100 Gb/s, e.g. 10 Gb/s on the access interfacing side, and aggregated pipes of up to 100 Gb/s on the Metro-Haul interface side. The AMEN DC would have storage capability measured in Terabytes (TB) of RAM and HD capacity, with processing power of the order of 1 TFLOPS on a multi-core chip (e.g. the Intel Core i9 Extreme Edition chip, which consists of 18 vCPUs). The Metro-Core Edge Node (MCEN) that interfaces with the photonic core can be expected to require another factor x10 increase in performance capabilities, as compared with the AMEN. Specifically, MCEN will be served by 100 Gb/s pipes from other AMEN and intermediate Metro-Haul nodes, and interface via 400G/s and 1 Tb/s pipes (i.e. comprising super-channels) into the photonic core. The storage capacity of the MCEN will be measured in the 10s of TB of RAM and HD, with associated compute resource of the order of 10 TFLOPS. Current topology descriptors are not anticipated to greatly change over the Metro-Haul timeframe, e.g. in terms of the average node degree sizes for the MCEN and AMEN, and intrinsic connectivity.

Description of relevant Use Cases and their requirements for Metro-Haul

A set of relevant 5G service Use Cases have been selected within a wide set available in literature. The Use Cases are analysed following a procedure defined to enable the derivation of the requirements imposed on the Metro-Haul architecture resulting from supporting the service use cases.

In a first step each Use Case is described and classified considering its belonging to a specific Vertical (e.g. automotive, Industry 4.0, etc.), the class of the communication services (CoS) involved. As far as CoSs, the 3GPP definition is assumed: Enhanced mobile broad band (eMBB), massive internet of things (mIoT) and ultra-reliable low latency connections (URLLC).

Secondly each Use case is analysed in terms of its Components and Functionalities and these elements are mapped onto the Metro-Haul architecture (i.e. functionalities and data flows are situated on nodes and links). In addition the requirements in terms of KPIs for the services associated to the Use Case (data rate, latency, availability and others for a total of 13 parameters) are assessed and assigned.

Finally, the Critical Tasks (i.e., those tasks that require challenging performances by the network) are identified and an evaluation of the amount of required resources has been performed (resources mean node throughput and link capacity in the network, computation and storage memory in the Data Centre part of the node). Furthermore, requirements of the identified Critical Tasks in terms of Monitoring, Data Analytics, Control, and Orchestration are also collected to complete the set of Use case requirements.

The Use Case selected are listed in table ES1 where the first column indicates the associated Vertical, the name of the Use Case in the second, and the CoS in the third. A Use Case can be characterized by more than one service. A service can be associated to more than one CoS in case a single CoS does not satisfy all the requirements for that service (e.g. service requiring high bandwidth and low latency is characterized by eMBB + URLLC).

Table ES-1 – 5G use case selected and analysed.

Vertical	Use Case	CoS
Media and Entertainment	Content Delivery Network	eMBB
	Live TV Distribution	eMBB + URLLC
	6DoF Virtual Reality	eMBB + URLLC
	Crowdsourced Video Broadcasts	URLLC
Cloud Services	Service Robotics	eMBB + URLLC
	Enterprise Access with NG Ethernet	BB[1] + URLLC
Utilities	mIoT Utility Metering	mIoT
Automotive	ITS and Autonomous Driving	eMBB + URLLC

[1] Enterprise Access with Next Generation Ethernet Use Case requires high level of performance on data rate, latency and reliability, but it does not involve the “mobile” access part of 5G infrastructure. The CoS is then assigned as a general indication (in particular assigning the class Broad Band, obtained removing enhanced Mobile from eMBB) as CoS eMBB is conceived for services relying on a mobile access.

Industry 4.0	Smart Factories	eMBB + URLLC + mIoT
Public Safety and Environment	RT LL Object Tracking and Security -	URLLC
Operator orientated	Secure SDN Control. Video Distribution	BB + URLLC [2]

Among the Use Cases included in Table ES-1 two are planned to be demonstrated over a Metro-Haul proof of concept during the last period of the project. These Use Case demonstrations will be design with minute detail and extensively described in WP5. All use case are End User orientated except the last one which is operator orientated.

Next, a brief description of one Use Case is included in this Executive Summary as exemplary means to illustrate the work, which shows the derivation of the Metro Haul requirements. Details of the other Use Cases is left in the main body of this report. The selected Use Case is the Content Delivery Network, which belongs to the Media and Entertainment Vertical type.

Content Delivery Networks Use Case

Most of the traffic in today's network is generated to support video delivery. Video-on-Demand is expected to represent approximately 85% of the global consumer traffic by 2018, and 75% of global mobile data traffic by 2019. The increasing importance of video traffic is shifting the network architecture host-centric (connection-centric) to content-centric, and this trend is expected to grow further in the coming 5G era. To cope with the continuous increase of video traffic, network operators are pushed to enhance their CDNs. A promising solution is to equip network edge nodes with storage and computing capabilities, allowing them to store (cache) and deliver contents as, e.g. popular video contents.

In the vision of Metro-Haul network architecture, the CDN caches popular video contents at AMENs and MCENs. The main functionalities to be located at AMENs and MCENs to enable efficient video caching and delivery are Storage capacity, Video interfaces to deliver video contents and Traffic inspection capabilities.

To provide a quantitative estimation of the network requirements the following case study characterizing a video-content catalogue of 10,000 contents with a popularity modelled with a Zipf distribution has been assumed. The contents sizes follow a power-law distribution with average size of 4.5 GB. The reference dense-urban scenario of the case study is characterized by 10,000 users connected to the AMEN with an average video delivery bit-rate of 8 Mbps. Furthermore, 6 AMENs are connected, on average, to the MCEN. Figure ES-2 depicts this scenario in which, for simplicity, the Metro network has only 6 AMEN and one MCEN connected by a mesh of links in the optical layer. Central office hosting MCEN is supposed to perform only the role of Metro Core node and not also the one of Access Metro. In Figure ES-2 the functionality of Content Cache and Deep Packet Inspection (DPI) is highlighted with specific icons on AMEN and MCEN, but they are supposed to be performed by generic storage and virtual machine service resources available in the node (i.e. not by specialized pieces of equipment).

[2] For Use Case Secure SDN Controlled Video Distribution, concerning CoS, it is the same as in the previous note.

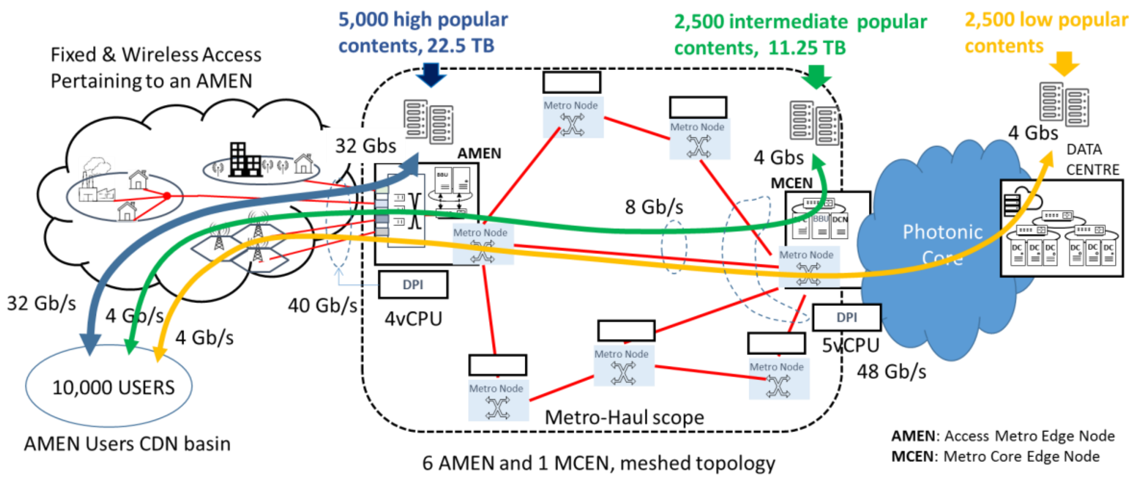


Figure ES-2 –CDN Use case mapped on the Metro-Haul architecture.

Cache in AMEN satisfies 80% of the video requests of users connected to it, as for the cache located at the MCEN, it satisfies 50% of the requests not satisfied by AMEN caches. The cache is calculated considering the popularity distribution and guaranteeing the 80% of the requests served by content cached in AMEN node, and another 10% of requests by contents cached in MCEN node. (The remaining 10% is served by Video Server in the Core.) These assumptions give 5000 content items in the AMEN, corresponding to 22.5 TB, and 2500 contents in the MCEN, resulting in 11.25 TB (the other 2,500 contents is in the core).

Under these assumptions, and assuming also that half of the user are connected to the CDN in the peak hour, the corresponding video interface capacity at AMEN is 32 Gb/s. The video-delivery interface capacity required at the MCEN to deliver videos to users is calculated taking into consideration the number of users of the 6 AMENs associated to it. It results in 4 Gb/s for each flow from the MCEN to each AMEN, and 24 Gb/s exiting from the MCEN in total. Note that this video interface capacity is sufficient to accommodate all video streaming traffic.

Data rate required on optical connections within the Metro networks to carry user video requests delivered by AMEN, but originated in MCEN and in the core depend on the networking of traffic (paths used in the mesh topology by MCEN-AMEN flows). With such considerations, and taking into account that path is direct (as in the example of Figure ES-2) or at most has a length of 2 or 3 hops, data rates on optical connections are of the order of one or few 10 Gb/s.

AMEN inspects all video traffic generated by users connected to it while the MCEN inspects only the video traffic of requests not served by the AMEN. As for the computing capacity needed for video traffic inspection at AMEN and MCEN, the computing capacity of DPI is assumed 10 Gb/s per core CPU. This gives the need in the example of 4 vCPUs in the AMEN and 5 vCPUs in the MCEN.

The critical tasks are defined as the jobs required by a Use case that, due to their stringent and highly demanding requirements, are critical to be performed in the Metro-Haul architecture. For CDN Use Case the following critical tasks can be identified:

- UHD/4K/8K video streaming for fixed and mobile users,
- Video-traffic inspection/monitoring and analysis and cache reconfiguration,
- Peak-hour video streaming and flash crowd phenomenon (i.e., sudden increases in request arrival rate for a new viral video content).

Assigning to each identified critical task the requirements for the Metro haul architecture in terms of networking, storage and computation capacities calculated as in the case study presented above, Table ES-2 can be provided.

Table ES-2 –CDN service requirements mapped over the Metro-Haul architecture.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
UHD/4K/8K video streaming	32Gb/s[3]	24Gb/s [4]	Nx10Gb/s	22.5TB	11.25TB	-	-
Video traffic inspection, analysis and cache reconfiguration	40Gb/s	48Gb/s	10Gb/s	-	-	4 vCPU	5 vCPU
Peak-hours/Flash crowd phenomenon	32Gb/s	24Gb/s		-	-	-	-

As previously mentioned, the critical task Peak-hours Flash crowd-phenomenon shares the video interface capacity with UHD/4K/8K video streaming task. However, it doesn't require any additional storage capacity for caching, as the new popular contents are stored replacing less popular contents. Figures in Table ES-2 are applicable to one Content provider serving 10,000 customers, assuming MCEN does not perform the role of AMEN. For scenarios in which Central Offices hosting MCEN include also AMEN functionalities (i.e., connection with the access), values in columns reporting MCEN requirements would result greater.

As far as functionalities required for Monitoring Data analytics, Management Control and end-to-end orchestration, the ones included in table have been identified.

Table ES-3 –CDN functionality requirements for Metro-Haul architecture.

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
UHD/4K/8K video streaming		Fast recovery mechanisms (protection and restoration)
Video traffic inspection, analysis and cache reconfiguration	Traffic monitoring for early detection of new popular videos	Local and Global reconfiguration of virtual cache
Peak-hours/Flash crowd phenomenon	Traffic monitoring for fast detection of a flash crowd phenomenon	Local and global reconfiguration of virtual cache

Split options for 5G RN transport architecture and their impact on Metro Haul

In traditional 4G (LTE) architectures, the Base-Band Unit (BBU) in a Distributed Radio Access Network (D-RAN) architecture is co-located with the Remote Radio Head (RRH). A transport network physically

[3] Shared with *Peak-hours/Flash crowd* task.
[4] Shared with *Peak-hours/Flash crowd* task.

connects the BBU with the Evolved Packet Core (EPC) server in the Point-Of-Presence (PoP). In a D-RAN, every base station has a dedicated housing facility, which is not shared with other base stations. Hence power consumption as well as investment and maintenance costs increase linearly with the number of base stations. A Centralised RAN (C-RAN) is an architecture where multiple BBUs are placed in a single physical location, called "BBU pool", to enable sharing the expensive BBU equipment among different RRHs, thus leading to cost and power efficiencies. The connection between BBU pool and RRHs is called "fronthaul" and requires relatively high and constant bit-rate as it transports digitized time domain IQ data. Moreover, since some of the functionality in the BBU have very stringent latency requirements, the distance limit for the fronthaul application is around 20km.

As the fronthaul requirements in terms of capacity and latency will become even more stringent in 5G Radio Networks, more flexible distributions of the functions in the 3GPP LTE RAN protocol stack between BBU and RRH are being investigated to relax bandwidth and latency requirements. The various options of this flexible distribution are referred to as *RAN functional splits* and are reported in Figure ES-3. Note that, in this new context, BBUs may be divided into Radio Unit (RU), DU (Distributed Unit), and CU (Centralised Unit).

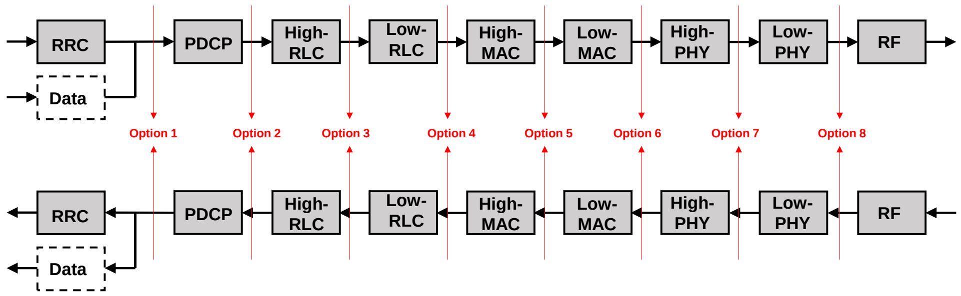


Figure ES-3 – BBU Split Options.

The different split options allows the mapping of the different BBU functions into the RU/DU/CU functional blocks as shown in Figure ES-4

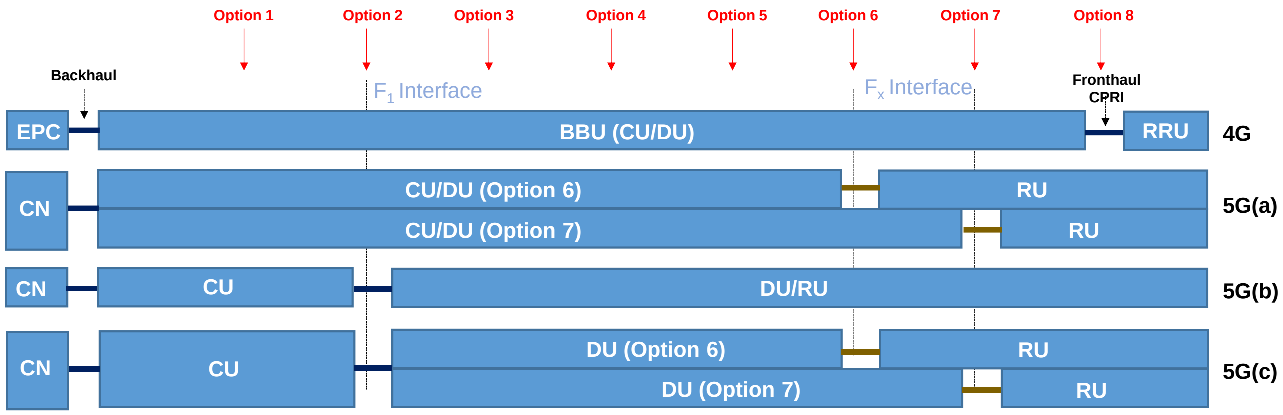


Figure ES-4 – Mapping of CU and DU functions according to the split points. 5G(a) low layer split (F1); 5G(b) high layer split (FX); 5G(c) cascaded split.

In the various split options, latency and bandwidth requirements vary significantly. For example the splits up to the MAC-PHY (i.e., options 1 to 6) have similar relatively-low bandwidth requirements, while in split options at the physical layer (i.e., from option 7 onwards) the required bandwidth increases significantly. Similarly, for latency requirements, while options 4 – 8 require latency values in the order of a few hundred microseconds, latency requirements for splitting options 1 – 3 are relaxed down to a range of several milliseconds.

From these latency and bandwidth requirements follows the adequacy to either “non-real-time” or “real-time” transport type of services, and it also follows the possibility to deploy different versions of 5G backhaul/fronthaul transport network architectures. In the main body of this deliverable we overview several of them, classifying them as network architectures using *i)* low-layer split options, *ii)* high-layer split options and *iii)* cascaded split option. Depending on the exact latency requirement, and also on the fibre distances involved in the metro network, the DU/CU box can be located either in the Local Exchange (AMEN in Metro-Haul) or somewhere deeper in the Metro Network (e.g. a MCEN) as summarized in Table ES-4.

Table ES-4– Location of functionality for the 5G network architectures.

Split Option	DU Functionality	CU Functionality	NC Functionality
Low Layer	RU	AMEN	MCEN
High Layer	RU	AMEN or MCEN	MCEN
Cascaded	AMEN	MCEN	MCEN

We have estimated the amount of GOPS (and hence CPUs) required by the functionality in the CU and DU in order to serve a given area, depending on the different split options, which has been used to dimension the AMEN and MCEN in terms of processing capacity according to the different 5G RN transport architectures, which are based on the chosen split options. We have considered two different geographical areas, namely, dense-urban and urban. We calculated the GOPS needed in the four split options and the GOPS needed in the Distributed Radio Access Network (D-RAN) to estimate the processing reduction in each case.

The results obtained from analysing the service Use Cases give us indication on the level of resources that would be necessary, rather than absolute numbers. Many of the Use Cases analysed in this work need to be mainly served from the AMEN due to latency requirements, thus AMEN, as expected, need to be provided with Storage and Computing facilities. The amount of storage and computing facilities depend fundamentally on the verticals being supported as there are big differences between service requirements, which can vary between 2 GB to 4 PB (Petabytes), using the assumptions made for each service use case. Computing facilities also vary very much between 2 vCPUs and 400 vCPUs, while switching throughput necessary in AMEN varies between 10 Gb/s and 600 Gb/s (video related services needing the highest throughput as expected).

MCEN dimensioning depends on the number of AMEN it is connected to. MCEN connections on the core network side are not taken into account in the tables provided. We see that, as expected, MCENs need the same type of functionality as the AMEN, but at a higher scale just because of the two previous points.

The optical layer needs to be flexible in order to meet the service requirements, i.e. a control plane needs to be built that can set up, tear down, or reconfigure the services that need to be supported. The connection capacities through the optical metro network need to be between 10 Gb/s and 100 Gb/s depending on the services being supported. Multicast capacity in the optical layer helps in limiting the total capacity needed in the metro network

Overall, the capacity needed in Metro-Haul access edge nodes (AMEN) to support all the service Use Cases is estimated to be in the order of more than 1 Tb/s of aggregated switched traffic with a dozen of 100 Gb/s links for theirs interconnections with the Metro core edge nodes and tens 10Gb/s for the interconnection with the users on the access side. Throughput is dominated by video services, in particular Live TV and 6DoF VR. A few tens of Tera-bytes storage capacity is estimated to be needed in each AMEN. It is worth noting that the object tracking use case alone requires a few thousand TB of storage capacity. MCEN nodes requires lower storage capacity when compared with each AMEN. When all service use cases are supported by an access node simultaneously, an aggregated computing capacity of around 1,700 vCPUs results, which in large part is due to the Smart factory service Use Case (assuming that factory servers are virtualized in the node) and to Service robotics and Autonomous driving, the last two ones with lower numbers. The computing capacity necessary in an MCEN is marginally higher.

- Resilience is also a necessary element, which needs the SDN/controller in order to link the optical layer in the metro network with the different access technologies including 5G RN, and the Data Centre facilities in the AMEN

- Traffic analysis and monitoring is also necessary to a) meet service requirements, b) pro-actively ensure service availability and reliability

Regarding the requirements of the set of service Use Cases analysed in this deliverable on the control and monitoring components of the Metro-Haul architecture, we have reviewed requirements collected from all the Use Cases and classified them into macro categories, identifying the major features for each category. Then, for each category, we have identified for which feature it is more demanding for the Metro-Haul network. This classification work has been done separately, first for the monitoring and data analytics subsystem, and then for the control and management plane and the planning subsystems.

Table of contents

List of Figures

List of Tables

1 Introduction

The overall Metro-Haul objective is to architect and design cost-effective, energy-efficient, agile, and programmable metro networks that are scalable for 5G access.

The Metro-Haul architecture must be capable of future service requirements, encompassing the design of all-optical metro nodes (including full compute and storage capabilities), which interface effectively with both 5G access and multi-Tb/s elastic core networks.

Metro-Haul has taken the existing 5G KPIs and determined their implication for the optical network with 5 targets:

- 100 x more 5G capacity supported over the same optical fibre infrastructure;
- 10 times less energy consumption;
- Latency-aware metro network in which latency-sensitive slices are handled at the metro edge ensuring the metro network adds no additional latency;
- End to end SDN-based management framework enabling fast configuration time to set up or reconfigure services handling 5G applications, specifically 1 minute for simple network path set-up and 10 minutes for full installation of a new Virtual Network Function (VNF) and 1 hour for setting up a new virtual network slice;
- Reduction in CAPEX of a factor of 10, plus a reduction in OPEX of at least 20%.

The deliverable D2.1 (this document) aims to deliver a preliminary set of requirements and KPIs for the Metro-Haul architecture. It provides inputs to the project to define solutions and strategies compliant to the fulfilment of the five general targets listed above, but also by taking into account requirements coming from three different and complementary lines of analysis.

The first line of analysis adopts a general “high level” perspective, it considers the trends emerging in recent technology and market outlook surveys. This top-down approach, starting from the predicted network evolution and traffic forecasts as a whole, derives the implications on Metro-Haul network architecture, with particular reference to new emerging node paradigm (Central Office as a Data Centre) and bandwidth requirements (size of interfaces and related traffic pipes between nodes). This approach and its consequences are discussed in Section 2.

A significant portion of this document, and indeed deliverable, is dedicated to the second line of analysis, which assumes as input the services demanded by customers belonging to a wide and relevant set of verticals (bottom-up). This second “service oriented” approach starts from the description and the analysis of specially selected 5G Use Cases. We will then investigate the implications of the Use Cases in relation to requirements for the Metro-Haul architecture. The Use Cases, collected and presented in greater detail in Section 4, belong to verticals that have been considered significant regarding their potential with respect to demand for connectivity and computing services, especially during the first phase of the 5G infrastructure rollouts.

To categorize and characterize the Use Cases and their related services a framework has been defined. This framework includes the specification of high-level properties (belonging to a vertical, Class of Service (CoS) required), the Service KPIs selected to characterize the services in terms of technical parameters (e.g. data rate, latency, etc.), and the identification of components and functionalities needed by each type of Use Case, together with their mapping within the Metro-Haul architecture. In addition, guidelines to derive from the Use Case description the Metro-Haul requirements at network level are also specified. Furthermore, this facilitates detailed Metro-Haul network KPIs to be defined. The framework, which is applied to Use Cases presented in Section 4, is discussed in Section 3.

Section 5 discusses the third line of analysis which reviews the impact of 5G Radio Network Access on Metro-Haul architecture. This issue is transversal to Use Cases and services because it deals with the way the aggregate traffic (any traffic type, any CoS components) is carried from the antenna to the mobile Core Network (CN). It considers that a split option is the mode to distribute the functions between the Radio Unit (RU), the Distributed Unit (DU) and the Centralised Unit (CU). It provides a limited set of split options to be identified and addresses their influence on Metro-Haul architecture, and analyses contributing factors and their impact, including latency constraints and fibre distances.

Section 6 summarize the results achieved and reports the main findings and indications from the three lines of analysis used to derive the Metro-Haul requirements

1.1 Metro-Haul Architecture

The Metro-Haul architecture (Figure 1) describes the network elements necessary to interconnect a number of access network nodes, the Access Metro Edge Nodes (AMEN), with the core network nodes, the Metro Core Edge Nodes (MCEN), via a high capacity dynamic and flexible optical network. Varying network types in different geographic locations will have different topologies (connection types, nodes and transmission distances). Transmission distances across these networks often vary between 50km and 200km.

The access network is out of scope of Metro-Haul, but it is assumed to that ultra-high capacity technologies will be available where 5G Radio Access Network (RAN) and fixed fibre networks play a predominant role. In these cases, typical distances between customer premises and the AMEN are of the order of 20km or less. Nevertheless, any access network terminates at an Optical Line Termination (OLT) located in an AMEN. Note that the 5G RAN may have some components (e.g. CU) either at the back (network side) of the OLT in the AMEN, or even in a MCEN in some cases.

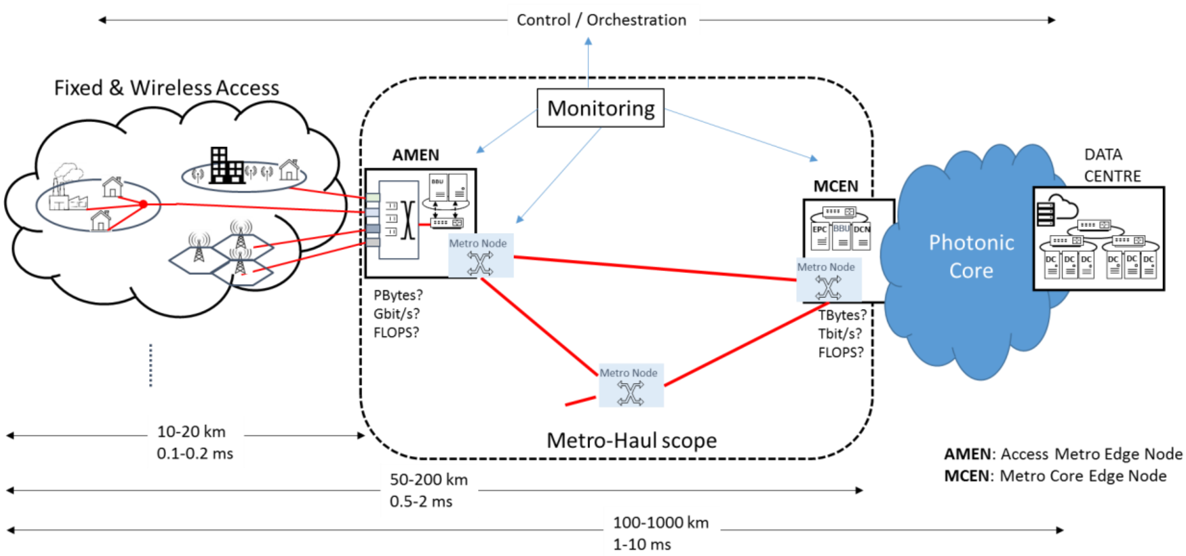


Figure 1 – Metro-Haul Network Architecture.

Control and monitoring functions will be required to manage the Metro-Haul network architecture. To implement a metro network that can meet the Metro-Haul objectives, both storage and computing resources will be necessary both in the AMEN and the MCEN. These resources may be used to host one or more controllers and/or network monitoring function that will have open interfaces capable of connecting to the various network components and nodes in the optical transport layer.

Additionally the controller needs (currently an issue that requires further study) an end-to-end view of the network in order to set up the network interconnection to give the customer (either still or in motion) access to the service resources it may involve and thus allow the service to use any necessary resources to make it dynamic, flexible and cost-effective. Therefore, part of the resources may involve compute, function, and storage resources (data centres) deep in the core network.

Finally, the controller will need information about the service traffic in order to effectively react to changing conditions, including: moving from access area to a different access area (e.g. moving from a RU connected to an AMEN to a different RU connected to a neighbouring AMEN).

This information is collected and made available to the controller by monitoring resources, which then may be used by the controller to dynamically and efficiently change the service configuration. The goal is provide continual service operation which is undisturbed by changing, planned or otherwise, circumstances. The same objective must be sought when a catastrophic failure occurs in order to guarantee a high availability for critical services.

Although the scope of the MH architecture overview is more focused on the description of *data-plane* components (i.e. AMEN, MCEN, Metro Nodes and optical links), mentioning the controllers gives us the opportunity to introduce the role and key components of the *control and management planes*.

The controllers introduced above may be considered the Software Define Networking (SDN) system that controls the assignment of both the computing and the transmission resources to provision the requested services. The SDN controllers interact with the data plane via the monitoring system, as already hinted above. The control plane is then completed by the service orchestrator. This component is able to receive requests from users in terms of service chains and translates the requests into requests of computing and transmission resources, fed to the SDN control architecture. The service orchestrator also interacts with the Operations Support System (OSS) of the network operator to receive configuration and policy inputs. Finally, we should mention that the Metro-Haul ecosystem includes a Network Planning subsystem. This final component is intended to support the control plane by suggesting strategic long-term actions such as periodic global network reconfiguration to improve the performance of the entire Metro-Haul ecosystem.

The control and management plane in Metro-Haul and their components mentioned above are studied within WP4, which also will define the details of the interfaces between the different modules.

2 Metro-Haul High-Level Requirements Definition

Estimating the dimensioning and evolution of a future 5G Metro-Haul network can be performed via a qualitative high-level (top-down) approach; whilst the alternative (bottom-up) quantitative approach is therefore undertaken by estimating the data, traffic, capacity requirements of the various mIoT (massive Internet of Things), CriC/URLLC (Critical Connections/ Ultra Reliable Low Latency Connections), and eMBB (enhanced Mobile BroadBand) 5G Use Cases that have been identified, and appropriately aggregating them together to reach an overall dimensioning requirement.

The high-level approach estimates trends in data-centre capacities, end-user bandwidth and data consumption, as well as other technology trends (e.g. 5G, SDN/NFV, fog computing) to assess the expected load on the Metro segment of the overall telecoms network. Assisting us in the top-down estimation, Cisco has recently published its Global Cloud Index: Forecast and Methodology, 2015–2020, White Paper [1], which provides its best-guess estimation of the evolution of some of these important technology trends and parameters. We also extrapolate some of the results from the Cisco white paper into the 2024 timeframe, by assuming similar growth rates. This allows us to gain understanding over a longer timeframe, towards 2024, and which is also appropriate for the 5G Metro-Haul design.

By 2020, Cisco expects there will be over 26 billion Internet-connected devices and over 4.1 billion global Internet users. The world entered in 2016 the so-called "zettabyte era", whereby global IP traffic reached 1.1 zettabytes (ZB), or over 1 trillion gigabytes. By 2020 global IP traffic is expected to reach 2.3 ZB; a factor x2.3 increase, while the number of networked devices and connections is expected to reach 26.3bn, with 82% of all IP traffic anticipated to be video.

The workhorse of future network evolution is the data centre (DC). The DC is highly scalable, ranging in size from a small micro-DC, up to hyperscale DCs[5]. In 2015 there were 259 such hyperscale DCs across the globe, and according to Cisco this is expected to grow to 485 by 2020. This represents a 13.4% CAGR, such that by 2024 assuming the same growth rate, we could expect there to be 802 such hyperscale DCs. In addition, by 2020 hyperscale DCs will be expected to house 47% of all installed data centre servers. Currently there are 24 hyperscale cloud operators, providing infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), software-as-a-service (SaaS), or other cloud services. Categorised by their revenues, they can be logarithmically (power-law) ranked as follows:

- $1B in IaaS/PaaS (e.g., Amazon/AWS, Rackspace, NTT, IBM)
- $2B in SaaS (e.g., Salesforce, Google, Microsoft, Oracle)
- $4B in Internet/search/social networking (e.g., Facebook, Apple, Tencent, Yahoo)
- $8B in e-commerce/payment processing (e.g., Amazon, eBay, Alibaba)

[5] What exactly constitutes a hyperscale data centre varies from expert to expert, and cloud operators do not share hard data about their infrastructure publicly. Generally only seven companies are strictly considered to be hyperscale – Google, Amazon, Microsoft, Facebook, Alibaba, Baidu, and Tencent - but other companies - including Apple, IBM, Twitter, LinkedIn, and eBay, among the others - have the characteristics of hyperscale if one looks at their way to operate in addition to the size of their facilities.

Traffic within hyperscale DCs is expected to quintuple (x5) by 2020 as compared to 2015. Hyperscale DCs already account for 34% of total traffic within all DCs, and are anticipated to grow substantially and account for 53% by 2020; i.e., there is an interesting trend towards greater centralisation of DC processing, rather than greater decentralisation. Such centralisation would also therefore tend to increase the loading of the metro and core segments of the network. Note that strict latency means that some functionality needs to be located close to the service user, i.e. in small DCs. This feature of future networks does not contradict the need for more Hyperscale DCs, as they act as content and information repositories, which can be shifted around in the network as how and when required.

By 2020 more than 90% of DC traffic will be cloud traffic. Although the amount of global traffic crossing the Internet and IP WAN networks is projected to reach 2.3 Zetabyte (ZB) per year by 2020, the amount of annual global DC traffic in 2015 was already estimated to have been 4.7 ZB, which is expected to triple to 15.3 ZB per year by 2020; representing a growth rate of 27% CAGR, see Figure 2. Assuming the same 27% CAGR, by 2024 we can expect global DC traffic to reach about 40 ZB. Consumer IP traffic is expected to reach 10.9 ZB (x2.5) by 2020, while business IP traffic will account for 4.4 ZB (x2), up from 4.3 ZB and 2.2 ZB respectively in 2016.

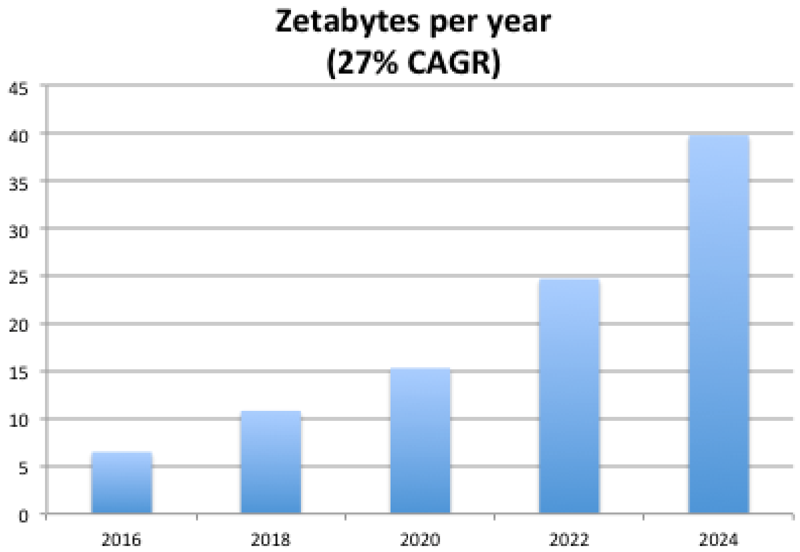


Figure 2 – Global Data Centre IP Traffic Growth

DC Storage Forecast

• By 2020, DC storage installed capacity is expected to grow to 1.8 ZB, up from 382 Exabyte (EB) in 2015, nearly a 5-fold growth; indeed the global modular data centres and related equipment market is expected to triple in size and reach $35.11 billion by 2020 [2].

• By 2020, the total global installed data storage capacity in cloud DCs will account for an 88% share of the total DC storage capacity, up from 64.9% in 2015.

Table 1 below, indicates that highest growth is expected to be represented by DC-DC traffic (see line 2 of table), rising from 346 EB per year, to 1,381 EB (x4) per year by 2020; representing a CAGR of 31.9%. Assuming a similar growth rate, this would mean that DC-DC traffic could increase to 4.18 ZB per year by 2024. Such DC-DC traffic is of specific relevance to the Metro-Haul project. Big data is a significant driver of traffic within the DC. While much of the big data traffic is rack-local, enough exits the rack such that big data is expected to be responsible for 17% of all traffic within the DC by 2020, up from 10% in 2015. Note that video does not itself drive a large volume of traffic within the DC, since minimal processing is performed on video relative to the large size of the video stream.

According to Cisco GCI study [1], traffic between DCs is growing faster than either traffic to end-users or traffic within the DC: By 2020, traffic between DCs is expected to account for almost 9% of total DC traffic, up from 7% at the end of 2015. The high growth of this segment is due to the increasing prevalence of content distribution networks (CDNs), the proliferation of cloud services and the need to shuttle data between clouds, and the growing volume of data that needs to be replicated across DCs.

Table 1 – Global DC Traffic 2015-2020 [1]

Data Center IP Traffic, 2015–2020							
	2015	2016	2017	2018	2019	2020	CAGR 2015–2020
By Type (EB per Year)							
Data center to user	744	933	1,164	1,438	1,772	2,183	24.0%
Data center to data center	346	515	713	924	1,141	1,381	31.9%
Within data center	3,587	5,074	6,728	8,391	10,016	11,770	26.8%
By Segment (EB per Year)							
Consumer	2,997	4,304	5,836	7,435	9,075	10,906	29.5%
Business	1,681	2,218	2,768	3,318	3,853	4,429	21.4%
By Type (EB per Year)							
Cloud data center	3,851	5,636	7,712	9,802	11,850	14,076	29.6%
Traditional data center	827	885	892	951	1,078	1,259	8.8%
Total (EB per Year)							
Total data center traffic	4,678	6,522	8,604	10,753	12,928	15,335	26.8%

Big Data is a key driver of overall growth in stored data, with big data expected to reach 247 EB by 2020 (see Figure 3, below), up almost 10-fold from 25 EB in 2015, and representing 27% of data stored in DCs by 2020, up from 15% in 2015. Assuming a continuation of the current 58% CAGR [1], the total big data storage volume is therefore expected to reach almost 1.6 ZB (zettabytes) by 2024.

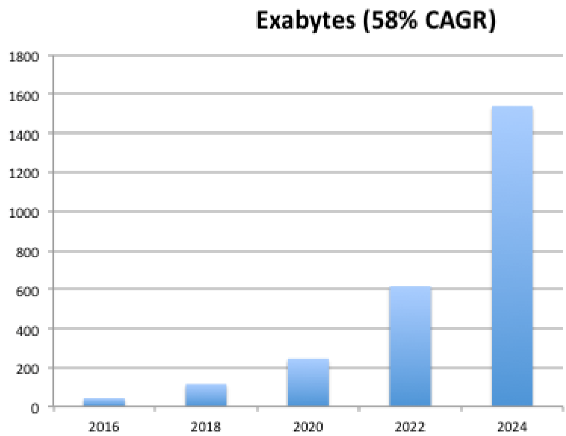


Figure 3 – Big Data Volumes

Although the data stored in DCs will be nearly 1 ZB by 2020 (growing nearly 5-fold from 2015 to 2020, a CAGR of 40%) the amount of data stored on end-user devices is expected to be 5 times higher: 5.3 ZB by 2020. Out of the combined 6.2 ZB of stored data in the world, most stored data will continue to reside in client devices, as is the case currently. Today, only 12% of total stored data is stored in DCs, but more data will move to DCs over time. Cisco GCI estimates that by 2020, 59% (2.3 billion) of the consumer Internet population will use personal cloud storage, up from 47% (1.3 billion users) in 2015.

In addition to the larger volumes of stored data, its origin will be coming from a wider range of devices by 2020. Currently, 61% of data stored on client devices resides on PCs, but by 2020, stored data on PCs is expected to reduce to 52%, with a greater portion of data on smartphones, tablets, and machine-to-machine (M2M) modules. Stored data associated with M2M is growing at a faster rate than any other device category.

Potential Effects of Internet of Everything (IoE) on Global DCs

Cloud services are being accelerated not only by the unprecedented amounts of data being generated by people, but also by machines and things. Cisco GCI estimates that 600 ZB will be generated by all people, machines, and things by 2020, up from 145 ZB in 2015. Gartner [3] estimates that the cloud market will grow from USD 58 billion in 2013 to USD 198 billion in 2020, representing an increase of factor ×3.41, and a 19% CAGR. Gartner is also promoting the DC-as-a-Service (DCaaS) model, where the role of IT and the DC combines to deliver the “right service, at the right pace, from the right provider, at the right price”.

One increasingly important vertical application for DCs, which also has important relevance to the access and metro space, is for self-driving (autonomous) cars and delivery vehicles. AT&T is investigating the deployment of edge DC networks for self-driving cars [4]. The next generation of such applications, powered by advances in machine learning, and virtual and augmented reality (AR), will require a great deal of computing power, with near-real-time responses from the computing systems, such that much of that computing power will need to be deployed at network edges, with smaller-capacity DCs scattered in and around densely populated areas to receive and process the data they will be receiving from self-driving cars or AR systems and sending data back.

AT&T, like many operators, wants to exploit the widespread and distributed infrastructure assets it has built out, comprising its current wireless network (COs, cell towers, and other types of base station sites), to house its future edge DC network as it prepares for the roll-out of 5G wireless technology, so as to enable the type of low latency and high speed data transfers necessary for such next-generation applications.

This strategy also complements AT&T’s decision several years ago to convert its COs into DCs (including central office re-architected as a datacentre (CORD) Open Source project); i.e. to transform its legacy network into a software-defined network, with automated network management, and virtualised network functions running on commodity servers, and delivered in a similar manner to cloud services. AT&T’s goal is to virtualize 75% of its network functions by 2020 (currently, 55% virtualization has been attained by 2017.)

2.1 Requirements for network infrastructure (nodes and links)

The techno-economic factors discussed earlier in the section and 5G use cases, demonstrate that DC traffic will continue to rapidly contribute to the overall traffic load and dynamics of future metro networking. Broadly speaking, we can see that overall DC-DC traffic will increase by a factor of x4 from 2015 to 2020, i.e. a 31.9% CAGR, with other associated computing metrics (e.g. storage, big data volumes) increasing by similar amounts. In addition, Cloud is causing a centralising trend that will also increase load on metro and core networks. Assuming a similar 32% CAGR for the subsequent 5 years to 2025, it appears that we can expect a 4x4 (i.e. x16) increase in traffic load in metro networking over the next decade.

From the point of view of 5G, which has already been mentioned above to be highly complementary to increasing the role of DCs in the network, a key emerging 5G technology is C-RAN (Cloud radio access network), which represents a centralised processing approach to radio access with multiple data servers, i.e. base band units (BBUs) or central units (CUs), located at a central office, either at an AMEN, or within the metro network at a MCEN.

Figure 4 shows the Metro-Haul network architecture, with the relationships between the key nodes and links (bandwidth pipes) within the scope of this study and over the relevant timeframes (2020-2025). We see a progression from micro-tiny-small DCs on the left-hand side, as they appear within the fixed and wireless access network, to small-medium-large DCs in the regional networking location (i.e. the main Metro-Haul scope), to the large-hyperscale DCs appearing in the photonic core.

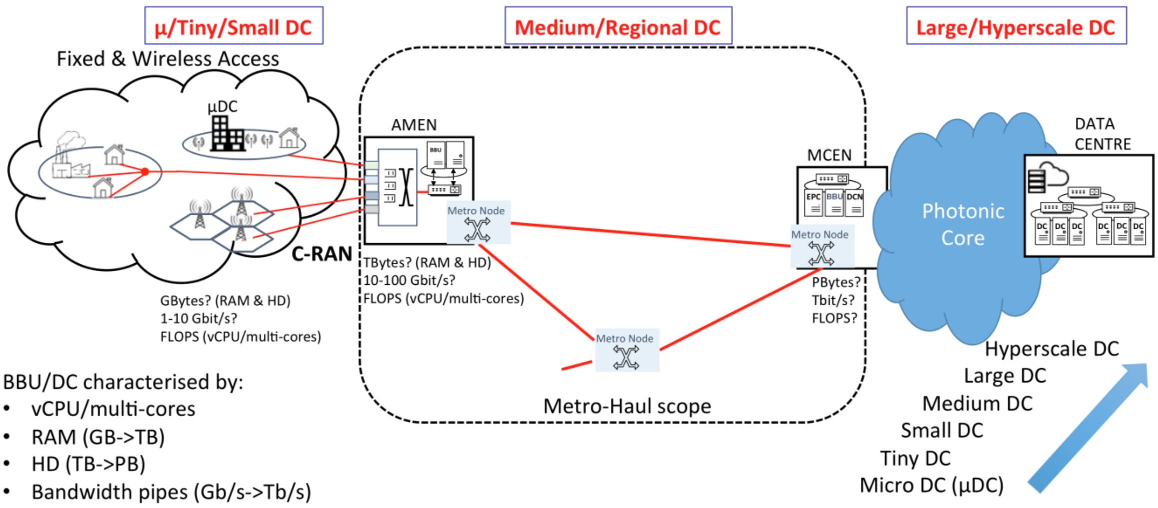


Figure 4 – Distribution and dimensioning of DCs across the Access-Metro-Core networks

2.2 Approximate dimensioning of Metro-Haul network capacity

The dimensioning of C-RAN architectures, particularly to the required specifications for the centralised BBUs or DUs, in order to accurately estimate the dimensions of DCs in the metro segment and associated metro network loading. With 5G KPIs suggesting end-user bandwidths of 1-10 Gb/s, with latencies of 1-5 ms, along with provision of the other 5G services (mIoT, CriC/URLLC, and eMBB) mentioned above, the demands placed upon the CUs will be considerable, i.e. the data processing, storage and exchange of data between DCs across the metro network will be large. Furthermore, the dimensioning of the CU of a C-RAN also depends on the reach of the RAN network and how many end-users it will service.

Unfortunately, the dimensioning (particularly of the CUs etc.) of such C-RAN architectures by vendors such as Huawei, Ericsson, Nokia, Fujitsu etc., isn't readily available on the Internet, e.g. in product specification brochures. This may be because C-RAN solutions are still in the relatively early stages of development and roll-out (i.e., they are still rapidly evolving, as the relevant standards continue to be defined, e.g., as for CPRI and for the functional splits), such that any commercial solutions will be necessarily bespoke designed and dimensioned according to the customer's specific needs. In particular, the storage capacity, processing power, and physical size (no. of racks etc.) of the CU servers will depend on the number of end customers being served (and the assumed statistical multiplexing gains) as well as the services being offered (i.e. 4G currently, evolving towards 4.5G and 5G), as well as the degree of future-proofing desired in the particular C-RAN solution. The key takeaways from these conclusions, are that any future infrastructure must be flexible and capable of coping with changing demands and unanticipated bandwidth.

However, if a 'typical' C-RAN CU dimensioning can be assigned for current solutions and assumed to be representative of an equivalent DC specification, then we can probably assume that the compute (storage, processing, data-pipes etc.) dimensioning of such a 2017 C-RAN solution would need to increase by a factor x4 by 2020, and a further factor x4 (i.e. x16 overall) by 2025. This again, assists in helping to understand the anticipated metro network dimensioning requirements through to 2025 to underpin expected 5G evolution and KPI performance. Figure 4 above indicates the dimensioning of the different DCs (ranging from micro, small, to medium, large and hyperscale DCs) located across the access, metro, and core networks, as expressed using the parameters of: storage (RAM), bandwidth pipes, and processing power (number of multi-cores) that are expected at each location. Together, these define the approximate performance requirements at each location, and assist in providing the high-level forecast of the Metro-Haul dimensioning.

In particular, we expect the AMEN to be served by pipes of 10-100 Gb/s, e.g. 10 Gb/s on the access interfacing side, and aggregated pipes of up to 100 Gb/s on the Metro-Haul interface side. The AMEN DC would have storage capability measured in TB of RAM and HD capacity, with processing power of the order of 1 TFLOPS (e.g. as extrapolated from the 2014 study described in [5] based upon RAN-as-a-Service considerations) and which can be embodied using the Intel Core i9 Extreme Edition chip, which consists of 18 vCPUs or multi-cores on a single chip exhibiting teraflop capability.

The MCEN that interfaces with the photonic core can be expected to require another factor x10 increase in performance capabilities, as compared with the AMEN. Specifically, MCEN will be served by 100 Gb/s pipes from other AMEN and intermediate Metro-Haul nodes, and interface via 400G/s and 1 Tb/s pipes (i.e. comprising super-channels) into the photonic core. The storage capacity of the MCEN will be measured in the 10s of TB of RAM and HD, with associated compute resource of the order of 10 TFLOPS.

In terms of the average node degree sizes for the MCEN and AMEN, these can be expected to remain fairly similar (i.e. constant) to the current topology descriptors, since intrinsic connectivity and topological (geographic) parameters are not anticipated to greatly change over the Metro-Haul timeframe.

3 Metro-Haul requirements driven by Use Cases

This section gives the fundamentals for the Use Case descriptions used in Section 4. The approach followed starts from the characteristics and needs of each Use Case, and then it derives the requirements for the network in terms of functionalities (what they are and where they are located) and required performance (with which level of QoS parameters the communication and IT services are required).

Firstly a categorization of the Use Case is presented. A wide, although not exhaustive, selection of verticals to categorize each Use Case have been identified.

Secondly the concept of service class is introduced, taking mainly into account the definitions established in 3GPP documents. A selection of service KPIs, also taken from the 3GPP documents, is then proposed to characterize the communication services required by each Use Case.

Assuming the above definitions a standard scheme to be applied to each of the proposed Use Cases for its description is defined. The scheme is composed by the following steps:

1. The Use Case is identified in terms of category (membership within the verticals and coverage requirement) and service class required by the communications services (a Use Case can require more than one service and each service can be associated to more than one class);
2. The components and functionalities of the Use Case is identified and mapped onto the Metro-Haul architecture. This task is very important because it allows to determine the impact of the Use Case requirements on the network;
3. The values of service KPIs (one or more services) associated to the Use Case are assigned;
4. Using a proposed methodology (see 3.5) the high-level Metro-Haul requirements, expressed by means of proper network KPIs, are then derived taking onto account the achievements of steps 2 and 3.

3.1 Use Case categorization: Vertical Industry and required coverage

In recent years many documents (white papers, regular articles, standards, and recommendations), often published by important institutions and consortia, present a wide set of 5G Use Cases associated to a variety of verticals that are expected to ask for communication services to the network of the future, i.e., the 5G infrastructure.

The NGNM white paper on "Perspectives on Vertical Industries and implications for 5G" [10] presents seven categories of Use Cases, each of them associated to a vertical industry. The seven families of Use Cases (or verticals) are: Automotive, Transport and Logistics, Health and Wellness, Smart Cities and Utilities, Agriculture, Media and Entertainment, Industry IoT. For each vertical an ecosystem and market analysis is proposed and a number of relevant Use Cases are listed. For example for the Media and Entertainment, the Use Cases mentioned are the Cooperative media production, the Collaborative Gaming and the New Media Experience (VR/AR). The required capabilities to support the Use Cases of each vertical are also reported in the document.

The 3GPP technical specification # 22.261 "Service requirements for the 5G system; Stage 1 (Release 16)" [7] analyses the Basic Capabilities (sec. 6) and the Performance Requirements (sec. 7) for a 5G network and in doing that presents some examples of scenarios involving Use Cases belonging to a number of specific verticals. Some verticals used in this 3GPP document are kept from the NGCM document cited above [10]. Although a number of Use Cases are referenced and employed to exemplify the service requirements, a systematic classification of Use Cases and verticals is not provided because it is out of the scope of the document.

The 5GPPP document "5G PPP Use Cases and performance evaluation" [12] propose a framework for Use Case classification. It defines 6 Use Case families obtained gathering the Use Cases analyzed in different 5GPPP projects (this is done relying on the project's results available up to beginning of 2016). The Use Case families are defined mainly on the base of their technical requirements and are the following ones: Dense Urban, Broadband everywhere, Connected vehicles, Future Smart offices, Low bandwidth IoT, Tactile Internet / Automation. The same document defines also a set of 5 relevant vertical industries: Automotive, E-health, Energy, Media and Entertainment, Factory of the future. Within each vertical a set of business cases are then identified and associated to one of the

Use Case families. For instance, the business case "Automotive driving" is associated to the "Automotive" vertical industry and belong to the Use Case family of "Connected vehicles".

Another recent source is the GSA white paper published in July 2017 and titled "How 5G changes perceptions of applications, coverage and architecture" [11]. In that report it is underlined as particularly critical for the Use Cases the aspect associated to the coverage: "Each specific Use Case can be defined in terms of a number of requirements; coverage is intrinsic to some applications, a limiting factor to development on others, and a peripheral requirement to others. It is important to understand this when discussing the concept of coverage in 5G networks more broadly". The 7 Use Cases mentioned in the report are: Fixed mobile broadband in rural and underserved areas, Enhanced Mobile Broadband, Cloud robotics, AI and industrial automation, Public safety and administration, IoT and human-IoT interaction, AR/VR and the tactile Internet, ITS V2X and autonomous vehicles. Comparing these definitions with the ones included in the other documents mentioned before, it becomes apparent that they should be more appropriately classified as super-families of Use Cases, rather than simple individual Use Cases.

Considering all the aforementioned references and being aware that it is not possible to make an ultimate classification in a such a dynamic technical ad socio-economic evolving environment, we propose the set of verticals included in the left part of Table 2.

For the Use Case categorization a second additional property is defined to identify the type of geographic coverage extension required by the services associated to the Use Case. Coverage extension means the area in which a service is expected to be accessible with the declared level of availability (e.g. 99.99%). The ten types of coverage area are listed in the right hand side of Table 2. Please note that in general there is not a 1 to 1 correspondence between the verticals in Table 2 (left) and the Coverages (right) on the same row. For instance if V2-Automotive vertical matches 1:1 with C2-Road networks (on the same row), V10-Cloud Services matches with C1-Everywhere and potentially also with C5-Indoor hot spot, depending on particular Use Case, but not with C10-Sky flight paths.

Table 2 – Verticals (on the left) and type coverage (on the right) used for Use Case's categorization.

Vertical Industry or Socio-economic sector	
V1	Mass Market
V2	Automotive
V3	Transports and logistics
V4	Utilities
V5	Public safety & Environment
V6	Industry 4.0
V7	Health and Wellness
V8	Agriculture
V9	Media and Entertainment
V10	Cloud Services (incl. Service Robotics)

Type of space Coverage required	
C1	Everywhere
C2	Road network
C3	Railway network
C4	Smart City
C5	Indoor hotspot
C6	Industrial Plant
C7	Hospital
C8	Agricultural land
C9	Crowd location
C10	Sky flight paths

3.2 5G Service types and classes

The 5G service macro classes adopted in this deliverable are the three that emerged in 3G PP, but agreed upon also in other contexts like 5G PPP. They are listed and described below:

- enhanced Mobile BroadBand (**eMBB**) It is a class of service introduced for handling enhanced Mobile broadband in 5G network. It is useful, but not limited, to the general consumer space mobile broadband applications including streaming of High Quality Video, Fast large file transfers etc. the Data rate required by a single connection belonging to this class could reach 1 Gb/s on average and 10 Gb/s at the peaks.
- massive Internet of Things (**mIoT**). This class allows the support of a large number and high density of IoT devices efficiently and cost effectively. Density of devices requiring connectivity could reach 1 million per Km^2.
- Critical Connections/ Ultra Reliable Low Latency Connections (CriC/**URLLC**). This class is created for supporting low latency ultra-reliable communications for critical applications including augmented (virtual) reality, advanced gaming, industrial automation, tactile interaction, (remote) control systems. Public safety and Critical Communication are included in this category. Requirement on latency for services belonging to this class could be of the order of a few milliseconds (1 ms in most demanding case) and availability 5 nines or higher.

Figure 5 gives a visual representation of the three classes of services as axis in a ternary plot, and how some example applications are proportionally mapped to each of them according to their requirements.

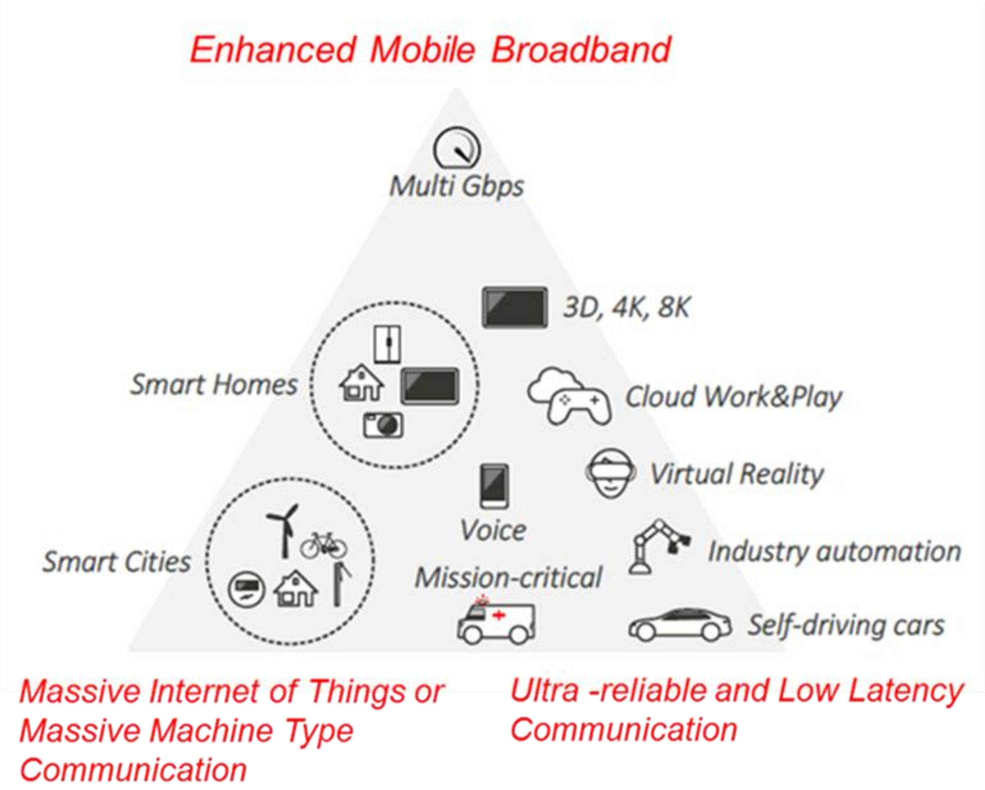


Figure 5 – Triangle representation of service macro classes with some examples of application placed according to their requirements.

It is important to highlight that this rather rigid classification, which use as specific performance indicators i) the data rate, ii) the density of devices/connections and iii) the reliability and latency (together), does not properly cover all the service profiles. Each service will typically have its mix of latency, data rate, availability and other requirements, which can impose the assignment of a hybrid class of service.

In Figure 6, taken from the 5G PPP 5G vision brochure [12], a similar, but more detailed representation with additional service performance indicators, is proposed. The radar diagram includes in the inner part (vertices of the dark blue polygon) the level of performance offered by the state of the art LTE technology and, close to the arrows outgoing the octagon vertices, the target values for the 5G network according to the 5G PPP vision. Performance indicators used in the diagram are 8 in total and, in comparison to the ones used to define the three macro classes above, they are enriched with: data volume (i.e., area capacity), service deployment time, energy efficiency and mobility.

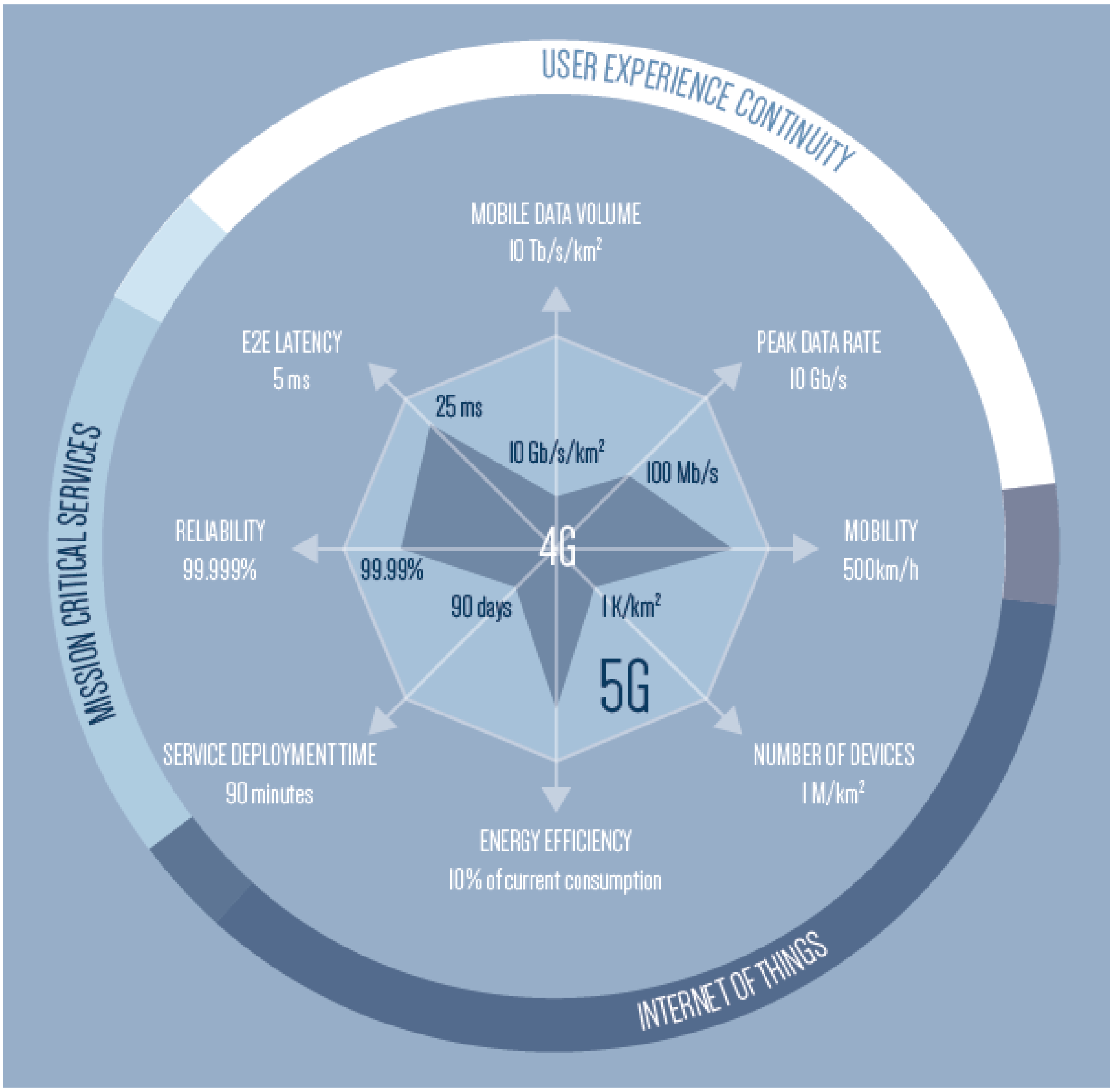


Figure 6 - Radar diagram of 5G disruptive capabilities (source: 5G PPP 5G vision brochure [12])

Comparing the service classification proposed and the outer ring of Figure 6 we can identify corependences of the service class **eMBB** with the USER EXPERIENCE CONTINUITY, of **URLLC** with MISSION CRITICAL SERVCES and of **mIoT** with the INTERNET OF THINGS (the latter with the nuance that not all IoT applications imply an high number/density of sensors/devices, as required by the attribute “massive” in mIoT class definition).

3.3 Identification and mapping of service Use Case components and functions

One fundamental task in the Use Case description is to identify the typical components and functionalities and map them on the architecture depicted in Figure 1.

Such components can be HW components like user terminals that, by extension, can be: connected sensors, vehicles on the road, service robots in a building, pieces of machinery in an industrial plant, video cameras in the streets and so on. Other type of HW components can be dedicated servers required by the service that can be placed in the customer premises or in the Operator's facilities (in Metro-Haul context, the Metro Central Office). In the last case they could be virtualized as virtual machines on shared facilities owned by the operator or by third part.

SW components are also fundamental and have to be identified and described with regards to their relation with HW components and positioning within the network architecture.

Functionalities like monitoring, control, data collection and processing (data analytics), particularly for the Use Case under description, have also to be identified in terms of which variables under observation or control, operations (what it is done with those observations) and positioning (where functionalities is necessary).

3.4 Key Performance Indicators (KPI) for services

The parameters and features selected to characterize a service active within a Use Case and recognized in the following as Service KPIs are reported in Table 2. The first 8 KPIs are parameters to be specified by a value (or a range) while the remaining 6 are features that can assume a binary value (for instance YES or NOT, for the feature "Relying on sensor network"). The service KPIs and their definitions are selected from available documents on this matter already mentioned in section 3.1, and in particular from the 3G PP Technical Specification # 22.261 [7].

Table 3 – Parameters and features considered as service KPIs in Use Cases.

N.	**Key Performance Indicators**	**Short definition**
1	Connection bandwidth (data rate) [Mb/s]	Peak and Average values of data rates should be provided (it could be useful or mandatory to provide also the user experienced data rate: the minimum data rate required to achieve a sufficient quality experience, with the exception of scenarios for broadcast like services where the given value is the maximum that is needed).
2	Latency [ms]	The time it takes to transfer a given piece of information from a source to a destination, measured at the communication interface, from the moment it is transmitted by the source to the moment it is successfully received at the destination.
3	Jitter [ms]	Variation in latency as measured in the variability over time of the packet latency across a network. Packet jitter is expressed as an average of the deviation from the network mean latency.
4	Reliability [%]	Percentage value of the amount of sent network layer packets successfully delivered to a given node within the time constraint required by the targeted service, divided by the total number of sent network layer packets.

5	Availability [%]	Percentage value of the amount of time the end-to-end communication service is delivered according to an agreed QoS, divided by the amount of time the system is expected to deliver the end-to-end service according to the specification in a given area. Communication service availability relates to the service interfaces, reliability relates to a given node or to a network. Reliability should be equal or higher than communication service availability.
6	Mobility – User Equipment speed [km/h]	Fixed (no mobility: office, home) or max speed in movement (Pedestrian or on a transportation mean: train, road vehicle, airplane, drone, etc.)
7	Area Traffic Capacity [Gb/s/km2]	Traffic collected per area. It can be specified with average and peak values and Up Link - Down Link separately. Relevant for hotspots and crowded area scenarios.
8	Relying on Sensor Network [Yes/No]	A Use Case/service relying on sensor network (can be massive or not, as specified by KPI number 9).
9	Massive Type [Yes/No]	High density sensor (i.e. millions of sensors per km^2)
10	Device Direct [Yes/No]	There are connection requirements between two UEs without any network entity in the middle.
11	Coverage Requirement [Standard/Extended]	Similar to application scenarios: Full Network, Office and residential, confined area (crowd), downtown, along roads, along railways, …
12	Energy Consumption critical [Yes/No]	Regards consumptions of sensors with autonomous battery (not connected to the energy mains network).
13	Type of User Equipment [Conventional/Special purpose]	Conventional (tablets, smartphones, laptops) or special purpose (e.g., devices installed on robots or drones).

While the majority of definitions given in Table 2 should be sufficient for the comprehension of the meaning of the related parameters and features, a specific explanation is necessary for the meaning of reliability and availability and their relationship. We propose the same definition given in Annex C of [7] where, by mean of Figure 7 (Figure C1 in [7] it is clarified that availability pertains the overall service and its specified QoS, and involve interfaces with the applications (but excluding them). On the other hand, reliability does not involve interfaces but only end-nodes and, by extension, the network portion on which the service is carried. In addition, Availability can be evaluated through QoS indicators related to the user experience (for instance the maximum waiting time for a response of an enquire on a web page) while usually reliability is evaluated on a parameter connected to the network operation like the bit error rate at transmission level or the packet loss at packet switching level. In summary, without going into deeper details, it can be shown that, in general and depending on the agreed QoS, the communication service availability can be of the same value as the reliability or much lower. (Annex C of [7] examples of Figures C2, C3 and C4).

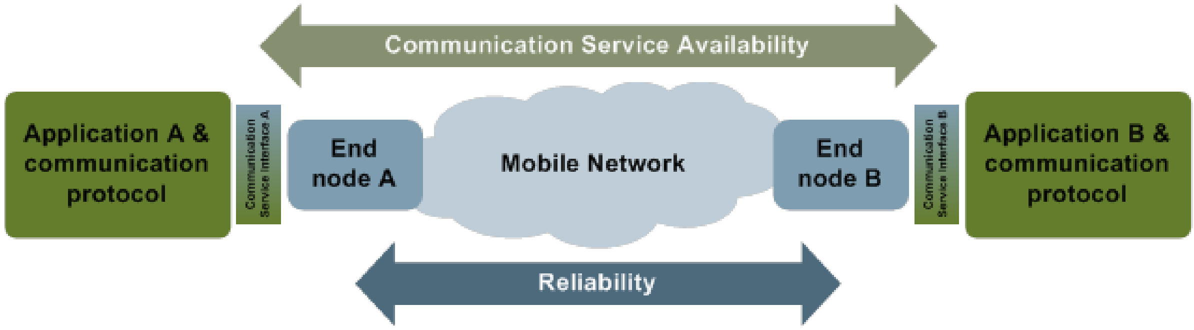


Figure 7 - Illustration of the concepts of communication service availability and reliability.

3.5 Guidelines to derive Metro-Haul network requirements from Use Case descriptions and service KPIs

In this section, we sketch the main procedure followed to produce a unified and clear view of the Metro-Haul network requirements and KPIs from a number of heterogeneous Use Cases, each of them having particular requirements in terms of the service KPIs detailed in Section 3.4. In particular, we propose to analyse each particular service Use Case individually and derive its specific requirements. After considering several Use Cases, an overall view of the Metro-Haul network requirements supporting all Use Cases is produced.

The first step consists in identifying (for a selected Use Case) the "crucial" or "critical" tasks to be accomplished within the Use Case and identify the associated service KPIs. A crucial task can involve communication, computing, data encryption or other jobs which are both essential and critical for service delivery, and result in challenges for the network in terms of required resources. An example of a crucial task is the communication between a robot and its remote control (*robot control loop*) which requires stringent latency, jitter and reliability constraints to the communication connection and computational and memory storage resources for virtual machines on servers located in the node premises. The objective of this task identification phase is precisely to isolate those tasks potentially requiring challenging metro network capabilities from those needing minor (or even negligible) Metro-Haul network requirements.

To facilitate the identification of crucial tasks from the service KPIs detailed in Table 3, the subset of relevant reference service KPIs in the Table 4 is proposed. Crucial tasks are identified taking into account the operations required by a Use Case that can have a strong impact on one or more of the service KPIs. Three main levels of requirement from a task to a service KPIs are identified: high-very high, medium and low with the corresponding value ranges being specified in Table 4. A task can be designated as crucial if it implies that at least one of the reference service KPIs is required to be at least at medium level.

Table 4 - Relevant reference Service KPIs and values ranges.

Relevant reference service KPI	*Level of requirement*		
	High- very high	**Medium**	**Low**
Data rate	100Mb/s – 1 Gb/s	1Mb/s-100Mb/s	< 1 Mb/s
Latency	<= 5 ms	5 ms – 100ms	>100 ms
Jitter	<= 10 µs	10 µs – 100 µs	> 100 µs
Availability	>= five 9s	four 9s – five 9s	< four 9s
Mobility (speed)	>= 100 km/h	10 -100 Km/h	< 10 km/h
Area traffic capacity	> 10 [Gb/s] / km^2	1– 10 [Gb/s] / km^2	< 1 [Gb/s] / km^2

Sensor density	>= 106 per km^2	<106 and >103 per km^2	<= 103 per km^2

Once crucial tasks have been clearly identified, the next step consists in mapping these tasks with Metro-Haul KPIs. This implies identifying the location and type (e.g. storage, compute, etc.) of crucial tasks onto the Metro-Haul architecture (i.e. AMEN, MCEN or connectivity in the metro network), highlighting the network elements or segments involved by the most challenging requirements.

Finally, the consolidation and aggregation of Metro-Haul KPIs over all the Use Cases is performed to produce an overall view of what it is expected in terms of resources and capabilities from a Metro network in a 5G era and compliantly to the Metro-Haul architecture.

4 Presentation of the selected 5G Service Use Cases

4.1 Summary of 5G Use Cases

A list of the 11 selected Use Cases for this deliverable and described in this section are reported in Table 5. The Use Cases are collected from proposals made by the partners on the base of their experience and knowledge and includes the two Use Cases that will be practically demonstrated by the project and which are extensively analysed in WP5. Even if this list of Use Cases does not guarantee an exhaustive coverage of all imaginable services, it provides a well-balanced samples among the different verticals and a good base for Metro-Haul network studies. All Use Cases are linked to end user (customer) services except the last one which is an operator orientated Use Case, in the sense that it enables the operator to provide services to the customers (in the specific case video services within the vertical Media & Entertainment) but it does not deal directly with a service to final customers. All the 5G Use Case listed in Table 5 involve the entire 5G infrastructure, including the access (mobile radio or fixed). Exception is made for Use Case N. 11 "Secure SDN Controlled Video Distribution", which relies on the metro and, potentially, the core of the network. Use Case N. 8 Enterprise Access with NG Ethernet, due to characteristics of the service (Enterprise high data rate fixed connectivity), does not use the mobile radio access.

Table 5 - 5G Use Cases for Metro-Haul network studies.

No	*Use Case name*	*Service Class*	*Vertical*	*Coverage extension*
1	mIoT Utility Metering	mIoT	Utilities	Smart City
2	Content Delivery Network	eMBB	Media & Ent.	Everywhere
3	Live TV Distribution	eMBB + URLLC	Media & Ent.	Everywhere
4	Service Robotics	eMBB + URLLC	Cloud Services	Smart City
5	Smart Factories	URLLC + eMBB+mIoT	Industry 4.0	Industrial Plant
6	6DoF Virtual Reality	eMBB + URLLC	Media & Ent.	Indoor hotspot
7	Intelligent environments to support Autonomous Driving	eMBB + URLLC	Automotive	Roads
8	Enterprise Access with NG Ethernet	BB + URLLC [6]	Cloud Services	Everywhere (not rely on mobile access)
9	Crowdsourced Video Broadcasts - Bristol Demo	URLLC	Media & Ent.	Crowd location
10	Real-time Low-latency Object Tracking and Security - Berlin Demo	URLLC	Pub. Saf. & Env.	Smart City

[6] Enterprise Access with NG Ethernet Use Case requires high level of performance on data rate, latency and reliability, but it does not involve the "mobile" access part of 5G infrastructure. The CoS is then assigned as a general indication (in particular assigning the class Broad Band, obtained removing enhanced Mobile from eMBB) as CoS listed in subsection 3.2 is conceived for services relying on a mobile access.

11	Secure SDN Controlled Video Distribution	BB + URLLC [7]	Media & Ent. (Operator perspective)	Everywhere (not relies on mobile access)

In this section the Use Cases listed in Table 5 are individually presented according the analysis strategy presented at the beginning of Section 3, with the definitions of service KPIs and guidelines to derive Metro-Haul requirements specified in the same deliverable section.

4.2 mIoT Utility Metering

4.2.1 Use Case description

The advent of Internet Technologies and mobile network capabilities such as LTE or even 5G, is connecting more devices to the Internet every day. Very often, new devices requiring connection appear, providing new services which were previously unthinkable. It is estimated that at least 50 billion devices will be connected by 2020 [5] with new emerging M2M scenarios, such as smart metering/monitoring, e-health, and autonomous vehicles.

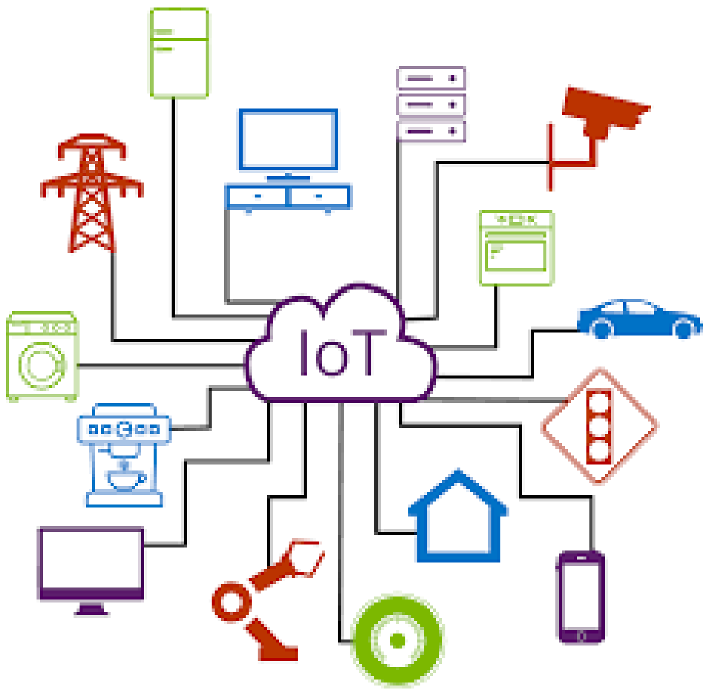


Figure 8 - Internet of Things concept.

Indeed, the shape and amount of traffic online will experience several changes. Many devices will operate autonomously without human interaction. In addition, the limited processing and storage capabilities of devices will reshape how network components are required to interact.

As a result, the amount of data generated from such interactions will generate traffic volumes never expected, combined with the speed IoT development will require enabling Internet infrastructure components development to provide reliable IoT services. Due to IoT, cohabitation of M2M and

[7] For Use Case Secure SDN Controlled Video Distribution as regard CoS, it is the same as in the previous note.

conventional user traffic together with quickly increase of connected devices to cellular infrastructure will unveil new challenges for a network in need of evolution.

In general, most IoT elements fall into Machine-Type Communication traffic (MTC), commonly used to identify traffic entirely generated and managed by machines, without human intervention. This traffic is typically uplink-dominant, characterized by low duty-cycles of short packets of periodic or event driven nature. Real-time, aggregation and synchronization requirements are usually application-dependent.

Regarding traffic patterns, there are three possibilities in MTC according to [13]:

- Periodic Updates (PU): A device transmits status reports or updates to a central unit on a regular basis. E.g. Smart-metering of water, gas, electricity, etc.
    - PU does not require real-time transmission and its traffic is typically similar and happens at regular intervals, which can be tuned.
- Event-Driven (ED): A specific event in the device triggers the communication initiation. E.g. a certain measurement exceeds a predefined threshold which requires an alarm, emergency notification, etc.
    - ED often requires real-time transmission, its shape being completely dependent on the triggering event.
- Payload Exchange (PE): This pattern comprises all cases where large amounts of data need to be exchanged between the sensing devices and a server. E.g. Constant size data as in telemetry, or variable size like the transmission of an image.
    - This traffic shape and periodicity depends on the sensor and the type of the event.

As an example, consider a smart metering application in Central London, where the household density is approximately 5000 households/cell. Assuming three hourly measurements (gas, electricity and water consumption) per household, the application generates 15,000 messages/hour in each cell [14]. If each message comprised 10 Kilobyte (KB) of information, each cell would experiment an average traffic of:

$$\frac{15{,}000 \ x \ 10KB}{3{,}600s} = 330Kbps$$

A number which should be multiplied by the total number of cells served simultaneously by a Metro-Haul node.

For the case of an ultra-dense geotype, such as Central London, one Central Office covers 2 km^2 ([15], Appendix 1), so each CO covers about 5-10 cells, serving 2,440 buildings containing approximately 15,820 household.

Despite its low average rate, spiky-like traffic should be expected, as most periodic updates will collide during the same time frame (on the hour, half past, etc.). This characteristic may pose challenges in the metropolitan network in case sensors provide their metering information in a synchronous manner. In particular, such a worst-case scenario in a Central London's CO would imply a peak traffic in the order of:

$$\frac{15{,}820 \ x \ 3 \ x \ 10KB}{1s} = 3.8 \ Gbps$$

4.2.2 Components and functionalities and their mapping on Metro-Haul architecture

IoT utility metering is a very versatile scenario where intermediate processing and network capabilities requirements may depend on the final metrics extracted. Any Metro-Haul node is enabled to process IoT metering traffic to transform it into a more optical network-friendly traffic pattern.

The main functionalities to be provided by the network to IoT traffic are traffic pre-processing, aggregation, and synchronization capabilities at the Metro-Haul nodes. Figure 9 shows their location in Metro-Haul architecture.

- Traffic pre-processing analyses and classifies the metering packets coming from final sensors following sets of rules that facilitate the aggregation. That is, packet inspection is performed to process relevant information for aggregation purposes.
- Traffic aggregation of utility metering packages consists on the definition and application of standard aggregation schemes and methods to combine messages into groups or new types of messages containing a group of relevant data chunks from different sensors to be uploaded to each service's central servers. The aggregation schemes could go from simple grouping of packets together into a larger packet or creating specific packets containing several measurements together to performing simple operations over sensor measurements such as summing or averaging by access (AMEN) or metro (MCEN) networks.
- Furthermore, synchronization and sanity checking are other key elements for smart nodes supporting IoT utility metering. Metro-Haul nodes might check sensor measurements and timestamps and directly require new measures upon anomalies. Leveraging on this functionality, other analytical functions, namely, alarm triggering, decision making, action execution or behaviour learning may extend the node abilities to achieve latency and network traffic load reduction. In addition, other application-specific capabilities would enable more flexibility, quick response to user actions and overall service improvement.

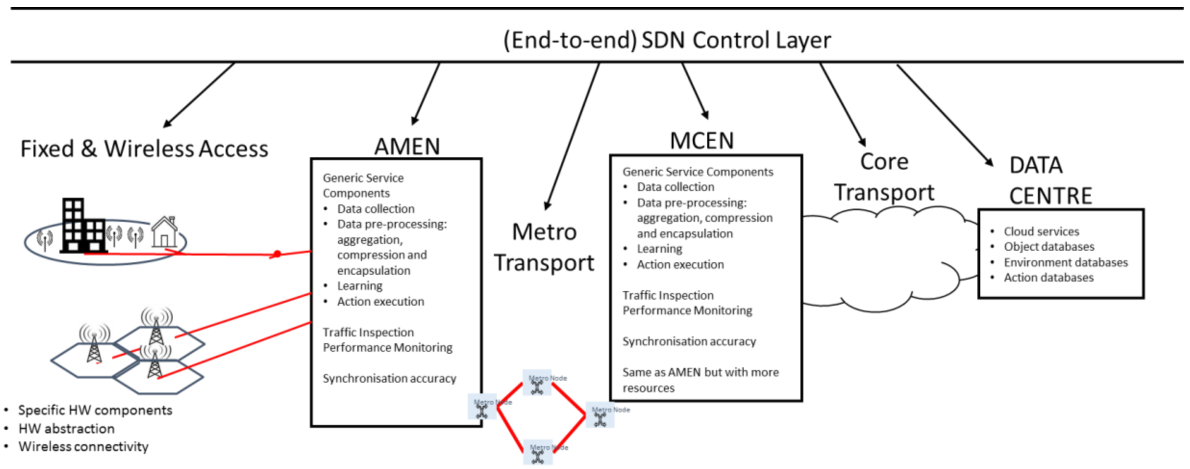


Figure 9 – Mapping of components on Metro-Haul architecture for mIoT Use Case.

In the light of these requirements, Metro-Haul nodes should be provisioned with medium storage capabilities, enough space to temporally allocate sensor messages and meta-data as well as accurate synchronisation capabilities, between AMENs, MCENs, sensors and Central services; reliable timestamp generation in AMEN and MCEN nodes would be essential. Moreover, NFV capabilities and processing power would be necessary to enable application-specific instances and perform the three main functions previously referred: (i) pre-processing and classification, (ii) aggregation and (iii) synchronization with sanity check for decision making and/or action execution.

4.2.3 Service Requirements

Table 6 - KPIs of service required by mIoT Utility Metering.

1	**Bandwidth (data rate) [Mb/s]**	Low, typically between 1 – 100 kb/s per sensor From few bytes per message to few KB From few Byte/s to few KB/s per sensor Spiky traffic pattern
2	**Latency [ms]**	Less than 5 ms on high-priority apps, up to 1 minute on low-priority IoT apps
3	**Jitter [ms]**	Unknown
4	**Reliability [%]**	high (five 9)
5	**Availability [%]**	high (five 9)
6	**Mobility - UE speed [km/h]**	0 km/s static sensors 100 Km/h vehicles
7	**Area Traffic Capacity [Gb/s/km^2]**	Ranging from 1,000 to 200,000 IoT devices per km^2 in big cities. About 5-10 Base Stations per km^2 => About 20,000 – 50,000 sensors per cell
8	**Rely on Sensor Network [Yes/No]**	Yes
9	**Massive Type [Yes/No]**	Yes
10	**Device Direct [Yes/No]**	Possible
11	**Coverage Requirement [Standard/Extended]**	Everywhere
12	**Energy Consumption critical [Yes/No]**	Yes
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Special purpose

4.2.4 Metro-Haul network requirements

The requirements identified are grounded on certain elements which must be included in the design and fabrication of Metro-Haul nodes. In general, such requirements aim to bring computing power to the metro network and enable in-node analysis and aggregation to reduce the network traffic load with every application central servers while reducing latency by solving many issues on the metro network directly.

The sensor and actuator control actions, including taking actions, reading measurements or updating configuration, requires a lightweight packet traffic processing layer feasible on Metro-Haul nodes. Each node may take critical decisions on the sensors and actuators connected to them without checking with the central server. For instance, if the service needs to update sensor parameters on a global scale, metro nodes might store that update and send it whenever any sensor from its access network becomes active.

Besides the control plane, nodes, through virtualized application appliances, might understand and modify application messages. This way, edge nodes can take actions or reduce traffic load by aggregating, sanity checking and improving the efficiency of usage of the transport network. Consequently, metro-core network traffic is simplified and reduced, since most time and resource-consuming tasks are decentralized and do not send traffic through metro networks.

The main components required for Metro-Haul equipment are therefore provision of computing resources, namely storage capacity for temporary caching records, processing capabilities to perform simple operations and application containers for application-oriented appliances to run.

In this light, the introduction of NFV capabilities in Metro-Haul nodes would enhance the easy and fast introduction of application-oriented appliances developed by each application developers. This way, application developers (in-node appliances) and network operators (NFV capabilities and resource provisioning) would work on a shared effort to improve network performance and usage while enhancing user experience and all these elements would be supported by Metro-Haul design and architecture.

In summary, regarding the mIoT Use Case, the following critical tasks have been identified:

1. Guaranteed delivery of messages, some with important latency and jitter bounds
2. Transport of spiky traffic patterns, as a consequence of certain sensors coordinating their transmission reports and updates
3. Data pre-processing and/or filtering and or/compression before transmission across the MH node (this function may be even virtualised at the AMEN or MCEN)
4. Aggregation of small packets into large ones to reduce overheads (this task may also be virtualised at the AMEN or MCEN)

These critical tasks demand the throughput, storage and computing capacity requirements, along with additional functionality regarding monitoring and data analytics plus e2e orchestration, management and control (i. e. Metro-Haul KPIs) as reported in Table 7.

In the example of utility metering for central London used in the text, let us assume 5,000 households per cell generating a total of 15,000 messages/hour (4.17 messages/sec). We consider that a single AMEN node collects the messages produced in 10 cells, i.e. a throughput of approximately 40 messages/sec on average. However, under the assumption that traffic is spiky, i.e. many messages will be concentrated in short periods of time (at o'clock), let us further assume a peak to average ratio of 10x.

Under this assumptions, 2-3 vCPUs should be enough to handle 400 messages/sec as long as the back-end processing requires simple tasks like storing a couple of measurement readings in a database plus updating some counters. If further processing is needed, the number of vCPUs should be revised.

Table 7 – Summary of Metro-Haul requirements (KPIs) by mIoT Utility Metering Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Guaranteed delivery of messages, something with latency and jitter bounds	10 Gb/s		Fat pipes for faster transmission				
Transport of spiky traffic patterns	10 Gb/s		Spiky				

			Capacity demand 10 Gb/s				
Data pre-processing and/or filtering and/0r compression				1 GB		1 vCPU	
Aggregation of small packets into large ones				1 GB		1 vCPU	

Task	Monitoring and Data Analytics	Management, Control and e2e Orchestration
Guaranteed delivery of messages, something with latency and jitter bounds	Transmission out of service detection	Fast recovery mechanisms
Transport of spiky traffic patterns	Transmission degradation detection/anticipation	Optimal Planning of capacity resources
Data pre-processing and/or filtering and/0r compression	VM monitoring	VM dynamic dimensioning (local reconfiguration)
Aggregation of small packets into large ones	VM monitoring	VM dynamic dimensioning (local reconfiguration)

4.3 Content Delivery Networks

4.3.1 Use Case description

Most of the traffic in today’s network is generated to support content dissemination, mostly associated to video delivery. As an example, Video-on-Demand (VoD) is expected to represent approximately 85% of the global consumer traffic by 2018, and 75% of global mobile data traffic by 2019 [16]. The increasing importance of video traffic, especially in mobile cellular network traffic, is shifting the Internet network architecture from its traditional host-centric (connection-centric) architecture, to a content-centric architecture in which the goal is to retrieve a video content, independently of its location. Moreover, VoD services keep evolving, and today this evolution is mainly driven by the adoption of new image resolutions such as 4K and 8K. With the advent of 5G, mobile users are expected to be provided with even higher bit-rates, allowing them to receive enhanced video qualities also on their mobile devices.

Solutions to improve content-delivery traffic and overall efficiency includes network operators deploying specialized Content Delivery Networks (CDNs). A CDN is a geographically distributed network of caches (in the form of proxy servers or data centers), with the objective of decreasing network traffic (as content is located closer to users) and thereby improving the experience of end users. Although video is not the only type of content supported by CDNs, the recent increase in video traffic made video delivery services the focus of CDN operators. To cope with the continuous increase of video traffic, network operators are still urged to enhance their CDNs, taking advantage of new architectural trends as Network Function Virtualization (NFV) and Edge Computing. A promising solution is to equip network edge nodes with storage and computing capabilities, allowing them to store (cache) and deliver contents as, e.g., popular video contents. Note that caching of popular video contents is particularly effective due to the nature of the VoD content popularity, which is known to follow a power law distributions (a commonly-used model is the Zipf distribution), where a small portion of popular video contents account for a high percentage of video requests (short head) and a big portion of video contents account for a low percentage of requests (long tail), as shown in Figure 10. For example, the caching of the most popular 10% of video contents of a given content provider at the edge nodes of the metro-aggregation offloads 70% of video data from the metro/core network. However identifying the optimal cache deployment (in terms of number, location and dimensioning of caches) remains an open issue as it depends on various topological parameters (e.g., size of the metro transport network, location of the MCEN), geotype area (e.g., urban, rural) and content catalog characteristics (e.g., popularity distribution, catalog size). This problem will be investigated during the project in the specific context of the Metro-Haul network architecture.

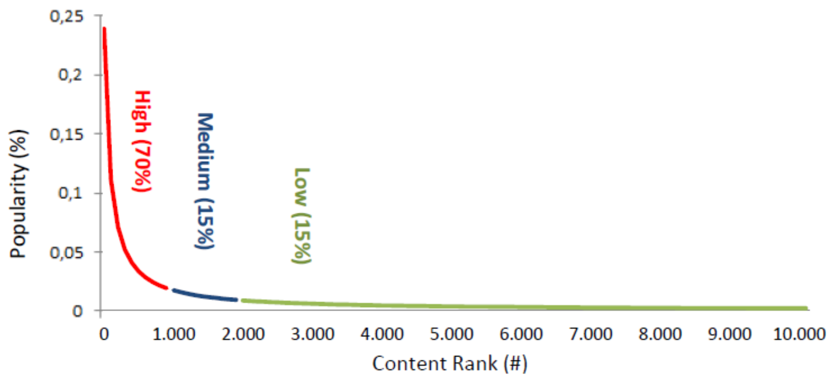


Figure 10 – Example of a Zipf popularity distribution of a content catalogue of 10,000 video contents.

4.3.2 Components and functionalities and their mapping on Metro-Haul architecture

In the vision of Metro-Haul network architecture, the CDN caches popular video contents at AMENs and MCENs. The main functionalities to be located at AMENs and MCENs (Figure 12) to enable efficient video caching and delivery are:

- Storage capacity to store popular contents. Storage capacity required is in general in the order of tens of TBs depending on the cache deployment and the type and location of node (AMEN or MCEN).
- Video interfaces to deliver video contents. The required interface capacity mainly depends on the area demography and the traffic patterns. Since the video delivery is one of the most bandwidth-demanding services (some user connections are hours-long) the video interface capacity is generally in the order to tens of Gb/s.
- Traffic inspection capabilities to inspect packets, analyse and classify video contents into various categories, i.e., to discover which contents are popular among users in specific areas or to detect which contents are no longer popular.

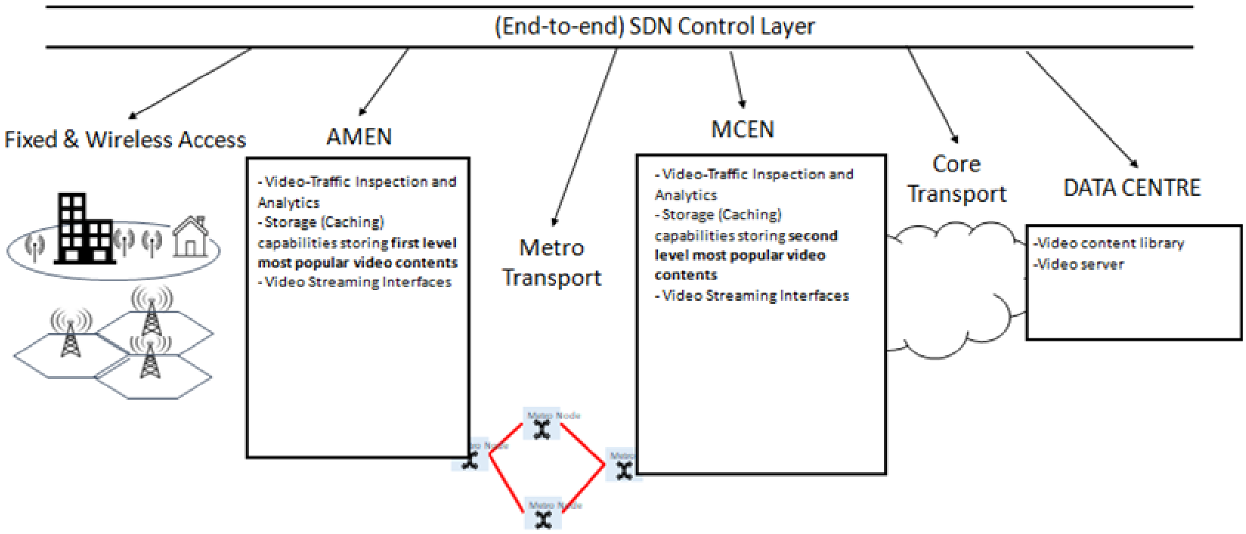


Figure 11 – Mapping of components on Metro-Haul architecture for CDN Use Case.

The AMEN and MCEN nodes should be equipped with enough computing capabilities to collect and inspect data in real time. The storage capacity of a cache is dimensioned to store popular video contents allowing it to serve a desired percentage of video requests. Moreover, the video delivery interfaces must account to serve the maximum number of users, mainly during evening hours which are characterized by the highest bandwidth consumption, such as to avoid any degradation of the service.

4.3.3 Service requirements

Table 8 shows the KPIs of this Use Case. Bandwidth needs for VoD services are due to the number of users and the video qualities requested by them. The bandwidth requirements per user depend on the encoding format and the rate of the video stream. For example, a high video quality (e.g. Full High Definition) requires a bit-rate of 12 Mbps. The latency requirements depend on the video application and the buffering capabilities and are not very stringent (4 to 5 seconds), as the video could take several seconds to cue up. Similarly, the jitter requirements are not very significant because of the video application buffering. However, some video equipment begins having troubles at a jitter higher than 10 milliseconds. The area traffic capacity depends on the area demography and as well the time of the day, as VoD is known to have peak hours during the evenings and weekends.

Table 8 - KPIs of service required by Content Delivery Network.

1	**Bandwidth (data rate) [Mb/s]**	Around 8 Mbps per user depending on video definition. Bit-rates vary during same session (Dynamic Adaptive Streaming over HTTP (DASH)). Expected to become higher with adoption of 4K and 8K definitions (around 20 Mbps)
2	**Latency [ms]**	Low (4-5 seconds of latency is accepted)
3	**Jitter [ms]**	Medium to high (10-20 ms)
4	**Reliability [%]**	Medium to high
5	**Availability [%]**	Medium to high
6	**Mobility - UE speed [km/h]**	0-5 km/h pedestrians 60-100 Km/h for users in vehicles, buses, trains
7	**Area Traffic Capacity [Gb/s/km^2]**	Medium to high (depends on area demography)

8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	No
11	**Coverage Requirement [Standard/Extended]**	Everywhere
12	**Energy Consumption critical [Yes/No]**	No
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Mobile phones, laptops, smart TV, tablets, phablets

4.3.4 Metro-Haul network requirements

To provide a quantitative estimation of the KPIs of this Use Case, we refer to three crucial tasks that could be identified as performance bottlenecks, as they have stringent requirements. Then we will show how Metro-Haul architecture will be able to support such challenging requirements. The crucial tasks are:

- UHD/4K/8K video streaming for fixed and mobile users
- Peak-hour video streaming and flash crowd phenomenon
- Video-traffic inspection/monitoring and analysis and cache reconfiguration

Streaming of high-definition videos requires high data rate, e.g., 12 Mbps per user for a 4K or 8K video, and is known to have peak hours during evening hours and weekends. This poses on the network a strong pressure in terms of bandwidth requirements. This can be further aggravated by flash-crowd phenomena (i.e. sudden increases in request arrival rate for a new viral video content) that might induce network congestion due to the huge amount of traffic to be transported in a short amount of time. Moreover, an effective classification and selection of content for the caches requires the AMEN and the MCEN to be able to monitor, inspect and analyse the transiting video traffic. This traffic monitoring allows to reconfigure the caches (store or remove a content) thus improving the cache efficiency. In Figure 12 we show a mapping of the selected critical tasks with their respective critical KPIs.

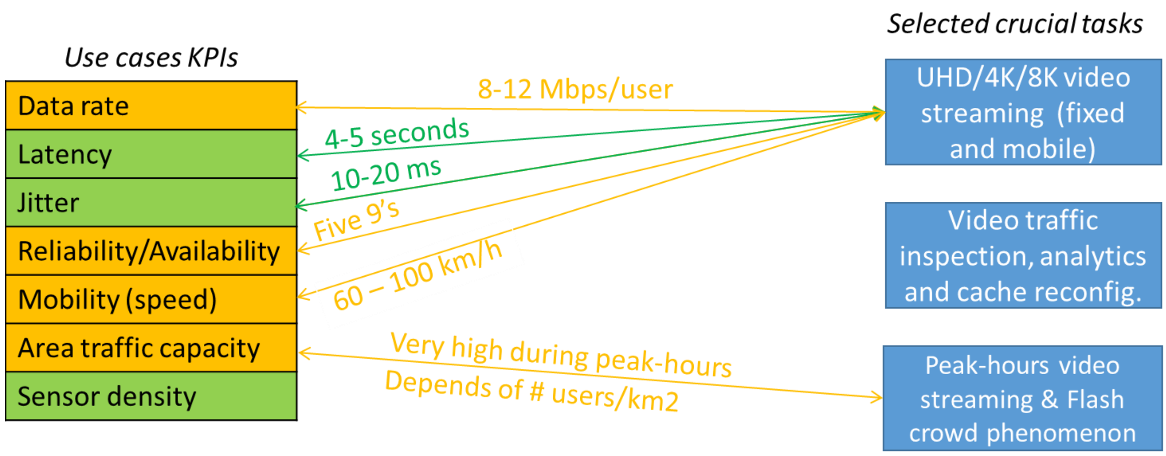


Figure 12 – Mapping of critical KPIs to the selected crucial tasks in the CDN Use Case.

These crucial tasks can be executed in the metro segment (i.e. they can be executed directly in the AMEN and/or in the MCEN). In fact, caching of popular video contents at the AMEN and MCEN allows to terminate video requests at network edge (instead of the remote video server) and this further allows to improve the QoS delivered to end-users.

Hence, in this Use Case, network requirements for AMEN and MCEN are similar. Both nodes must host video caches capable of inspecting, analysing, storing and delivering video contents *in real time*, which requires a high amount of processing and storage capabilities. Depending on the considered geotype (e.g., urban, rural) and the characteristics of the video-content catalogue, the network requirements (amount of storage capacity to host caches and video-delivery interface capacity) at AMEN and MCEN can vary. To provide a quantitative estimation of the network requirements we consider the following case study characterizing a video-content catalogue and a specific geotype. The video-content catalogue is characterized by its size (number of contents) and its popularity distribution. We consider a catalogue size of 10,000 contents following a Zipf distribution characterized by a skew parameter α = 0.8.

The content size range between 2GB (e.g., short duration episodes) and 9 GB (e.g., long duration movies). The contents sizes follow a power-law distribution with average size of 4.5 GB. We assume a dense-urban geotype where the maximum number of users connected to the AMEN U_{max} is 10,000 and the average video delivery bit-rate b_r is 8 Mbps. Furthermore, we assume there are 6 AMENs connected to the MCEN.

Given the abovementioned data and assuming that the desired hit-ratio H_{amen} of a cache located at AMEN is 0.8 (i.e., cache satisfies 80% of the video requests of users connected to it), the cache capacity (in terms of its size (GB)) is calculated considering the popularity distribution, i.e., we sum the capacity needed to store the most popular contents whose aggregated popularity sums up to 80% (a hit-ratio of 0.8). In our case study, the number of contents which account for 80% of the requests is 5000 requiring a storage capacity $Scap$ of 22.5 TB. The corresponding video interface capacity $Icap, amen$, assuming half of the maximum users connected to AMEN request a video during peak hour, can be calculated as follows:

$$Icap, amen = \frac{Umax/AMEN \bullet br \bullet Hamen}{2} = 32\ Gb/s$$

As for the cache located at the MCEN, assuming its desired hit-ratio is 0.5 (i.e., it satisfies 50% of the requests not satisfied by the AMEN caches), its overall hit-ratio H_{mcen} is $0.2\ x\ 0.5 = 0.1$. Similar to the AMEN cache capacity calculation, MCEN cache capacity is calculated according to popularity distribution by finding the number of contents whose popularity sums up to 10% starting from the first content that is not cached in the AMEN (i.e., starting from content number 5001). According to the considered content catalogue, the number of contents is around 2500 (ranging from 5001 to 7500). Thus, the required storage capacity at the MCEN cache is 11.25 TB. As for the video-delivery throughput capacity required at the MCEN, it must be enough to satisfy video requests of video contents stored/cached in the MCEN cache originated from users of AMENs connected to the MCEN. Therefore, the video-delivery interface capacity required at the MCEN to deliver videos to users is calculated taking into consideration the number of users of all AMENs connected to the MCEN, the average bit-rate of all users and the MCEN cache's hit-ratio H_{mcen}:

$$Icap, mcen = \frac{\#AMENs \bullet\ Umax/AMEN \bullet\ br \bullet\ Hmcen}{2} = 24\ Gb/s$$

Moreover, to perform video traffic inspection, AMEN and MCEN must be capable to inspect all the video traffic which passes through the nodes. Thus, in this example, the required video traffic inspection and analysis throughput capacity at AMEN is $\frac{Umax \bullet\ br}{2}$ = 40 Gb/s while that at the MCEN is $\frac{\#AMENs \bullet Umax \bullet br \bullet (1-Hamen)}{2}$ = 48 Gb/s. Note that the AMEN inspects all video traffic generated by users connected to it while the MCEN inspects only the video traffic of requests not served by the AMEN. The optical throughput at the AMEN in this case study is 40 Gb/s, relative to the total amount of video traffic transported to end-users, while that for the MCEN is 48 Gb/s (to account for 4 Gb/s of video traffic generated from the MCEN and 4 Gb/s video traffic generated from the data centre to each of the 6 AMENs). Considering that the path length from the MCEN to the AMEN is of a maximum of 3 hops, the optical throughput at either nodes is in the order of few tens of Gb/s. As for the computing capacity needed for video traffic inspection at AMEN and MCEN, we consider the computing capacity of a similar function, i.e., deep packet inspection (DPI), performed by a DPI-engine [17], that is 10 Gb/s per core CPU. In the considered case study, the number of CPUs required at the AMEN and the MCEN is 4 and 5, respectively.

Table 9 summarizes the network requirements of the selected tasks at both the AMEN and the MCEN for one content provider (one content catalogue).

Table 9 – Summary of Metro-Haul requirements (KPIs) by Content Delivery Network

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
UHD/4K/8K video streaming	32 Gb/s[8]	24 Gb/s [9]	Nx10 Gb/s	22.5 TB	11.25 TB	-	-
Video traffic inspection and analysis	40 Gb/s	48 Gb/s	10 Gb/s [10]	-	-	4 vCPU	5 vCPU
Peak-hours/Flash crowd phenomenon	32 Gb/s	24 Gb/s	-	-	-	-	-

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
UHD/4K/8K video streaming		Fast recovery mechanisms (protection and restoration)
Video traffic inspection and analysis	Traffic monitoring for early detection of new popular videos	Local and Global reconfiguration of virtual cache
Peak-hours/Flash crowd phenomenon	Traffic monitoring for fast detection of a flash crowd phenomenon	Local and global reconfiguration of virtual cache

4.4 Live-TV Distribution

Live-TV is also one of the stringent and more popular services that a telecom network needs to support (forecasts show that 82% of the global IP traffic will be related to video traffic by 2021 [20]). In this Use Case, we will introduce the peculiarities of the live-TV service and illustrate how the deployment and dynamic adaptive reconfiguration of virtualized cache components within the

[8] Shared with *Peak-hours/Flash crowd* task.
[9] Shared with *Peak-hours/Flash crowd* task.
[10] Per content provider.

Metro-Haul infrastructure can strongly help to make possible achieving the stringent requirements of such service.

4.4.1 Use Case Description

In live-TV distribution, the bandwidth needed to convey a video stream is actually determined by its quality. To allow adapting video stream to a wide range of qualities, uncompressed video streaming formats are used before video production. Uncompressed video streaming supporting 4K Ultra-High Definition (UHD) TV format ranges from 6 to 48 Gb/s, according to ST 2036-1 [18], while uncompressed real time 8K UHD transmission needs 72 Gb/s connections [19]. Those uncompressed video streams are then processed and adapted to compressed streams for video distribution, requiring up to hundreds of Mb/s per individual user depending on its quality, i.e., standard definition (SD), high definition (HD) or UHD (Figure 13).

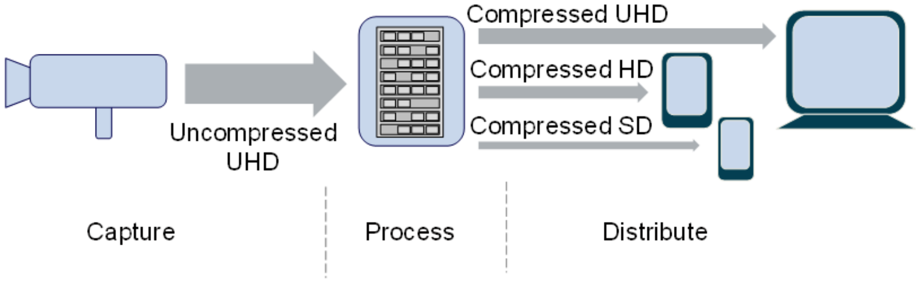


Figure 13 – Live-TV distribution schematic stream process.

Video signal processing can be performed in commodity hardware inside large data centres placed in the core segment and distributed directly towards the end-users (*centralized* approach). In that case, a large number of small flows need to be conveyed to the metro networks, which are commonly used to aggregate users' traffic. On the contrary, if several small DCs or computing nodes performing signal processing are placed closer to end-users, uncompressed UHD video signals need to be conveyed from the signal source to each of the nodes holding computing resources for video processing (*distributed* approach). Authors in [21] presented a scalability study (based on network CAPEX) where centralized and distributed architectures for live-TV distribution were evaluated for a realistic scenario in terms of metro and core topologies and forecasted service demand. The main conclusion extracted from that study is that a distributed approach taking advantage of computational resources close to end-users, i.e. in metro nodes, entails significant overall network cost reduction in terms of L2 equipment and optical transponders.

The distributed approach can be fully exploited by placing *virtualized caches nodes* in those network nodes with computing capabilities. Virtualizing caching capabilities facilitates rapid distribution and/or scaling of cache nodes in a cost-efficient and scalable manner. For instance, as a result of using the standardized MPEG Dynamic Adaptive Streaming over HTTP (MPEG-DASH) [22] technique, multimedia contents can be served by HTTP servers. Another cache component must be in charge of generating DASH segments in several qualities and the related Media Presentation Description (MPD) files. However, the component that requires the most computational effort is video transcoding / transrating, although it can be implemented in software and performed in real-time (see [23] for available software implementations).

A virtualized leaf cache node would consist of the following components running as software inside VMs: *i*) the *packager* is in charge of live-TV preparation, including stream transcoding/ transrating, segmentation and MPD generation, *ii*) the *HTTP server* component serves end users' segment requests, *iii*) *the Cache Manager* is the entry point of the cache node; it receives users' requests, identifies which contents will be locally stored (if any), and redirects users' requests to the appropriate HTTP server. Each component usually consists of a pool of resources for load balancing and redundancy purposes.

One of the main benefits of using a dynamic, adaptive virtualized infrastructure is the capability to fit with actual demand needs without incurring into expensive resource overprovisioning. In this regard, authors in [24] proposed an architecture fulfilling the ETSI NFV guide lines that allows controlling virtualized components while collecting and pre-processing monitoring data from both network and computing resource utilization as well as users' experience metrics. Local, e.g. at computing nodes, and global (network-wide) re-optimization problems to reduce resource utilization while ensuring the highest quality to end users were presented. Data stream mining, regression and machine learning techniques were proposed to produce estimation of future scenarios for such re-optimization. Numerical results over a realistic scenario showed huge reduction (up to 83.5%) in the total number of allocated HTTP servers and bandwidth resources compared with a static approach based on dimensioning for the peak users' demand. The effect of allowing adding and releasing new leaf cache nodes was also analysed; remarkably network costs reduction as high as 33% can be achieved by placing transcoding close to the end users.

The next subsection presents a mapping of virtualized cache infrastructure components and functionalities for live-TV distribution on the Metro-Haul architecture.

4.4.2 Components and functionalities and their mapping on Metro-Haul architecture

Figure 14 illustrates a proposal of mapping where AMEN nodes can be mainly used to host leaf caches that are adapted according to on-demand users' needs, whereas MCEN is a good place for those caches that are fixed, e.g. to give uninterrupted service during night hours to a small but widely spread amount of end users. In the example, we assume that contents can be produced either out of the metro domain (coming from an external source) or within the metro domain; in any case, we could assume that uncompressed streams (one 72 Gb/s flow for each TV channel) are distributed from MCEN to AMENs with active leaf caches.

To allow re-optimization based on data analytics, both network and computational resources need to be monitored. According to the general approach of the data analytics architecture proposed in the Metro-Haul, local Knowledge Discovery in Database (KDD) and Data Mining (DM) capabilities in AMEN nodes can be used for local reconfiguration, e.g. using predictive models to anticipate resource capacity violation and triggering scaling up of HTTP servers. Global KDD and re-optimization targeting overall reconfiguration should be executed in a centralized way.

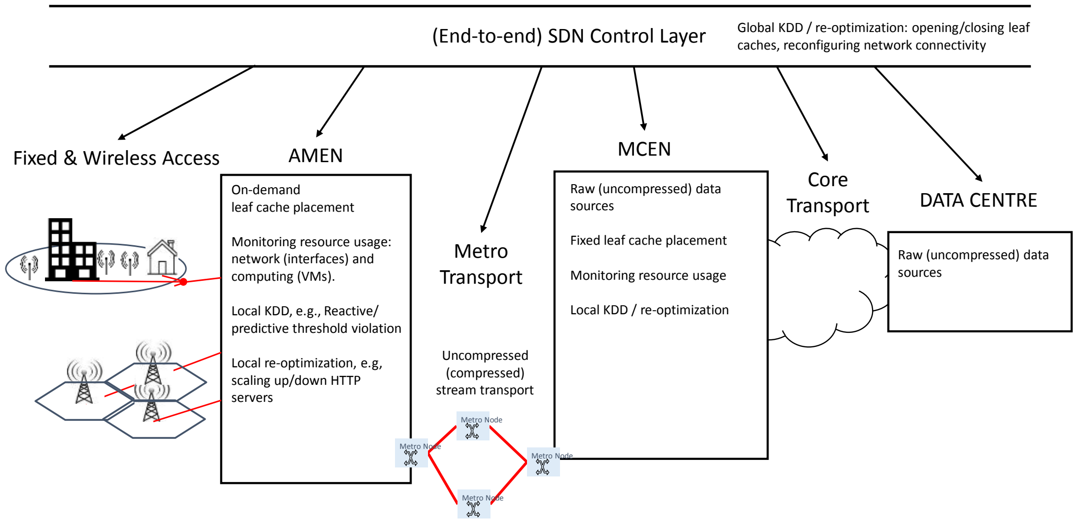


Figure 14 – Components and functionalities mapping

4.4.3 Service requirements and KPIs

The transport of uncompressed raw flows (one per TV channel) poses the strictest requirement in terms of network bandwidth. However, although compressed flows are small, the need of one stream per individual user makes significant the data rate that HTTP servers need to manage. Just an illustrative example, an AMEN node serving 1000 users of on-average moderate quality (e.g. 10 Mb/s) should provide HTTP servers being able to deliver processed flows at 1-10 Gb/s.

Among several KPIs, those related with end users' quality of experience are the most critical. Live-TV is very affected by jitter and packet loss, effects that lead to a poor experience characterized by undesirable effects such as pixelisation, tilling, frame freezing, or blue screen. Although those effects could be partially dimmed by the adaptive bitrate streaming of MPEG-DASH, video quality downsizing also represents an undesirable loss of experience. Therefore, jitter and reliability are critical KPIs that must be kept under control by means of a network infrastructure allowing the delivery of thousands of individual end user flows at a wide variety of rates with a minimum constant throughput to avoid video quality downsizing.

Table 10 – KPIs of service required by Live-TV distribution

1	**Bandwidth (data rate) [Mb/s]**	Uncompressed video stream (per channel): • High (~6 - 18 Gb/s for moderated colour depth 4K UHD) • Very high (~48 - 72 Gb/s for high colour depth 4K-8K UHD) User streams (per user) • Low (2.1-4 Mbit/s for SD and HD) • Moderate (10 Mbit/s for full HD) • High (25 Mbit/s for 4K UHD)
2	**Latency [ms]**	High (tens-hundreds ms)
3	**Jitter [ms]**	Very stringent (<1 ms)
4	**Reliability [%]**	High (five 9)
5	**Availability [%]**	Moderate (99.99%)
6	**Mobility - UE speed [km/h]**	3 km/h (Pedestrians) 50 km/h (Vehicles)

7	**Area Traffic Capacity [Gb/s/km^2]**	[Normal] 10 Gb/s/km^2 (peak – downlink) [11] [Hotspot] 500 Gb/s/km^2 (peak – downlink) [12]
8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	No
11	**Coverage Requirement [Standard/Extended]**	Standard
12	**Energy Consumption critical [Yes/No]**	No
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Conventional

4.4.4 Metro-Haul network requirements

To meet a desired service performance in terms of crucial service KPIs, there are crucial tasks to be deployed in the Metro-Haul infrastructure that directly impact on service quality. Specifically, the most relevant tasks are:

- *Uncompressed flow distribution*: captured raw video streams are distributed through the metro segment from the entrance point, e.g. placed in a MCEN, to all leaf cache nodes, e.g. placed in various AMEN. Those flows require high rates (up to 72 Gb/s per channel) and high availability to avoid disrupting video service to a large number of users due to the stream interruption. Then, metro transport network needs to deal with such type of flows in a reliable way but also exploiting optical link capacity resources. In this regard, the adoption of advanced modulation formats and/or even innovative efficient transmission techniques such as signal overlap with protection capabilities [25] need to be fostered to better support this task.
- *Video processing*: although this task is actually composed by a concatenation of several sub-tasks, we can define video processing as all the processes needed to transform an uncompressed raw flow into several compressed flows, one per each of the qualities and formatting that users require, e.g. SD, HD,4K UHD, 8K UHD, etc. This task is the main responsibility of packagers in virtualized cache nodes. Note that this task requires real-time processing of large uncompressed video streams with the objective to produce processed segmented video streams at a constant bitrate. Hence, computational resources (i.e. VMs) allocated for packagers need to guarantee that required performance, avoiding undesired negative impacts such as an increase of overall e2e latency due to computation performance degradation.
- *End user video delivery:* segments are sent to end-users upon request by HTTP servers. Note that HTTP servers need to continuously deal with individual requests sent by end-user video clients, so a pool of VMs is required to support all the demand. In fact, this pool can be adapted to the actual needs as proposed in [24], provided that monitoring and analysis of servers' activity and pro-active leaf cache node reconfiguration is available. However, it is important to mention that an improper dimensioning of HTTP servers and/or load balancing

[11] Based on a hypothetical case of 150,000 subscribers in an area of 100 km2 (Barcelona) in prime time (70% audience) and average video quality ~10Mb/s.

[12] Based on a hypothetical case of ~10,000 subscribers in a crowd event (i.e. last season soccer match at stadium) watching simultaneous live-TV contents (another soccer match at other stadium), video quality ~4 Mb/s.

could negatively impact on jitter and reduce overall throughput offered to deliver video segments. In that case, quality of delivered video would be degraded in order to keep serving all user requests.

The following table presents some tentative figures of key Metro-Haul requirements. Values have been taken based on the studies and analysis performed in [24]. Note that only relevant cells for this Use Case have been filled.

Table 11 – Summary of Metro-Haul requirements (KPIs) by Live-TV distribution

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Uncompressed flow distribution	[6, 72] Gb/s per TV channel						
Video processing (*packager*)	1 Gb/s per packager					1 VM (8 vCPUs)per TV channel	
End-user segments delivery (*server*)	1 Gb/s per server					1 VM (8 vCPUs) per ~500 users	

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Uncompressed flow distribution	Quality of Transmission monitoring and degradation detection/anticipation	Reconfiguration mechanisms
Video processing (*packager*)	VM (CPU, interfaces) monitoring	Local and global reconfiguration of virtualized cache
End-user segments delivery (*server*)	VM (CPU, interfaces) monitoring Applications monitoring (user activity, video quality)	Local and global reconfiguration of virtualized cache

For uncompressed flow distribution, it is necessary to support several 100 Gb/s connections among MCEN and AMEN to support a realistic number of TV channels. Moreover, to guarantee a reliable service, optical connections need to be monitored to detect/anticipate transmission degradation that could eventually disrupt the service.

To dimension video processing, assuming 1 packager is required per each TV channel, each packager is instantiated on a 8-core 4 GB RAM VM equipped with 1 Gb/s interface. Regarding video delivery, one video server can support around 500 users of average quality (assuming video quality adoption scenarios in [20] and also redundancy of HTTP servers) if it is instantiated in a VM with similar characteristics. Note that by monitoring the activity of VMs (CPU usage, interfaces), local and global reconfiguration of leaf cache nodes can be performed with the aim of reducing overall resource utilization (computing and network) while keeping the desired service performance in terms of relevant KPIs.

Based on those initial figures, one can easily evaluate the impact of this particular Use Case in terms of the required capacity at the metro network. Let us then assume a live-TV service providing 35 different channels (according to the scenario in [24]) with a similar demand scenario than that used in Section 4.3.4 for VoD, i.e. 10.000 end-users in the peak. For the sake of simplicity, let us assume that 6 leaf cache nodes are created in 6 different AMEN nodes (each giving service to a proportional number of users in the peak). An 18 Gb/s uncompressed flow per each TV channel to be distributed from the same MCEN entrance point to any AMEN hosting a virtualized leaf cache can be considered as a reasonable assumption. Thus, this service might require (in the hypothetic peak) around 600 Gb/s connectivity (supported by 100 Gb/s connections and interfaces) between AMENs and MCEN and near 50 VMs (including packagers and servers) per AMEN node.

4.5 Service Robotics

4.5.1 Use Case description

The recent and significant development of M/N-EMS (Micro/Nano - Electro Mechanical Systems) and embedded systems has facilitated rapid commercialisation of Service Robotics, i.e. the robotics excluding industrial automation applications [26]. Service Robotics includes terrestrial robots for professional and personal tasks, and drones for professional tasks only.

The evolution of the mobile network to 4G and then 5G and the development of cloud computing infrastructures has resulted in the emergence of cloud robotics, i.e., the new paradigm in robot applications where a consistent part of the robot "intelligence" is located no more on the robot itself but in the cloud. This brings to an increase of the usually limited robot functionalities and computational power with a highly reduced cost.

This Use Case exploits a novel approach towards endowing robots with advanced perception and action capabilities, thus enabling them to carry out useful tasks autonomously in circumstances that were not planned for explicitly at design time. This is possible thanks to the knowledge provided by robots in their normal operations and uploaded in the cloud, which is used to improve the control strategies and programs in the remote-control logic which are made available to all the robots relying on the cloud infrastructure.

The core of the innovation proposed involves the development of a Cloud Robot Infrastructure (CRI). Figure 15 depicts a simplified model for the Service Robotics employing the concept of CRI.

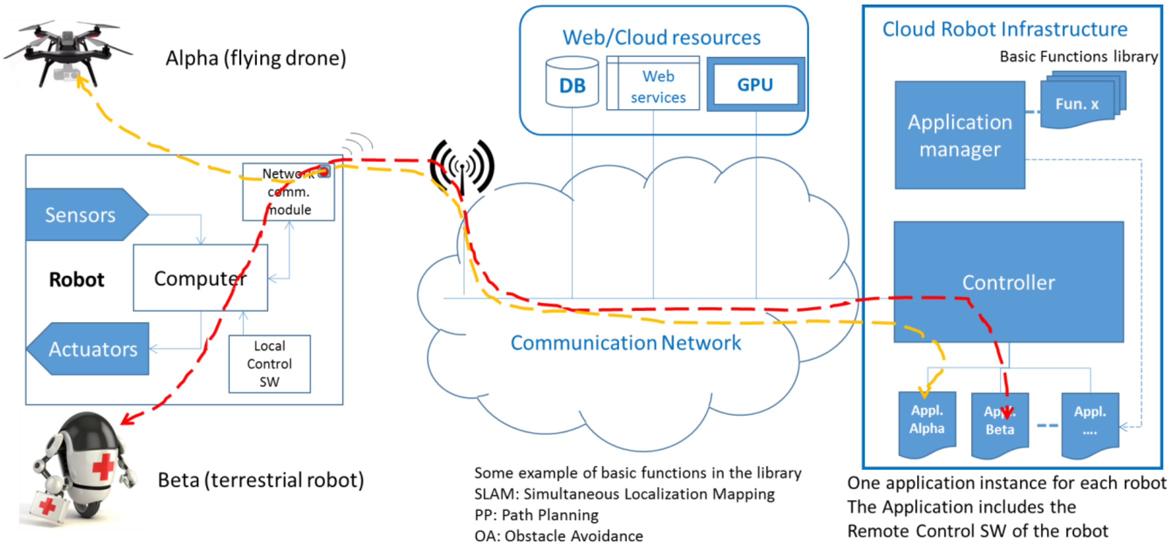


Figure 15 – A simplified model for HW and SW systems, computational tasks and communication in Service Robotics. (DB: Data Base; GPU: Graphics Processing Unit)

On the left of Figure 15 a generic model for a robot which includes: sensors, actuators, a computer with the local control SW and a communication module using a radio interface (5G or other). The generic model is suitable for both terrestrial and aerial (drone) robots.

The CRI, shown on the right side of Figure 15, requires an Application Manager that, for each robot, creates a dedicated application (Alpha, Beta, etc.) which includes the main and heavier part of the control SW of the robot (a smaller light part remains on board). In addition, a main Controller has in charge the coordination of the remote robot applications with the robots on field, and with application's needs that can be met locally on the server (library of basic functions) or in the Cloud (DB, web services, GPU computing services). The CRI allows a robot to maintain on-board in the location where the robots operate, their hardware (sensors, actuators) and the basic control logic and communication software, while off-loading the majority of software tasks to a server in a remote location (a specific data centre or even in the cloud). The advantage is on one side that the robot is lighter in terms of on-board functionalities and energy demand (the last point is especially important), and on the other side that the dedicated control software of the robot can be instantiated by the infrastructure relying on a continuously updated library of basic functions. The Robot Operating System (ROS) is the common SW environment (emerged as a standard de facto) used in the robot application's programming and operations [30]. RoboEarth is a successful project that demonstrated that the model described above can effectively be implemented. It facilitated an experimental environment in which Robots can store and share information, offload computational tasks and collaborate to achieve a common goal [31].

In the middle of Figure 15 it is depicted the communication network composed by a fixed part, for the interconnections of CRI with other parts of the model, and a radio part for the mobile connection of the Robot to its controller. To properly work, this model needs an efficient and reliable communication infrastructure and the mobile radio connections between the robots and the network must have adequate requirements. 5G network, with its suitable performances on latency, reliability and data rate sustainable on mobile connections, is a candidate to cover the needs of communication for Service Robotics.

4.5.2 Components and functionalities and their mapping on Metro-Haul architecture

According to the simplified model shown in Figure 15, the main components of Service Robotics are the following ones.

- Robot (terrestrial or airborne)
  - Computer
  - Local Control SW
  - Communication module
  - Sensors
  - Actuators
- Cloud Robot Infrastructure
  - Application Manager
  - Controller
  - Remote robot control SW instances (dedicated Robot applications)
  - Library of basic functions (SLAM, PP, OA, etc.)
- Web and Cloud resources (DB, web services, GPU)
- Wireless Connectivity

The mapping of the functionalities is shown in Figure 16 where, the servers of CRI are placed, in general, on the AMEN or MCEN closest to the robot to be controlled, due to the stringent latency requirements for the closed loop motion control.

Other web and cloud resources can be placed far from the controller and then they can also be downloaded from data centres reachable on the Core Transport considering that a specific service can be subjected to communication performance requirements. For instance, if a Controller of the CRI placed on an AMEN node requires a GPU service to cope with an additional heavy computation load, such service must be obtained under a maximum delay between the submitted task request and received results. Such delay can be lower than the maximum latency required by motion control loop on the radio access segment but in any case, it involves constraints (e.g. distances) in the fixed part of the network.

The service architectures is only at an experimental stage and there are not yet commercially available solutions, so discussing detailed solutions is currently out of scope for this deliverable. However, it is important to notice that all network components have potentially challenging requirements.

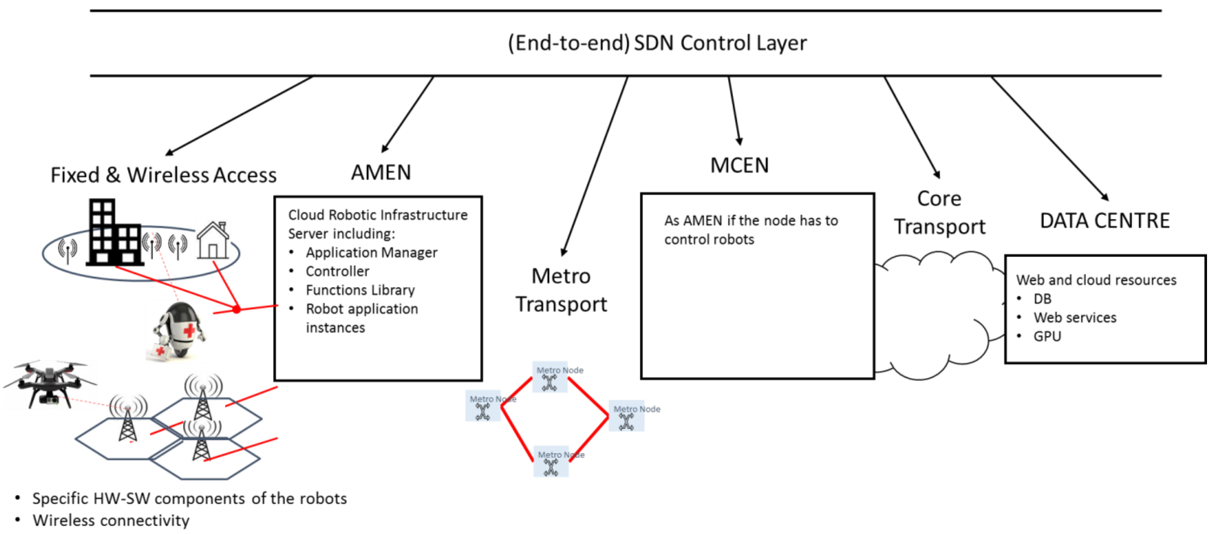


Figure 16 – Components and functionalities required by Service Robotics mapped on Metro-Haul architecture.

4.5.3 Service Requirements

Service requirement for Service Robotics are reported in Table 12 . The most stringent requirement is on the motion control loop that require 1 ms of latency and 10 µs for latency at most. Service robots are very different from each other and these values can be slightly different for different types of robots (for instance tighter value for Jitter could be compulsory or, on the opposite side, loose values of latency could be allowed).

Concerning the data-rates, they strongly depends on the presence of a payload to be carried on a connection between the robot and the control in addition to the control data flow. It is foreseeable that many robots, terrestrial and aerial, will have one or more video cameras mounted to send video in real time to the controller or to a human user. This vide stream images could be panoramic 360-degree HD video, justifying the 100 Mb/s of bandwidth required.

Regarding other parameters it is important to notice that, due to the mission critical characteristics of the robot control, availability and reliability must be high to avoid the loss of control, for an extended period of time, of a potential harmful and dangerous uncontrolled object Concerning area traffic capacity the most likely scenario that can be imagined is the one with a few remote controlled robots located in a given area (in the order of tens or hundreds per km^2).

Table 12 - KPIs of service required by Service Robotics.

1	**Bandwidth (data rate) [Mb/s]**	Low without payload (1 Mb/s) Medium to High with a payload 360° camera (100 Mb/s)

2	**Latency [ms]**	Stringent (1 ms for control)
3	**Jitter [ms]**	Stringent (10 µs for control and possibly for payload)
4	**Reliability [%]**	high (five 9)
5	**Availability [%]**	high (five 9)
6	**Mobility UE speed [km/h]**	5 km/h terrestrial robots 60 Km/h drones
7	**Area Traffic Capacity [Gb/s/km^2]**	Low
8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	Possible
11	**Coverage Requirement [Standard/Extended]**	Everywhere
12	**Energy Consumption critical [Yes/No]**	Yes
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Special purpose [2 logical connections, for control and payload]

4.5.4 Metro-Haul network requirements

According to the Use Case description and the methodology presented in subsection 3.5, the crucial tasks identified for Service Robotics are categorized below:

Task 1: Remote robot control loop and control process. It implies the task for controlling a Service Robot from the CRI and in particular from a process running on an edge node that can be an AMEN or a MCEN. It implies two main requirements: a) communication in the access part between the service robot and the controller with stringent latency, Jitter and availability, b) computation and storage capacities at AMEN or MCEN node for robot control process. A third requirement, the needs of SW updates within the CRI and the bandwidth access to web and cloud services, are made available from task 3 specified below. This task requires a function for monitoring the transmission quality and enables the triggering of an intervention in case of degradation below acceptable levels for critical parameters. In addition a function is necessary to monitoring the status of VMs allocated to the robot control process, along with a function to allow the dynamic reallocation of resources.

Task 2: Payload transmission and processing. When a robot mounts video cameras or other recording devices, data have to be transferred to and processed in the remote controller placed In AMEN or MCEN node. Requirements of this task are a) communication between the service robot and the controller on the access part with high data rate and the support of a mobility at a moderate speed, and b) computation and storage capacities at an AMEN or MCEN node for payload data processing. As tasks 1 above this task also requires transmission degradation detection and VMs resource monitoring and management.

Task 3: Access to Web and Cloud services (DB, GPU etc.). Communication needs across the metro network that the robot control loop and the payload processing tasks have with the web and the cloud are of medium to high data rate (high data rate in case of bulk video streams), medium level of reliability and medium level of latency (the last two only in case of need of requiring GPU services to support the robot controller or the video signal processing). This task requires as Task 1 and 2 transmission degradation detection (applied to the data flows exchange in the metro segment) and in addition a functionality of connectivity reconfiguration in the metro network in case of failures or other reasons (e.g. centralized servers relocation).

The connection of the crucial tasks with service KPIs are reported in Table 13.

Table 13 – Impact of crucial tasks on Service KPIs in Service Robotics Use Case.

Service KPI	Value	Crucial Tasks requiring most stringent KPI value
Data rate	100Mb/s per flow	**Payload transmission and processing** **Access to Web and Cloud services**
Latency	1 ms	**Remote robot control loop and control process**
Jitter	10 µs	**Remote robot control loop and control process**
Availability	Five 9s	**Remote robot control loop and control process**
Mobility (speed)	60 Km/h	**Remote robot control loop and control process** **Payload transmission and processing**
Area traffic capacity		(can be critical only in case of very high density of robots in the area covered by a metro node)
Sensor density		(not relevant)

According to an estimation of 100 robots present in the area covered by a metro node (AMEN or MCEN) the resulting network KPIs are as in Table 14. These requirements could be higher in case of higher density of remotely controlled robots and it depends on the level of penetration of this Use Case. Values one or two orders of magnitude higher could be expected in case of a very high level of service penetration (i.e. 1,000 to 10,000 robots under the control of a single metro node). The variability in throughput (1-100Gb/s) required by the task of Access to Web and Cloud services depends on whether the payload is processed and locally used or if it has to be transferred to other places in the metro or in the core. In the last case, if a high share of robots requiring payload transfers outside the metro node, the throughput is very high (100 Gb/s).

Table 14 – Summary of Metro-Haul requirements (KPIs) by Service Robotics Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Remote robot control loop and control process	100 Mb/s			100 GB		100 vCPU (1 vCPU per robot)	
Payload transmission and processing (e.g. video signals)	10 Gb/s			10 TB		100 vCPU (1 vCPU per Payload flow)	
Access to Web and Cloud services (DB, GPU etc.)	1-100 Gb/s		1-100 Gb/s				

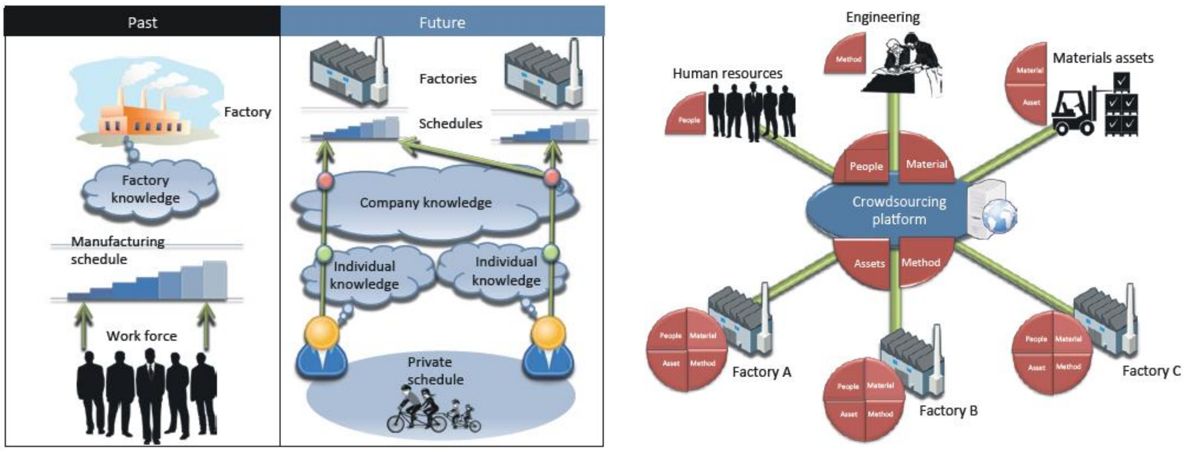
Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Remote robot control loop and control process	Transmission degradation detection/anticipation. VM monitoring	VM dynamic dimensioning (local reconfiguration)

Payload transmission and processing (e.g. video signals)	Transmission degradation detection/anticipation VM monitoring	VM dynamic dimensioning (local reconfiguration)
Access to Web and Cloud services (DB, GPU etc.)	Transmission degradation detection/anticipation	Global connectivity Reconfiguration

4.6 Smart Factories

Due to the fast evolution of the socio-economic scenario at a world scale, and thanks to the opportunities offered by new technologies, the factories are greatly changing the way they are built and operated [32].

In the past the model of manufacturing was very straightforward and simple: starting from a business plan, the manufacturing scheduling could be defined and workforce assembled. Knowledge, production means, and human resources were all located in the factory itself as a physical location. This implied that the relocation of the developed knowledge to other factories was rather difficult, and manufacturing flexibility was very limited. This model is visually depicted on the left part of Figure 17 (a) below [32].



(a) (b)

Figure 17 – Comparison between Past and Future factory model (a), and the concept of Crowdsourcing model (b).

In future, the factory model is expected to change as in right part of Figure 17 (a). Future relations between humans and non-human resources factory will become more flexible using advanced IT, with potential advantages for both production needs and people participating in the production cycle. Also, the sharing of knowledge across platforms will be improved and learning cycles will be reduced due to data storage, semantic technologies and the ability of the worker to merge and analyse the company's experiences with human experience to boost innovation. Smart robotic technologies will be able to contribute to the improvement of ergonomics in production and to support more and more load intensive and routine tasks, freeing workers into high-cognitive activities. Another line of evolution will concern the customer integration, with the implementation of concepts like joint product design, engineering, manufacturing and delivering, which will be customer specific, or customer-driven. The manufacture Industry is quickly moving from a model based on the Factory knowledge, whose schedule is based on workforce allocation, to a model based on the Company knowledge, whose schedule is based on private schedules (and, at the end, on customer's up to date needs).

In Figure 17 (b) [32] a technique known as Crowdsourcing is used, which is claimed as particularly appropriate to implement future manufacturing systems. The term crowdsourcing is a blend of “crowd” and “outsourcing” and describes the process of obtaining ideas, services or content from a large, collaborative group of participants rather than from traditionally specified employees, contractors or suppliers. According to this model, customers or factory operators address orders to a site of a crowdsourcing service (the Crowdsourcing platform in Figure 17 (b), which provides services as engineering supports, temporal human resource employing, purchasing parts or facilities, etc.). In response, a member of the crowdsourcing pool proposes a plan to implement the order, potentially including quotations, and gets it if the plan satisfies the customer or factory operator. With Crowdsourcing the management model radically change: it is no longer a top-down management model, but it becomes rather a model based on collaborative cooperation between parties.

Looking more specifically at the production system, trends in manufacturing are moving towards seamless integration of physical and digital worlds. This enables fast integration, feedback and control loops throughout distributed manufacturing infrastructures. According to the vision reported in [33], the smart manufacturing of the future will be a system in which “stand-alone plants can also communicate with other factory sites, merging vast industrial infrastructures already in place with cloud computing and IoT. The result is a complex but vibrant ecosystem of self-regulating machines and sites, able to customize output, optimally allocate resources and offer a seamless interface between the physical and virtual worlds of construction, assembly and production.”

To promote and foster this transformation toward smart manufacturing, many local initiatives, private or supported by governments, have been launched. Among them, are the Manufacturing Leadership Coalition (SMLC), the Industrial Internet Consortium (IIC) and the advanced Manufacturing Partnership (AMP) in the USA, e-Factory in Japan, Industry 4.0 in Germany, Intelligent Manufacturing in China and Impresa 4.0 in Italy.

4.6.1 Use Case description

The Use Cases related to the smart manufacturing represent the introduction of the Internet and IoT technologies within the production of the manufacturing processes. They rely heavily on the cyber physical systems (CPSs) or smart devices. A CPS can be both the evolution of the current programmable logic controller (PLC) and new type of sensors and actuators, including robots. Automated guided vehicles are assumed to be extensively used for material transfer. Video cameras for quality monitoring and surveillance are also present [34]. The goal of this Use Case is to create a smart factory, modular and in islands, which can be remotely controlled. The Smart Factory shall have a virtual image in the internet cyber world. IoT technologies help the smart devices to communicate and cooperate with each other and with humans in real time.

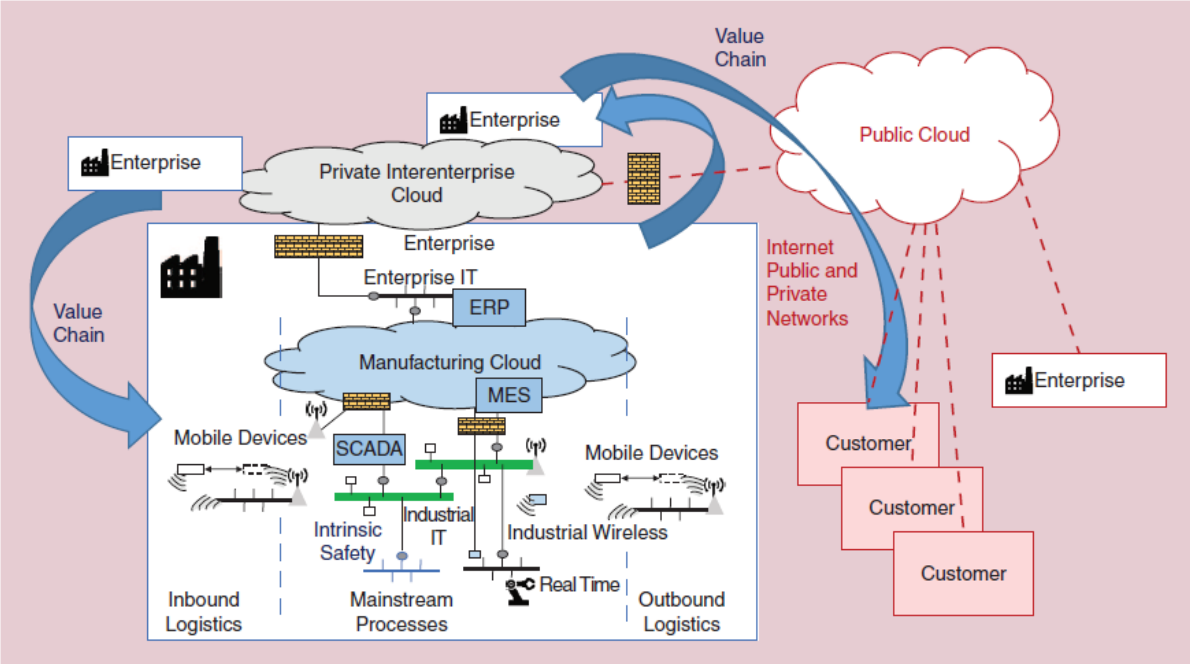


Figure 18 – The complexity of communication in today industrial automation systems. [35]

Figure 18, taken from [35], shows the complexity of communication in industrial automation systems as it is today in the plants of most advanced manufacturing companies. Without going into exhaustive detail on all the aspects discussed in [35], there are multiple application domains with different requirements, e.g., regarding real time, mobility, safety and security, explosion protection, availability, and so forth. It is important to observe that, according to the scenario shown in Figure 18, in today's factories the communication resources are confined to private networks and clouds and protected by firewalls at many levels.

Due to confidentiality, security and industrial intellectual property protection, the challenge in the 5G era is making it possible and secure to provide communication, computational and storage resources outside the factory and enterprise borders. This will open an opportunity to network providers to become suppliers for a set of services to a manufacture enterprise while enlighten the enterprise to have its own dedicated ICT infrastructure to build, operate and update.

Regarding the impact on communications, and in particular the role of the 5G network, the model is articulated into two levels:

- Intra-factory: it involves the communication within the systems and machineries inside the plant. From the 5G communication point of view it involves local data exchanges and all the needs should remain confined within the reference local edge node. (AMEN or MCEM, if it plays the role of an edge node). The radio access segment and related requirements (latency in particular, but also bandwidth and handling a large number of connections required by a high number of sensors) is fundamental for intra-factory communication. 5G is the candidate to replace a costly and rigid cabling system and it competes with other local private wireless network (for communication) and IT (for the cloud part) solutions.
- Inter-factory: it concerns the communication between different plants and between a plant and its headquarters, which coordinates the whole production system (implementing for instance the model of crowdsourcing platform, if the related business model is applied). This communication in general involves both the communication in the metro segment (involving

AMEN and MCEN) and outside the metro segment by the use of the core-backbone network (crossing the MCEN).

In this context we assume that all the issues regarding security, confidentiality and intellectual property protection have been solved by appropriate technical and legal means, and the enterprise is open to purchase communication services from a network operator for both intra-and inter-factory communication, computational and data storage necessities.

4.6.2 Components and functionalities and their mapping on Metro-Haul architecture

To imagine and represent a smart factory in its whole variety and complexity is out of scope for this document. This Use Case considers a limited number of components that could be used as part of a factory and defines for them the requirements in terms of communication and other needs. In this sense, the smart factory Use Case can be more appropriately defined as a non-exhaustive collection of a set of sub-Use Cases. Each sub-Use Case has its own service requirements in terms of KPIs expressed for a single service instance (e.g. latency for connecting a robot with its controller) and all together (e.g. the area traffic capacity required by the complete set of robots of the factory). For the 5G network, the aggregated demand to be carried on and the set of requirements to be fulfilled come from the complete set of heterogeneous types of services.

Components of a Smart factory that are taken into consideration for the Intra-factory model are:

1. Robots
2. PLC lines (or their evolution)
3. Sensors for process monitoring
4. Video cameras (ultra-high definition) for quality check
5. Video cameras (high definition) for plant surveillance
6. Automated guided vehicles for material transportation
7. Factory Application Platforms

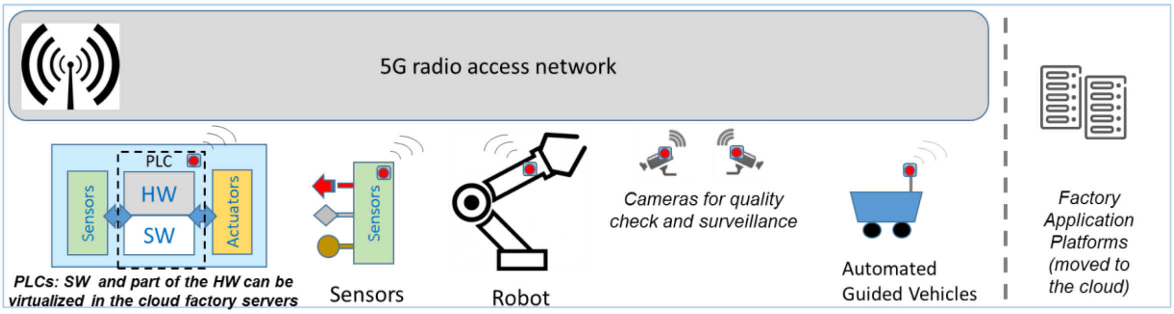


Figure 19 – Components of the model considered in the Smart factory Use Case.

The components of the model shown in left-centre part of Figure 19 are assumed to rely on wireless connections provided by the 5G radio access network. Factory Application platforms that today take part of the private enterprise cloud (see Figure 18) are depicted in right part of Figure 19: they are assumed to be moved to the network provider premises, in locations where service constraints allow theirs installation (typically at the edge, corresponding in Metro-Haul architecture to an AMEN node).

Video cameras for quality check are a means to remote control the manufacturing quality. This can be done by the processing and analysis of the images collected with the aim of evaluating the conformity of surfaces and volumes of machined parts with the standard defined in the design of the same part. This is done the purpose of monitoring the process and raise warnings or alarms which allow to tune process parameters or conduct maintenance actions. Video cameras for plant surveillance are used for remote supervision of the plant by human operators and, possibly, by some automatic mechanisms involved in emergency actions (for instance in case of a failure of the mechanical parts of a robot).

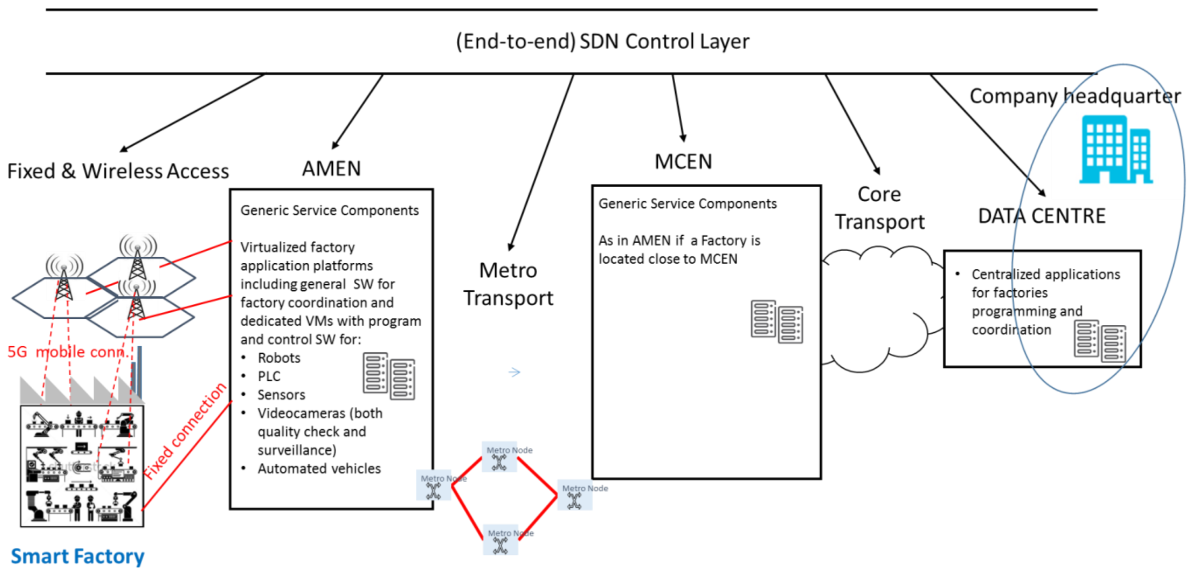


Figure 20 – Components of the model considered in the Smart factory Use Case.

Figure 20 shows a mapping of the Use Case components on the Metro-Haul architecture. Assuming that the intra-factory communications are provided by the 5G network, 5G radio/mobile connections are assured by the radio cells close to the factory that have to be appropriately dimensioned for the traffic load generated by radio mobile connections. An AMEN node (or MCEN, if it acts as edge node) is supposed to host the virtualized factory platforms including SW for factory coordination and control SW for all the pieces of machinery to be controlled (robots, PLC, sensors, video cameras and automated guided vehicles).

For additional needs of communication not requiring mobile access (for instance for interconnecting the factory LAN with the Internet), a fixed connection between the factory and the closer AMEN node is also present.

4.6.3 Service requirements

Table 15 reports the service KPIs for the six identified smart factory components. Values of parameters come after elaborating and mediating among many sources.

A typical industrial closed-loop motion control application is based on individual control events. Each closed-loop control event consists of a downlink transaction followed by a synchronous uplink transaction, both of which are executed within a cycle time. Control events within a manufacturing unit may have to occur isochronously. Factory automation considers application-layer transaction cycles between controller devices and sensor/actuator devices. Each transaction cycle consists of (1) a command sent by the controller to the sensor/actuator (downlink), (2) application-layer processing on the sensor/actuator device, and (3) a subsequent response by the sensor/actuator to the controller (uplink). Cycle time includes the entire transaction from the transmission of a command by the controller to the reception of a response by the controller. It includes all lower layer processes and latencies on the air interface as well the application-layer processing time on the sensor/actuator [7].

Latency jitter availability and reliability values for Robots and PLC lines are taken from Table 7.2.2-1 of [36]. For robots the values are the ones defined for Motion control, while for PLC lines the values are the one for Discrete automation in the Table 7.2.2-1 of [36]. More stringent values, such as the lowest one reported in Table 3 of [34] (for instance a latency of tens of microseconds and Jitter of tens of nanoseconds) are also possible, but they are too challenging to be achieved with the support of 5G network and therefore not considered in the requirements listed in Table 15. In those cases a dedicated infrastructure relying on another technology for communication (i.e. not 5G) must be used.

To evaluate the bandwidth, communication required to control a single or set of robots, we ascertained that a closed loop control signals for 100 axles requires less than 100 Mb/s [36] (which is the capacity of the Ethernet bus used to transport the control signals). As a robot has less than 10 axels (normally 4 to 6) we can infer that the control of a robot requires at most 10 Mb/s, but probably reasonably much less than these values, as the Ethernet bus is not used at its full capacity. Therefore, we assigned 1 Mb/s as the target value of bandwidth required by a robot.

Concerning availability and reliability, they are quantified with six or five 9 requirements, except sensors and video quality check devices which tolerate a lower level (quantified in four 9) of resiliency as they do not have to handle stringent closed loop real time tasks.

Video surveillance does not require a very high resolution streams and may typically be 5 Mb/s (compressed HD 1080p) per single camera video signal flow is considered enough for the purpose. Latency requirement on Video surveillance (10ms) is required, in response to prompt emergency actions triggered by this source. Differently, Quality check require a very high resolution for detecting imperfections on parts under monitoring, which justifies and a 100 Mb/s (compressed UHD 8k) requirement per single flow. Latency is not an issue for this component.

The range of values for the area traffic ((Gb/s)/km^2) reported in Table 15 depends on the density of machinery in factories in each specific context, and are taken from the dimensioning exercise reported in Appendix A1. We assumed a low density for an automotive factory as the one of example reported, and a value two order of magnitude higher for a factory with a very high density of small sized automation equipment.

Table 15 – KPIs of service required by Smart factory service components.

	KPI	Robot	PLC line	Sensor	Video Surveillance	Video Quality Check	Automated Vehicle
0	5G Service Category	URLLC	URLLC	mIoT	eMBB	eMBB	URLLC
1	Bandwidth (data rate) [Mb/s]	1 Mb/s	100 Kb/s	10 Kb/s	5 Mb/s	100 Mb/s	1 Mb/s
2	Latency [ms]	1 ms	10 ms	100 ms	10 ms	-	50 ms
3	Jitter [ms]	10 µs	100 µs	-	-	-	100 µs
4	Reliability [%]	Very High (6 nine)	High (5 nine)	Medium (4 nine)	High (5 nine)	Medium (4 nine)	High (5 nine)
5	Availability [%]	Very High (6 nine)	High (5 nine)	Medium (4 nine)	High (5 nine)	Medium (4 nine)	High (5 nine)
6	Mobility - UE speed [km/h]	0 km/h	0 Km/h	0 Km/h	0 Km/h	0 Km/h	20 km/h
7	Area Traffic Capacity [Gb/s/km^2] (typical)	0.2-20 (Gb/s)/ km^2	0.01-1 (Gb/s)/ km^2	0.02-2 (Gb/s)/ km^2	0.1-10 (Gb/s)/ km^2	4-400 Gb/s per km^2	0.05-5 (Gb/s)/ km^2
8	Rely on Sensor Network [Yes/No]	NO	NO	YES	NO	NO	NO
9	Massive Type [Yes/No]	NO	NO	NO	NO	NO	NO
10	Device Direct [Yes/No]	YES	YES	NO	NO	NO	YES
11	Coverage Requirement [Standard/Extended]	Industrial Plant	Industrial Plant	Industrial Plant	Industrial Plant	Industrial Plant	Industrial Plant
12	Energy Consumption critical [Yes/No]	No	No	NO	NO	NO	NO
13	Type of User Equipment (UE) [Conventional/Special purpose]	Special Purpose	Special Purpose	Special Purpose	Special Purpose	Special Purpose	Special Purpose

4.6.4 Metro-Haul network requirements

Intra-Factory level requirements

The more challenging network requirements are for the Intra factory connections and this involves the segment between the factory site and the AMEN node where the real-time control SW for the factory machinery are placed.

As a general requirement the radio access network (BBU loop, see Figure 21) must guarantee a maximum delay of 1ms with a jitter of 1µs. The connections and bandwidth required by a factory of medium size is estimated in about 2000 simultaneous connections and 4 Gb/s of traffic (see dimensioning exercise of Appendix A1). Factory platforms could be back-upped on (or distributed among) two or many AMEN nodes in case reliability and/or availability requirements could not be fulfilled with a single AMEN node.

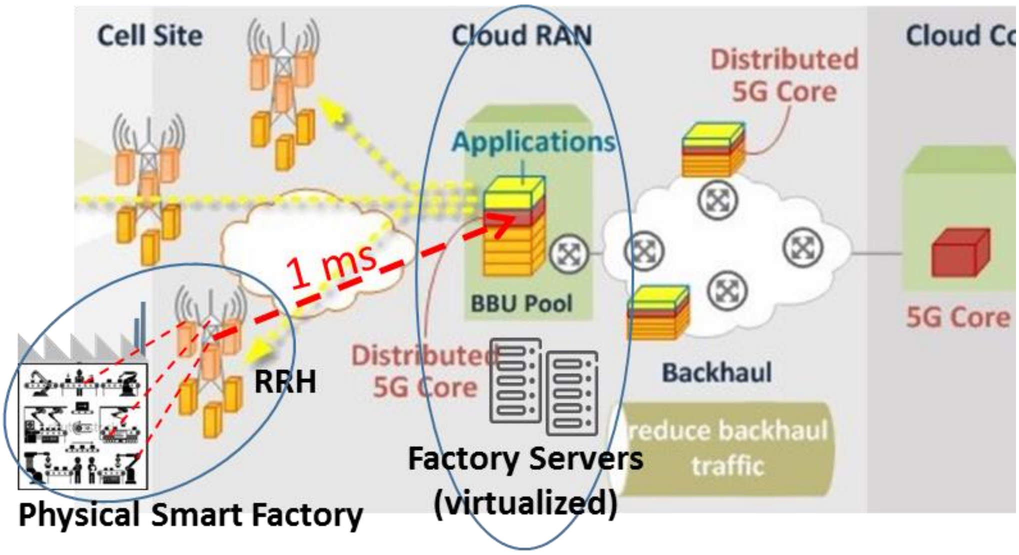


Figure 21 – BBU loop between factory site where wireless connected machinery work, and factory servers located in the node hosting the BBU pool (LTE cloud RAN model is assumed, but concept can be extended to other architecture and splitting options). In metro Haul BBU pool and factory servers are located in AMEN (or MCEN) node.

A high amount of processing and storage capacity is required on the AMEN and MCEN to host the edge server capabilities for machinery control and coordination. For a medium size single factory (Appendix A1) the requirements are estimated at more than 2000 simultaneous connections, 4 Gb/s of aggregated bandwidth, 40 TB of storage memory and 700 vCPU of computing capacity. In case of a greater smart factory or multiple smart factories served by a single AMEN node, these requirements increase accordingly.

Inter-Factory level requirements

For communications between the Company headquarter (that in general relies on a private data centre or cloud for its computational and storage facilities) and the factories, high bandwidth and secure connections are required. Central Company servers elaborate the production planning and communicate their production programs to the factories. On the other end, the central server receives feedbacks from factories about the status of production in real-time, updates the production programs accordingly, and coordinate the orders to suppliers and deliveries to customers. In addition SW release for machinery control and management are also uploaded on the factory servers by centralised Company servers or from places where engineering and SW developing centres are located. It is not easy to predict the amount of the required bandwidth for connections between the company centralized servers (connected to the metro or to the core) and factory servers (located in the AMEN) but it could be estimated in the range of hundreds of Mb/s to tens of Gb/s.

According to the Use Case description and the methodology presented in subsection 3.5, the crucial tasks identified for all the Smart Factory sub-Use Cases are the following five.

Task 1: Motion Control (Robot, PLC, vehicle). It implies the task of remote controlling all the machinery within the factory from processes running on an edge node that can be an AMEN or a MCEN. Different pieces of machinery have different service requirements on KPIs: the most stringent ones are considered to characterize this crucial task (for instance latency for controlling a robot is less than the latency required for controlling a PLC). This critical task implies two main requirements: a) communication on the access part between the piece of machinery and the controller with stringent (PLC and vehicle) or very stringent (robot) latency, Jitter and availability and a medium level of requirement on mobility (vehicle only), b) computation and storage capacities at AMEN or MCEN node for control processes. A third requirement, the needs of SW updates from SW development centres and Systems providers is made available from task 5 specified below. This task require a function for the monitoring of transmission quality and enabling the triggering of an intervention in case of degradation under acceptable levels of critical parameters. In addition, a function is necessary to monitoring the status of VMs allocated to the control processes, along with a function to possibly allow the dynamic reallocation of resources. Due to the very high availability requirement a very fast recovery mechanism for both the communication and the process itself is also necessary. Solutions as back-up images of the process on another AMEN node and a redundant radio coverage of the plant site could be a solution.

Task 2: HD Video Surveillance, Communication and Processing. Requirements of this task are a) communication on the access part between the cameras and the image processing made at the metro node with medium level of requirement for both latency and data rate, and b) computation and storage capacities at AMEN or MCEN node for image data processing. This task requires transmission degradation detection and VMs resource monitoring and management. A very fast recovery mechanism for both the communication and the process is also necessary.

Task 3: UHD Video Quality Check Communication and Processing. Requirements of this task are a) communication on the access part between the cameras and the image processing made at the metro node with high level of requirement for data rate, and b) computation and storage capacities at AMEN or MCEN node for image data processing. This task requires transmission degradation detection and VMs resource monitoring and management. To avoid the loss of data for a too extended period of time, a fast restoration mechanism for communication is necessary.

Task 4: Sensor's Connections and Data Processing. Requirements of this task are a) high connection number (a sensor density of thousands of sensors per km^2 can be achieved) on the access part between the sensors and the data processing made at the metro node, and b) computation and storage capacities at AMEN or MCEN node for sensor's data processing. This task requires detection of out-of-service transmissions, VMs resource monitoring and management and fast restoration mechanisms in case of failures.

Task 5: Factory Coordination Processes. This crucial task coordinates all the processes of the factory site located in the AMEN node and it communicates with centralised Company servers or with places where engineering and SW developing centres are located. Main requirements of this task are a) communication at a high data rate across the network between AMEN node and other centralized locations b) computation and storage capacities on the dedicated VM at AMEN or MCEN node. Other requirements are a fast restoration mechanism for both communication and process itself, and also a VM monitoring and management including the possibility of dynamic allocation of VM resources for all the processes coordinated (i.e. tasks 1 to 4 above reported).

The connection of the crucial tasks with service KPIs are the ones reported in Table 16.

Table 16 – Impact of crucial tasks on Service KPIs in Smart factory Use Case.

Service KPI	Value	Crucial Tasks requiring most stringent KPI value
Data rate	100Mb/s per flow	**Payload transmission and processing** **Access to Web and Cloud services**
Latency	1 ms	**Remote robot control loop and control process**
Jitter	10 µs	**Remote robot control loop and control process**
Availability	Five 9s	**Remote robot control loop and control process**
Mobility (speed)	60 Km/h	**Remote robot control loop and control process** **Payload transmission and processing**
Area traffic capacity		(can be critical only in case of very high density of machinery or video cameras in the area covered by a metro node)
Sensor density		(not relevant)

According to an estimation of 100 robots present in the area covered by a metro node (AMEN or MCEN) and the dimensioning calculation reported in Appendix 1, the resulting network KPIs are as in Table 14. The requirements on network KPIs could be higher in case of higher density of remotely controlled robots and it depends on the level of penetration of this Use Case. Values one or two orders of magnitude higher could be expected in case of a very high level of service penetration (i. e. 1,000 to 10,000 robots under the control of a single metro node).

Table 17 – Summary of Metro-Haul requirements (KPIs) for the smart factory Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Motion control (all machinery)	0.2 Gb/s		nx10/100 Gb/s (shared)	230 GB		550 vCPU	
Sensor's connections and data processing	0.02 Gb/s		nx10/100 Gb/s (shared)	20 GB		20 vCPU	
HD Video surveillance Comms. and Processing	0.1 Gb/s		nx10/100 Gb/s (shared)	2 TB		20 vCPU	
UHD Video Quality Check Comm. and Processing	4 Gb/s		nx10/100 Gb/s (shared)	40 TB		80 vCPU	

Factory coordination processes	Few to tens Gb/s		10/100 Gb/s	1 TB		50 vCPU ≈7% of other tasks	

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Motion control (all machinery)	Transmission degradation detection/anticipation VM monitoring	Very Fast recovery mechanisms
Sensor's connections and data processing	Transmission out of service detection VM monitoring	Fast restoration mechanisms
HD Video surveillance Comm and Processing	Transmission degradation detection/anticipation VM monitoring	Very Fast recovery mechanisms
UHD Video Quality Check Comm. and Processing	Transmission out of service detection VM monitoring	Fast restoration mechanisms
Factory coordination processes	Transmission out of service detection VM monitoring	Fast restoration mechanisms Global Reconfiguration VM dynamic dimensioning (local reconfiguration

4.7 6DoF Virtual Reality immersive experience

Next-generation VR (Virtual Reality), AR (Augmented Reality) and MR (Mixed Reality) experiences – term XR (Extended Reality) encompasses all those [39]- will have 6 degree of freedom (6DoF) for the next level of immersion, allowing users to move within and intuitively interact with the environment. 6DoF content is an order of magnitude richer in naturalness and interactivity than 3 degrees of freedom (3DoF) video. Current 3DoF experiences, such as 360° video, allow the user to rotationally look around from a fixed position. 6DoF experiences, which are available in video games today, allow the user to move spatially through the environment just by walking or leaning the head forward. 6DoF head motion tracking is required to enjoy 6DoF content in an intuitive manner [40].

4.7.1 Use Case Description

Synthetic (computer-generated) 6DoF environments can be rendered in the cloud today, while 6DoF video, sometimes referred to as "point cloud video," is in its early stages. Sports, tourism, education, and other forms of immersive video will flourish as 6DoF technologies evolve.

This Use Case is akin to streaming a video game, so end-to-end latencies must be on the order of a few tens of milliseconds. A trade-off between latency and bandwidth exists for 6DoF content delivery.

Current mobile VR head-mounted display (HMD) systems support a 3DoF architecture, where the HMD detects the user's rotational movement, providing the main benefit of looking at the world from a fixed point. 6DoF systems that delivers a more immersive VR experience, where user motion is not confined to rotations around a single viewpoint are still under development and 6DoF devices are in a pre-commercial stage. This will enable highly accurate 6DoF motion tracking of head movements on the mobile platform. [41]

The following text, get from [40], explain well what is the main reason of latency requirements in VR applications: "*head movement while wearing an HMD with head tracking is another specific interaction with distinct latency requirements that warrants more discussion. Motion-to-photon (MTP) latency is the time between an action (a head movement) and reaction (the display is updated* *based on the movement). When a user moves their head, the brain expects an instantaneous visual and aural update, so delaying that even minutely can be problematic. A MTP latency below 20 ms is currently targeted for many VR user experiences, but studies have shown that achieving a MTP latency of less than 15 ms makes the delay imperceptible to nearly all users. Considering the strict MTP requirements associated with HMD usage, ABI Research expects that AR and VR applications will keep MTP processing on the device (the HMD itself). There are hybrid-computing scenarios where 3D rendering happens off the device. PCs and game consoles tethered to VR HMDs are available today, while network edge or cloud rendering may be possible in the future. For these hybrid scenarios, ABI research expects that the processing will be intelligently split between cloud-based 3D rendering, which requires low network latency, and additional processing on the untethered mobile device to ensure high quality visuals at a fixed low MTP latency.*"

Among all the possible 6DoF XR services the selected one for this Use Case regards an immersive experience that allows the user to virtually move within a live event (for instance within a stadium during a football match) by means of VR HMD 5G mobile connected."

4.7.2 Components and functionalities and their mapping on Metro-Haul architecture

The identified service implies the following components and data processing tasks:

- VR head-mounted display 5G mobile connected worn by the user. Bandwidth and latency of data streaming received from the network and concerning the evolving immersive environment are strict requirements (200Mb/s and 5 ms or 1Gb/s and 20ms). [40]
- VIDEO collection and processing at the location of the event: many VHD video cameras film the event and provide video streams from different shooting points. A server (Event located server) combines the video camera's streams and produce the VR immersive video streaming flow to be transmitted in quasi real-time to the edge locations where an Edge server allow the users can enjoy the immersive experience. Location of the event can be far or very far from the user. Image processing on the server and distance from event location to edge site can cause important delays (hundreds of milliseconds or even seconds) and a consequent shift in time between the event and the experience. The data rate flow of the whole immersive video stream is estimated in the order of few to tens Gb/s, and so a connection from 10 to 100 Gb/s is necessary from the Event located server to the Edge server.
- VR immersive video streaming flow on network EDGE: a local Real-time Video Content Server located close enough to the user (for latency constraint fulfilment) acquires the VR immersive video streaming flow of the live event from the event location server and make it available to the users. A dedicated stream of 100Mb/s - 1Gb/s [40] is provided to each individual user taking into account the HMD movements.

The mapping of the components and the required processes on the metro Haul architecture is shown in Figure 22.

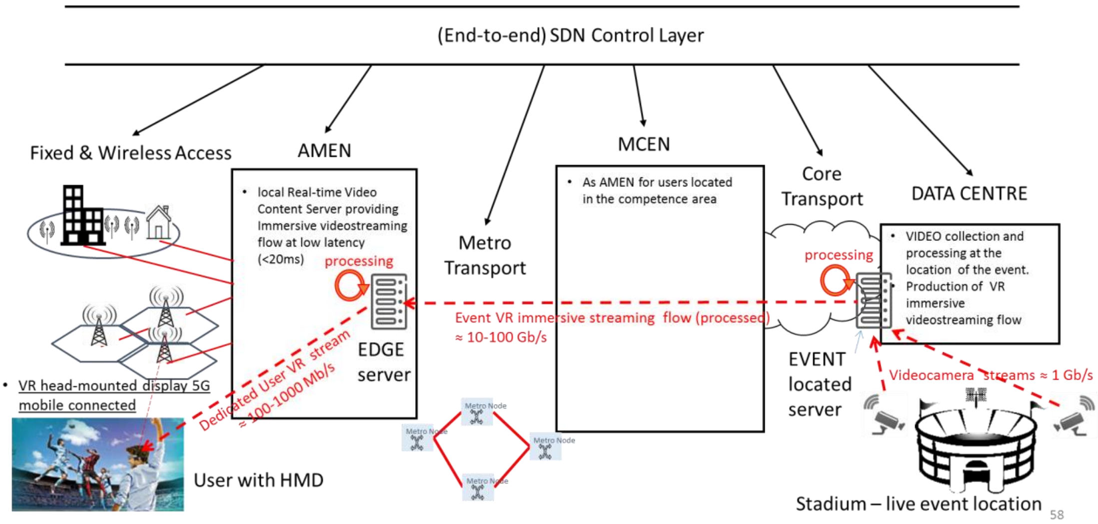


Figure 22 – Components of the model considered in the 6DoF VR immersive experience Use Case.

4.7.3 Service requirements

The service envisioned in this Use Case has both the requirements of categories Ultra Reliable Low Latency Connections (URLLC) for stringent requirements on latency (less than 20 ms) and enhanced Mobile Broad Band (eMBB) for requirement on data rate (up to 1 Gb/s) [40]. On the other hand, the requirements on availability are not required to have high values because service continuity for safety reasons is not an issue.

It should be emphasized that trade-off between latency and bandwidth exists for 6DoF content delivery. With lower end-to-end latency for communicating the user's movement, fewer viewpoints around the user need to be sent. For example, when end-to-end latency is on the order of 1 ms to 5 ms, then the bit-rate for 6DoF content can be reduced to ~100 Mbps to 200 Mbps. However, if the end-to-end latency of the system is 5 ms to 20 ms, then the bit-rate is closer to ~400 Mbps to 600 Mbps, because more viewpoints around the user need to be sent. The result, 6DoF content will most likely be limited to edge streaming scenarios (i.e. the achievable latencies) during the 5G timeframe.

Table 18 - KPIs of service required by 6DoF VR immersive experience.

1	***Bandwidth (data rate) [Mb/s]***	*Very High (100 Mb/s to 1 Gb/s, depend on Latency)*
2	***Latency [ms]***	*Stringent (1 ms to max 20 ms)*
3	***Jitter [ms]***	*Unknown (but expected low)*
4	***Reliability [%]***	*medium (four 9)*
5	***Availability [%]***	*medium (four 9)*
6	***Mobility - UE speed [km/h]***	*10 km/s*
7	***Area Traffic Capacity [Gb/s/km^2]***	*High*
8	***Rely on Sensor Network [Yes/No]***	*No*
9	***Massive Type [Yes/No]***	*No*
10	***Device Direct [Yes/No]***	*No*
11	***Coverage Requirement [Standard/Extended]***	*Hotspot*
12	***Energy Consumption critical [Yes/No]***	*No*

13	***Type of User Equipment (UE) [Conventional/Special purpose]***	*Comm. devices on headset (HMD)*

4.7.4 Metro-Haul network requirement

The service architecture of this Use Case cannot be accurately described because many technical details and solutions are not yet readily available, and its expected they will developed over the coming years. A very generic scheme of service architecture put in relation with the network hierarchy is reported Figure 23, taken from [40].

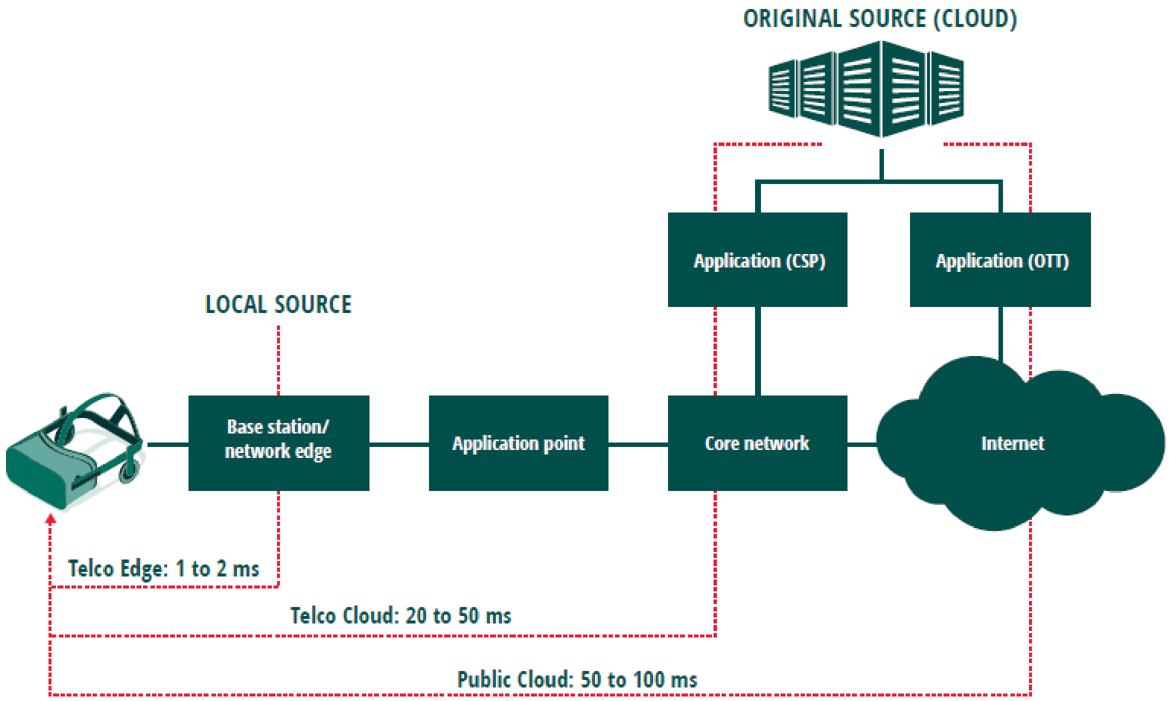


Figure 23 – Original source and Local source of VR contents and their relation with network tiers.

What might be imagined as impact of this Use Case in terms of resources in relation to the Metro-Haul architecture is the following.

EDGE (metro node), which hosts Local source of 6DoF VR interactive streams, will require high processing, storage and data rate capacities for the local video server (called local source in Figure 23) that have to receive the 6DoF VR immersive live streaming from the event location (≈2-10 Gb/s) and make available the dedicated streaming instances to the users (N x 100/600 Mb/s, on average, where N is the number of simultaneous users served by the edge node, for N=100 -> 10/60 Gb/s).

EVENT LOCATION (could be metro node or a separate data centre) will necessitates of very high processing capacity of the server (called original source in Figure 23) which receives the video streams from the shooting cameras (M x 1 Gb/s), processes the data and transmits the VR immersive streaming (≈ 2-10 Gb/s) to the edge nodes by a multicast scheme.

According to the Use Case description and the methodology presented in subsection 3.5, the crucial tasks identified for the 6DoF VR Use Case are the following three.

Task 1: Local Source Processing and Interactive VR Stream Delivery. It implies the task of receiving the content from the original source and making it available to the users by handling each individual interactive steam. This critical task implies two main requirements: a) communication on the access part between the HMD and the local server with stringent or very stringent latency b) computation and storage capacities at AMEN or MCEN node for 6DoF image processing.

Task 2: Original Source Processing. It involves the heavy processing of video streams from the live event and their 6DoF rendering in the format to be interactively used by the final user. This process can be in the core or in a metro node (AMEN or MCEN) which generally is not the same where the Local source process operates. Requirement of this process are a) very high data rate connections due to the size of video streams to be received, handled and transmitted, b) VM resources in terms of storage and computing, c) VM monitoring and dynamic resource reconfiguration.

Task 3: High Data Rate and Reliable Communication between Original and Local sources. Requirements of this task are a) very high data rate multi cast connections between Original and Local sources (due to the typically large size of streams for 6DoF rendering), b) Transmission degradation detection/anticipation, c) very fast recovery mechanism in case of failure and d) global reconfiguration of multicast connections. Latency and Jitter are probably other requirements for this task, but even though they are unknown, they should not be particularly critical because they can be handled with buffering and with a fixed delay of seconds which allows to not compromise the quality of the perceived service.

The connection of the crucial tasks with service KPIs are the ones reported in Table 19.

Table 19 – Impact of crucial tasks on Service KPIs in 6DoF VR experience Use Case.

Service KPI	Value	Crucial Tasks requiring most stringent KPI value
Data rate	Up to 1Gb/s per flow	**Local Source processing and interactive VR stream delivering** **Original Source processing** **High Data Rate and reliable communication between Original and Local source**.
Latency	1-20 ms	**Local Source processing and interactive VR stream delivering.**
Jitter	unknown	
Availability	Four 9s	**Local Source processing and interactive VR stream delivering** **Original Source processing** **High Data Rate and reliable communication between Original and Local source**.
Mobility (speed)	<10 Km/h	(possibility of users travelling at high speed (on car, train etc.) not considered)
Area traffic capacity		(can be critical in the access part in case of high density of users)
Sensor density		(not relevant)

The resulting network KPIs for the Use Case are as in Table 20. Numbers are given for 100 users simultaneously served by a metro node (500 Mb/s data rate, 0.2 vCPU and 500 GB of memory per active user). Service is assumed in an early phase of its availability on the market. Original Source processing is assumed to be located in a MCEN node (but it could be outside in the core) and 100 video cameras with 1 Gb/s uncompressed flow are assumed to be shooting the event. 50 vCPU and 100 TB of storage memory is required to process such original video streams to produce the 6DoF rendering live event multicast stream- The delivery live event stream from the Original source to the Local source requires a bandwidth in the range between 10 and 100 Gb/s.

Table 20 – Summary of Metro-Haul requirements (KPIs) by 6DoF VR experience Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Local Source processing and interactive VR stream delivering	60-150 Gb/s			10 TB		20 CPU	
Original Source processing		100 Gb/s	100 Gb/s		100 TB		50 vCPU
High Data Rate and reliable communication between Original and Local source	10-100 Gb/s		10-100 Gb/s				

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Local Source processing and interactive VR stream delivering	Transmission degradation detection/anticipation. VM monitoring	Fast restoration mechanism VM dynamic dimensioning (local reconfiguration)
Original Source processing	VM monitoring	VM dynamic dimensioning (local reconfiguration)
High Data Rate and reliable communication between Original and Local source	Transmission degradation detection/anticipation	Fast restoration mechanism Global connectivity Reconfiguration

4.8 Intelligent Transport System and Autonomous Driving

4.8.1 Use Case description

This section discusses the requirements for the development of an intelligent transport system including autonomous driving, advanced driver assistance systems (ADAS) and intelligent cooperative traffic technologies by Metro-Haul reporting on an approach starting from the service requirement projections with some discussions on the potential impact on the Metro-Haul network architecture and systems. Focused on the end-user case, this subsection will provide the definition of intelligent environments for automotive scenarios, including the service requirements

As of 2016, Autonomous driving capabilities are currently less than 1% of vehicles sold (considering vehicles with basic to partial autonomous capabilities). Currently, over 80% of top original equipment manufacturers (OEMs) have announced plans for highly autonomous vehicles to be on the road by 2025 [27]. In terms of shared mobility, there is a strong projected growth increase of 80% in the number of customers expected to be using shared mobility infrastructure, once self-driving taxis become available. In the domain of connectivity, it is expected that 100% of all cars manufactured especially in the premium segment will have a range of embedded connectivity solution to enable in-vehicle communication but also connectivity to the surrounding infrastructure.

Hence it is clear that the automotive industry is undergoing key transformations as more vehicles become connected. Therefore, it is foreseen that a strong drive for connectivity is in mission critical applications. According to the world health organisation, there were about 1.25 Million road traffic fatalities worldwide in 2013. In addition, 20-50 million injuries or disabilities were caused from road traffic incidents. These have a detrimental impact on the global economy resulting in approximately 1 – 5% economic impact on global GDP.

This has been duly recognised by the industry and has resulted in the forming of a number of organisations such as 5GAA [29], and within IEEE focused on Mission critical 5G for Vehicle IOT, as well as other major stakeholders coordinating technological development between the Automotive sector and the Telecommunications industry. The aim is now to co-develop the vehicle platforms with the required hardware and software solution in combination with upcoming connectivity and networking solutions including the end-to-end solutions for intelligent transportation. Many of the vehicles being developed will indeed incorporate 5G technology as the next generation of mobile technology to support reduced latency, increased reliability and high throughput vehicle to everything messaging in a highly dense connectivity environment which the road presents.

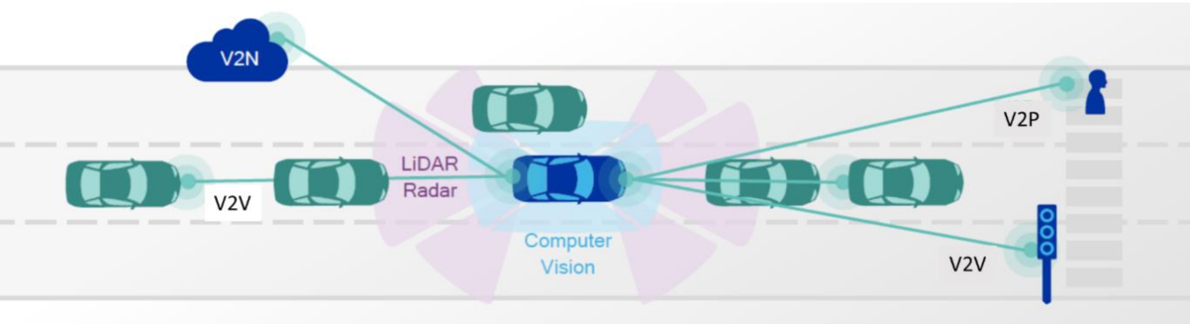


Figure 24 – V2X supporting Vehicular communications as a non-line of sight sensor for moving objects.

Vehicles are being improved to exchange information with other vehicles (V2V) to avoid collision, with the road-side infrastructure (V2I) for traffic signal timing and priority. In addition, vehicles are being rolled out with vehicle to internet servers (V2N) for real-time traffic routing or for Fleet management services. Vehicle to pedestrian (V2P) information exchange also enables safety alerts to pedestrians and cyclists. In this section, we focus on V2X as a service delivery platform for the above applications and the requirements and KPIs for V2X Use Cases which will form the basis for the Metro-Haul 5G Use Case definition. Figure 24 shows an example of V2X communication.

V2X services can be divided into 2 main cases, Firstly, Direct Communications (so called PC5 Interface) where building upon a 5G device to device design with enhancements for high-speed, high density communication with improved synchronisation between terminals/cars with low latency is the main objective. Here, communication over 100s of meters can be supported with periodic broadcast messaging specifically tailored for latency sensitive Use Cases e.g. for V2V safety. The second case is a network wide communication; here using 5G broadcast to send messages from a V2X server to vehicles can enable messages to be event based, unicast type messages. These are more latency tolerant Use Case mainly for situational awareness.

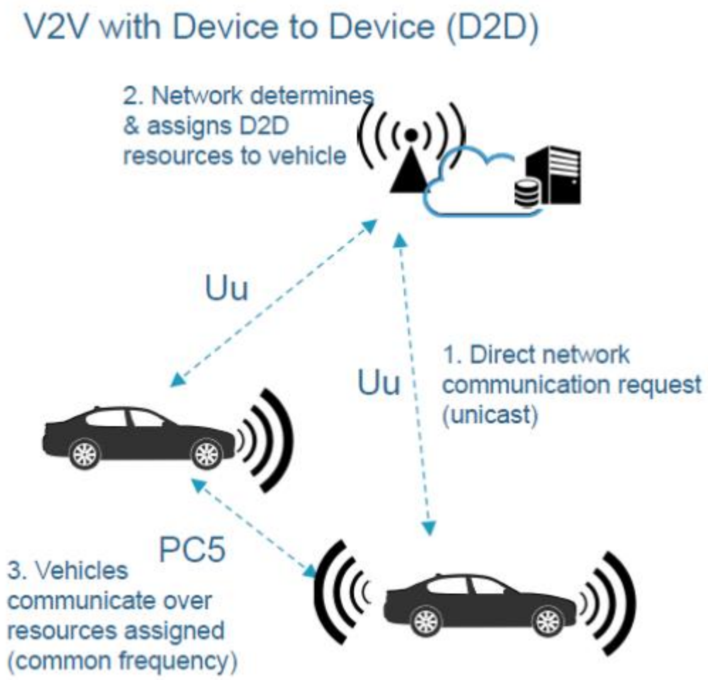


Figure 25 – Model of direct communications between vehicle through P5C interface and resource request and assignment.

4.8.2 Components and functionalities and their mapping to Metro-Haul architecture

The IEEE has been working on optimising 5G for V2X and has been studying the implications and challenges. Here the focus is more on the Metro-Haul network; the key requirement is to support dynamic mobility and high relative velocities between transmitter and receiver vehicles. This has to be done with extremely low latency. Hence the components and functionalities of this Use Case includes:

1. Automated vehicles + sensors: The vehicle (including the individual sensors in it) can be abstracted and virtualised to enable messaging from a sensor to be escalated.
2. Infrastructure: Traffic lights, Parking spaces, Unexpected features on road (e.g., obstacles)
3. Fleet and platooning management applications at the back-end server

At the User, due to the nature of the services, there will be a trade-off between the number of messages with the level of local nodes (for reliable transmission of services), priority of the message due to the situation of the vehicle. At the Edge (Metro-node), the level of local caching of information to be collected may depend on the number of users/vehicles. Metro node resources must support delivery of thousands of individual secured end-user flows at very low rates to maintain secure, reliable and constant throughput over the duration of a journey. Depending on situation (e.g., emergency vehicles) network wide signalling may be required to set infrastructure and ensure other vehicles do not approach high priority vehicles. Given the KPIs provided for V2X, no additional optical layer requirements are needed.

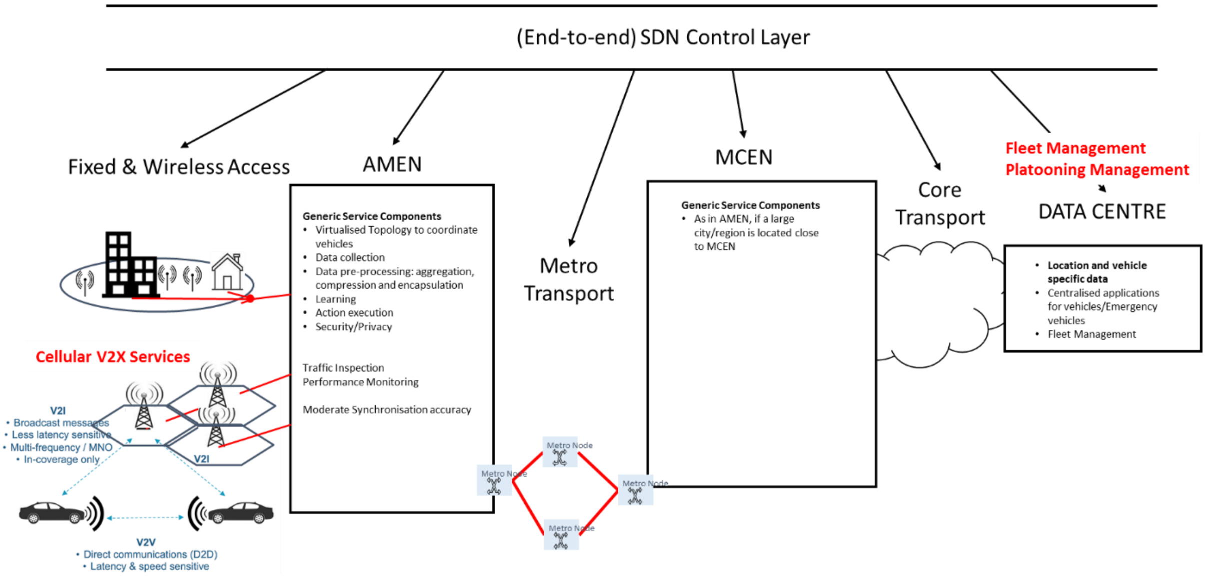


Figure 26 – Mapping of components on Metro-Haul architecture for Autonomous Driving Use Case

Figure 26 shows the mapping of main service components needed by the automotive drive on the Metro-Haul architecture.

4.8.3 Service requirements

Figure 27 shows the connectivity demands including bandwidth and latency (further summarised in Table 21 of future connected vehicles [13]. Depending on the processes that will be outsourced from the vehicle to the communication infrastructure or processed locally at the vehicle, the requirements in terms of latency may be stringent.

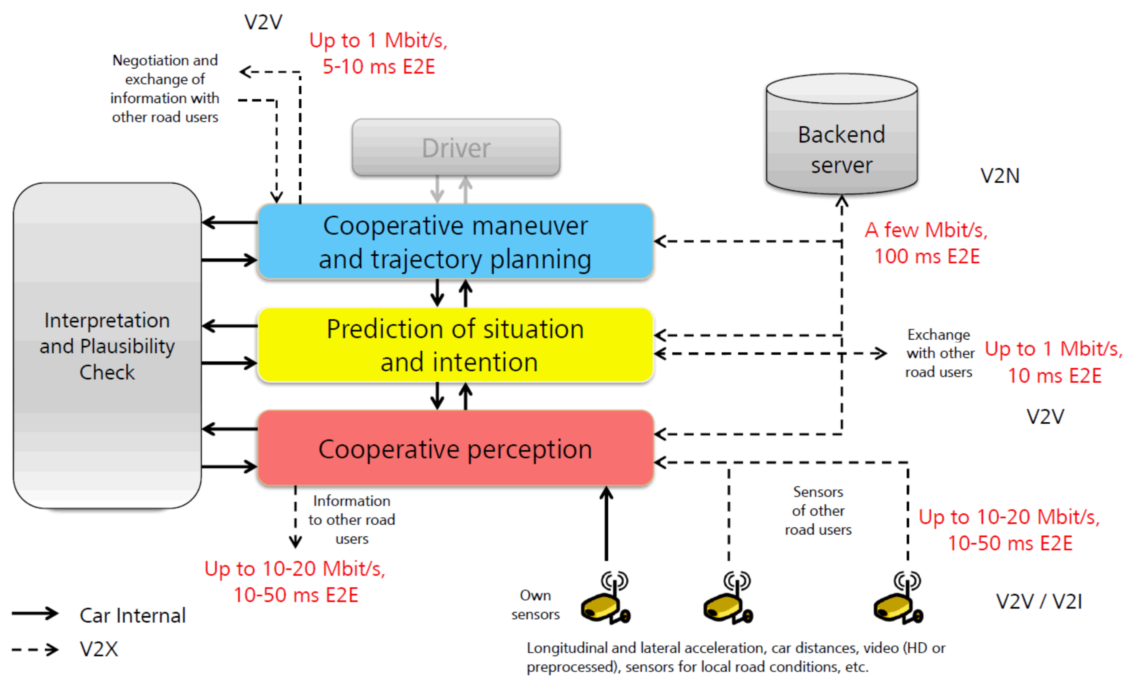


Figure 27 – Connectivity requirements for the future connected vehicle.

To further focus on the key Use Cases for V2X, automated overtaking, cooperative collision avoidance, high density vehicle platooning, see-through and vulnerable road user discovery and bird's eye view Use Cases are presented.

The service category is expending to be a combination of enhanced mobile broadband (eMBB) for V2N, ultra rel. low latency connections (URLLC) for V2V and Critical Connections (CriC) for V2V/V2I/V2N for road emergencies and critical public safety situations. Table 21 provides the KPIs for V2X Use Cases which Metro-Haul will have to abide by to be able to support precise application specific requirements. These requirements are provided by the 5G Automotive Vision outlined by 5G-PPP in collaboration with key stakeholders from Automobile Manufacturers, Suppliers, Telecom Operators, Vendors and Research Institutes.

Table 21 – KPIs of service required for V2X automotive in Metro-Haul.

1	**Bandwidth (data rate) [Mb/s]**	V2N to Backend servers: A few Mbits/s V2V informative mode: Up to 1 Mbit/s V2V cooperative mode: 10-20 Mbits/s V2I/V2V prediction of intention: 1 Mbit/s Teleoperator/Platoon video stream: 10-20Mbits / stream
2	**Latency [ms]**	V2V collaborative mode: 5-10 ms E2E V2V information mode: 10ms E2E V2I: 10- 50 ms E2E V2N: 100ms E2E
3	**Jitter [ms]**	Unknown
4	**Reliability [%] – Scenario dependent**	when vehicles are 10 meters away travelling in the same direction in a sparse road :99.999% When Vehicles are 500 meters away travelling in opposite directions in a crowded road : 99.99%
5	**Availability [%]**	high (five 9s)
6	**Mobility - UE speed [km/h]**	0 km/s static sensors 200 Km/h vehicles
7	**Area Traffic Capacity [Gb/s/km^2] Dependent on vehicle density**	Ranging from 1000- 1400 devices per km^2 in ***Amsterdam.*** About 10-50 cars by Road side unit => About 200 – 1000 sensors per cell (depending on the number of devices)[13]
8	**Rely on Sensor Network [Yes/No]**	Yes
9	**Massive Type [Yes/No]**	Yes
10	**Device Direct [Yes/No]**	Yes
11	**Coverage Requirement [Standard/Extended]**	Everywhere

[13] Note that the sensor density calculations is based upon one sensor connection per car fora car population of 383 cars per 1000 people for a city of 200km^2 e.g. Amsterdam. – Including sensors on infrastructure e.g. indicate road works or occurring travel events.

12	**Energy Consumption critical [Yes/No]**	Yes (must compensate for high penetration losses 15-20dB)
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Special purpose

Table 21 provides only indicative values for KPIs. The above KPIs can be more accurately assigned (especially latency and reliability) once the scenario is investigated in more detail.

4.8.4 Metro-Haul network requirements

To meet a desired service performance dictated by the service KPIs, there are several aspects that should be considered in the Metro-Haul network architecture and infrastructure being defined. These aspects can have a significant impact on the QoS delivered. This architecture will be delivering networking as well computing functions. These challenging Use Cases were introduced initially by the V2X work for 3GPP but is now being adapted to cater for future 5G based V2X scenarios:

***Automated overtake procedures**:* During this procedure, the semi-autonomous or fully-autonomous (in the most ambitious case) vehicle must, upon detecting a slow-moving vehicle, requests updated information on the condition of the roads (traffic, accidents and/or infrastructure), and information from the cars around the vehicle. Hence for this task, traffic management information is distributed from the V2N server (operated by servers for the vehicle service provider or the city or region).

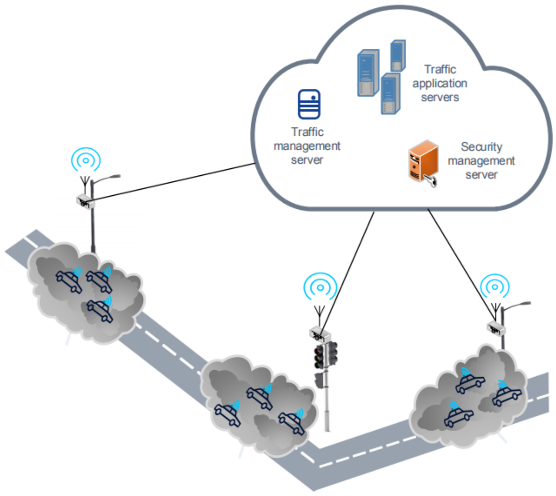


Figure 28 – Interaction of Vehicles to Infrastructure RSUs and Application servers at the edge

These traffic flows will interconnect the infrastructure (so called Road Side Units RSUs, near with the client vehicle via the metro segment to a local V2N server or local compute station see (Figure 28 for an example). It is expected that messages of size 50-300 bytes are used for the periodic broadcast of messages. For the event-triggered messages, 1200 bytes payloads (excluding security bits) are transmitted. In a semi-urban, urban and rural environment, it means that connection between a single road side unit can be between 10 to 50 vehicles in a cell.

From the AMEN, where the V2N servers, compute nodes close to the RSUs and the local/edge V2N servers are placed, a dynamic connection to the MCEN and major coordination between V2N servers is required. The critical issue here is that the AMEN side of the Metro-Haul network must be able to deal with such dynamic increases in the size and number of flows depending on road conditions whilst maintaining strict latency and reliability requirements due to safety. In this regard, the Metro-

Haul network should be capable of bandwidth variable flows (perhaps by adaptive spectrally efficient coded modulation schemes for robust data connections). In addition, highly flexible optical nodes with fast switching and restoration capabilities are required.

Collaborative Collision Avoidance*:* Currently, advanced driver assistance offers radar-based sensor feedback to the driver without knowledge of what new obstacles beyond what forward and rear sensors may provide. Hence this enables a new feature that provides information from infrastructure, from other vehicles and from the overall network management platform regarding obstacles/collisions and where/how to take evasive action (e.g. avoid detected obstacle/blind spots or take avoiding action, or different routes, etc.).

Firstly by collaboratively coordinating the routes of all vehicles, the trajectories of all vehicles/roach users are managed to avoid collision. This requires regular transmission of cooperative awareness messages (CAM messages). These messages are increased in priority if the client vehicle is in danger or collision or has collided so that vehicles in the vicinity of the client vehicle are aware.

From the Metro-Haul network, the latency aspect of this Use Case is particularly critical. This task may require the *edge computing resources* local to the collision event to process the information, flag and reprioritise the message and disseminate in a low-latency and highly reliable manner – this is done in collaboration with the vehicles in the area. Maximum tolerable latency is 100ms as shown in Figure 28.

Vehicle Platooning: In case of convoy management requirements, the involved vehicles plan a path for efficient vehicle traffic flow using the monitoring of data from installed sensors on the vehicles and along the infrastructure road side units to gather information on the environment from neighbouring vehicles (outside of the convoy but also using the roads).

Therefore, due to the higher level of automation required for this task, including the need to deal with highly complex road situations from the Metro-Haul network, ultra-reliable communication links (>99%) is required. In this Use Case, a much lower latency performance is required (between 1-10ms).

Vehicle Platooning can also be co-operated by a remote operator (vehicle teleoperation), in this case, the Metro-Haul network has to ensure high availability of the connections to take control when the platoon of vehicles are in a highly dense or highly unpredictable situation. In this case, uncompressed video streams from the vehicles must be highly available to avoid disrupted video service between the operator and the vehicles (high data rates > 50Gb/s per optical channel in which several video streams of 10-20Mbit from each vehicle is transmitted) is required.

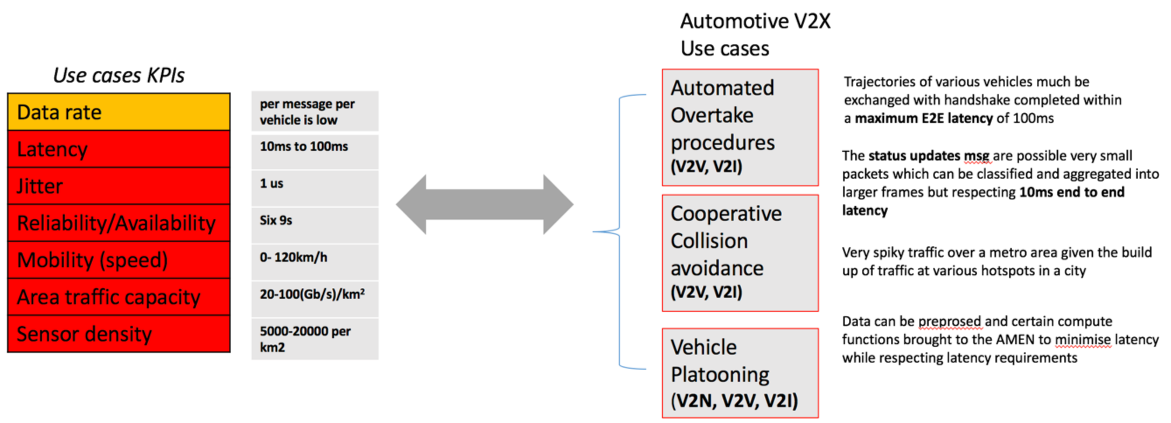


Figure 29 – Selection of critical KPIs and Use Case tasks

Summary of Tentative Network requirements

Optical Layer

- Highly reliable data flows between RSU, AMEN and V2N main servers
- High reliability
- Dynamic data path to cater to change in vehicles connected to each RSU
- Redundancy/resilience – fast restoration
- Low loss guarantee
- High capacity data flows for reliable video streams of 10-20Mbit (Optical Channels of 10-50Gb/s)
- Low latency 10-100ms end to end

Wireless Network

- Multiple technologies (apart from 5G RN), such as centimetre (Cellular V2X) and millimetre wireless frequency bands, IEEE 802.11p and Vehicular Visible light communications are too the means by which the vehicles will communicate with each other and to the road side units. Hence the traffic will be aggregated or backhauled at a very large scale at the edge of the optical network. The user density depends on the urban or rural location of the road side unit. However, a density of 1000-1500 users per km2.

Computing requirements

- Onboard Vehicle computational capabilities
- Edge compute services are required at the AMEN to compute messages from local road side units and to plan, reroute and execute commands or reprioritise messages according to ongoing/live traffic scenarios. These edge compute resource are required to minimise latency 10 – 100ms and to support reliable service.
- Further compute resources are required at the V2N servers close to the MCEN for the overall management of traffic

Table 22 – Summary of Metro-Haul requirements (KPIs) for 3 main Automotive Use Cases

Task	Throughput			Storage		Computing capacity	
	AMEN	MCEN	Optical	AMEN	MCEN	AMEN	MCEN
Automated overtaking procedures	0.2 Gb/s		Shared 100G	100 GB		1 vCPU per 2 sensors	
Cooperative collision avoidance	0.01 Gb/s		Shared 100G	100 GB		1 vCPU per 10 sensors	
Vehicle Platooning	10 Gb/s		Shared 100G	5-10 TB		1 vCPU per 10 sensors	

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Automated overtaking procedures	Transmission degradation detection/anticipation	Fast-restoration mechanism

Cooperative collision avoidance	Transmission out of service detection	Fast-restoration mechanism
Vehicle Platooning	Transmission degradation detection/anticipation	Fast-restoration mechanism

4.9 Enterprise Access with NG Ethernet

This Use Case is motivated by some ideas contained in the project proposal. For instance, it is stated that "These optical node designs will include line and client side flexibility in dynamic network operation, in-line with Ethernet standard evolutions targeting both higher and lower rates than 100G (e.g., 25 GbE)." Moreover, a figure depicted an enterprise access to an AMEN through an NG Ethernet Interface. It should be noted that, although 5G networks are primarily envisioned for wireless access, they will be convergent networks able to provide fixed services as well.

At the same time, the motivation of this Use Case is to test Naudit's active and passive probes to be developed in the context of this project.

4.9.1 Use Case Description

Large enterprises (e.g. Banks) need very high bandwidth and low latency access to the Metro Network to support their needs to operate their IT infrastructure. For example, some common services with these requirements include:

- Interconnect two Data Centres (DCs) (primary and secondary) or a private and a public cloud, as well as central premises. This is a clear case of high bandwidth consumption (several tens of Gb/s). In fact, an expensive dark fibre is usually contracted for this.
- High available large databases, with low delay and high bandwidth interconnections for data replication. In this case, low delay is needed to keep data base coherence with synchronous replication. A similar case would be the replication of a disk enclosure.

Given the latency requirements (see below), such connections could be classified as ultra-reliable low latency (URLLC).

To provide such access, it is common that network operators provide a router with a standard interface at the edge of the enterprise premises. Current link standard is clearly Ethernet, with speeds ranging from 10 Mbps to 400 Gb/s. Flex Ethernet has also been proposed [32]. Current high speed Ethernet interfaces are flexible. For instance, a 100 Gb/s interface can be configured as: 1x100 Gb/s, 1x40 Gb/s, 4x25 Gb/s, or 4x10 Gb/s interfaces.

Given the cost of such high bandwidth and low delay service, large enterprises need to check the SLA related to the service they are paying to the operator. Thus, it is necessary to check both capacity and delay, as well as availability and reliability provided by the network between Data Centres, connected to the network through new Ethernet flexible standards. Both active and passive monitoring probes are advisable for this task, which make this Use Case a good scenario where to test Naudit's probes to be developed at Metro-Haul.

To stress the importance of this Use Case, and take some requirements that have to be met, we provide some facts and figures:

The average cost of DC outage was around 8K$ per minute in 2016. [43]. This cost can be caused because central premises' work depends on their connection to the DCs. Thus, it is very important to keep availability of DCs. For this, enterprise DCs are usually deployed in pairs (primary, secondary). Enterprises also combine public and private clouds. Tier IV DCs (fault tolerant) have to assure an availability of 99.995% (26 minutes downtime per year). For instance, in Spain, BBVA and Telefónica have such DCs (with Naudit probes deployed on them). [44].

Synchronous replicating services need very low latency. For instance, 20 ms is the upper limit when synchronously replicating DB across DCs [45]. Other example is the replication of disk enclosures/cabinets, where needed latency is even lower, around 5-10 ms [46].

Traffic in large enterprises is very variable, with high rates in the peak hour and large valleys at nights. This fact foster for the adaption of the bandwidth links to their needs and schedule network resources dynamically. Operators can also benefit from this solution, providing more bandwidth to other Use Cases at other times when it is not necessary for large enterprises. In conclusion, the network provided by Metro-Haul can be very valuable, lowering Large Enterprises network costs when compared to dark fiber links.

4.9.2 Components and functionalities and their mapping to Metro-Haul architecture

Figure 30 provides a view of the components of this Use Case.

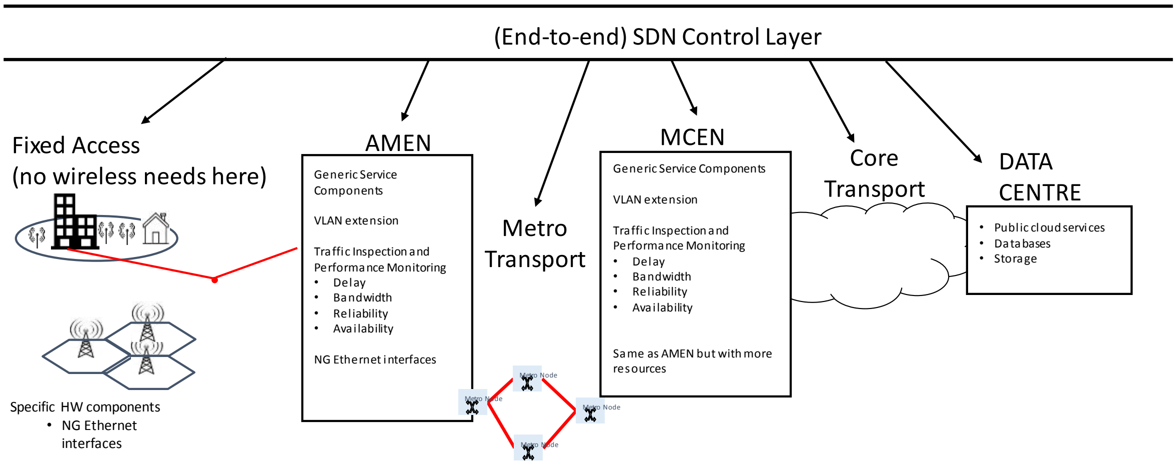


Figure 30 – Mapping of components on Metro-Haul architecture for Enterprise Access Use Case

As shown, in this Use Case no wireless access is necessary, but rather a high bandwidth fixed access for large enterprises. It is expected that, to connect these enterprises with the Metro-Haul network, access routers with NG Ethernet interfaces, ranging from 10 to 100 Gb/s data rates will be deployed.

At the enterprise DCs and central premises critical electronic services will be deployed, including:

- Services provided to internal and external users
- Large database replication between data centres
- VLAN extension across different premises

At the AMEN/Metro network it will be necessary to have:

- NG Ethernet interfaces, ranging from 10 to 100 Gb/s speed
- Active and passive probes to measure provided QoS
- 100 Gb/s VLAN extension along a metro network

At the MCEN/Core network it will be necessary to have:

- Routes to other infrastructure, such as Cloud PoPs, with NG Ethernet interfaces in the PoP

- Active and passive probes to measure provided QoS
- 100 Gb/s VLAN extension along a whole country network

4.9.3 Service requirements

Table 23 shows the KPIs of this Use Case. Bandwidth needs are due to the number of users in the enterprise. Latency needs are due to the replication services to be deployed on top of the network. Availability of the network is needed in order to achieve five nine availability in the interconnection of the enterprise.

Table 23 – KPIs of service required by Enterprise Access Use Case.

1	**Bandwidth (data rate) [Mb/s]**	Ultra-High (10 to 100 Gb/s on peak)
2	**Latency [ms]**	Stringent (5 ms to max 20 ms)
3	**Jitter [ms]**	10% of latency (0,5 to 2ms)
4	**Reliability [%]**	High (five 9) due to the latency requirements and the time it takes to retransmit a packet, affecting also throughput.
5	**Availability [%]**	high (five 9)
6	**Mobility - UE speed [km/h]**	No mobility needed in this Use Case
7	**Area Traffic Capacity [Gb/s/km^2]**	High
8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	Yes
11	**Coverage Requirement [Standard/Extended]**	Large enterprise premises, fixed
12	**Energy Consumption critical [Yes/No]**	No
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Large servers (database, web architectures…)

4.9.4 Metro-Haul Network requirements

In order to meet a desired service performance in terms of crucial service KPIs, there are crucial tasks to be deployed in the Metro-Haul infrastructure that directly impact on service quality. Specifically, the most relevant tasks are:

- *Aggregated enterprise access to the network*: Given current and future bandwidth needs at large enterprise premises, it is expected that such landline links to the Metro Network will have to be of very high capacity, ranging from 10 to 100 Gb/s, with future connections of 400 Gb/s. The availability of the network has to be very high, given the importance of connectivity in large enterprises such as banks. In fact, they already have data centres with five 9s, so the connection to them has to be also five 9s to assure the continuity of the service. Large enterprise premises are usually in large buildings or campuses, where most employees are located. This implies a high density of flows from the employees connecting to the data centre servers.

- *Data replication*: A very concrete task in large enterprise is to assure the availability of critical data, which is usually stored in databases or disk enclosures/cabinet. For this, it is necessary to provide a synchronous data replication between data centres, which requires both high data rate and low latency. Thus, for this case we foresee the need of 10 to 100 Gb/s links with very low latency ranging from 5 to 20 ms at most. Data replication could also benefit from caching policies with copy-on-write to reduce the perceived latency.
- *VLAN extension*: In order to interconnect Enterprise premises and data centres, it is necessary to have VLAN extensions that allow the replication of services in independent locations, with a seamless connectivity across the network. Such service has to be available with five 9s in order to assure the continuity of the service.

The connection of the crucial tasks with service KPIs are the ones reported in Table 24.

Table 24 – Impact of crucial tasks on Service KPIs in Enterprise Access with NG Ethernet Use Case

Service KPI	Value	Crucial Tasks requiring most stringent KPI value
Data rate	10 to 100 Gb/s	**Aggregated enterprise access to the network** **Data replication**
Latency	5-20 ms	**Data replication**
Jitter	0,5-2 ms	(A 10% of latency is admissible)
Availability	Five 9s	**Aggregated enterprise access to the network** **Data replication** **VLAN extension**
Mobility (speed)		(not relevant)
Area traffic capacity		**Aggregated enterprise access to the network**
Sensor density		(not relevant)

In order to meet the service requirements, it is expected that the Metro-Haul network meets the following requirements:

- Access routers have to provide NG Ethernet interfaces, ranging from 10 to 100 Gb/s.
- SLA checking, through both active and passive network probes.
    - Active network probes need to measure bandwidth up to 100 Gb/s and latencies under 1 ms. To have a good measurement of packet loss (reliability) it is necessary to send at least ten times the inverse of such packet loss. For instance, in order to measure a loss of 0,001%, 1E6 packets will be sent.
    - Passive network probes need to provide consumption time series with 1s granularity, being able to differentiate among VLANs or other tags (e.g. MPLS). Other records, such as NetFlow/IPFIX are also advisable.

- Recovery mechanisms and reconfiguration: in order to keep KPIs, once the monitoring and data analytics systems detect a problem, the network has to be reconfigured, providing recovery mechanisms as well.
- Network slicing: In order to interconnect different premises of a large enterprise, or for multicast data replication, the network has to provide network slicing for the VLAN extension service. Probably, a filterless architecture, as proposed in WP3, is useful to meet this requirement.

Table 25 presents some tentative figures of key Metro-Haul requirements. Only relevant cells for this Use Case have been filled and they provide the mappings between these tasks and Metro-Haul KPIs.

Table 25 –Metro-Haul requirements for enterprise access with NG Ethernet.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Aggregated enterprise access to the network	100 Gb/s	>100 Gb/s	>100 Gb/s			Virtual and Physical functions for Monitoring tasks	
Data replication	100 Gb/s	>100 Gb/s	>100 Gb/s	Caching policies with Copy on Write for data synchronization	Caching policies with Copy on Write for data synchronization	Virtual and Physical functions for Monitoring tasks	
VLAN extension	100 Gb/s	>100 Gb/s	>100 Gb/s			Network slicing	

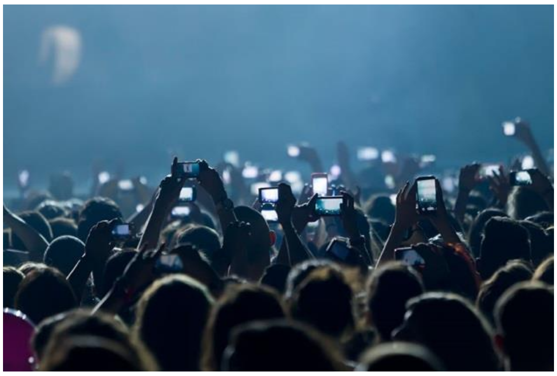
Task	End to End latency (RTT) and jitter	Monitoring and Data analytics	Management, Control and e2e Orchestration
Aggregated enterprise access to the network	<5 ms [10%]	Active measurements: Capacity, delay, packet loss (reliability) and availability.	Recovery mechanisms
Data replication	<5 ms [10%]	Passive measurements: Time series and Netflow records	Recovery mechanisms
VLAN extension		Passive measurements: Time series and Netflow records	Global reconfiguration

4.10 Crowdsourced video broadcasts

This Service Use Case will be demonstrated at the end of the Project in the UK. Further detailed definition of the demo is part of the Metro-Haul Work package 5.

4.10.1 Use Case description

Consider a sporting event in a crowded stadium. Often, spectators will film on their smartphones, later posting these videos to YouTube, Facebook, etc. Even if their content is generally mundane, these videos capture the unique perspective of the observer, and views potentially missed by professional videographers, even if present. Unfortunately, given a multitude of sources, such video content is difficult to browse and search.



Crowdsourced live streaming (CLS) platforms allow general users to broadcast their content to massive viewers, thereby greatly expanding the content and user bases. CLS services enriches social interactions across communities by allowing users to share common interests and stories, Furthermore, if organisations hosting such events could easily capture all those video content in an interactive and highly shareable visual story about the event this would boost its social reach immensely.

Services like Watchity, Periscope etc. provide such cloud based CLS systems to extract content-specific metadata from crowd sourced videos which identifies and forms a "clusters" of synchronized streams with related content surrounding the event. These clusters are made available online so that event organisers can capture different views and combine them to portray their brand or message.

Synchronizing large, high volume streams and capturing them at a cloud service require a high capacity network infrastructure which adapts and scales dynamically to accommodate multiple synchronized streams. Moreover, the heterogeneous quality, format and information of source stream require computational resource to transcode it into multiple industrial standard quality versions to serve viewers with distinct configurations.

4.10.2 Components and functionalities and their mapping to Metro-Haul architecture

Crowdsourced video broadcasting workflow depicted in Figure 31 includes components acting as follows:

- Capture devices (smartphones with GPS, Compass, Accelerometer, Gyroscope) upload content along with sensor data through the crowdsource video sharing application
- The video app will analyse the content and its associated data (location, position, angle) to filter and aggregate multiple videos into clusters
- As the clusters increase the APP will request the orchestrator to provide computing resources and low latency path to these computing resources to aggregate and synchronize its video streams as close as possible to the sources. Service orchestrator converts this request and feeds it to respective controllers to provide edge compute resources along with low latency paths
- The crowd source video APP will also request orchestrator, via APIs, for high capacity network paths to reach the cloud provider where the clustered videos will be further edited and be prepared for consumption by users. Orchestrator will translate the request to network requirements and feed the path setup configurations to involved device controllers.

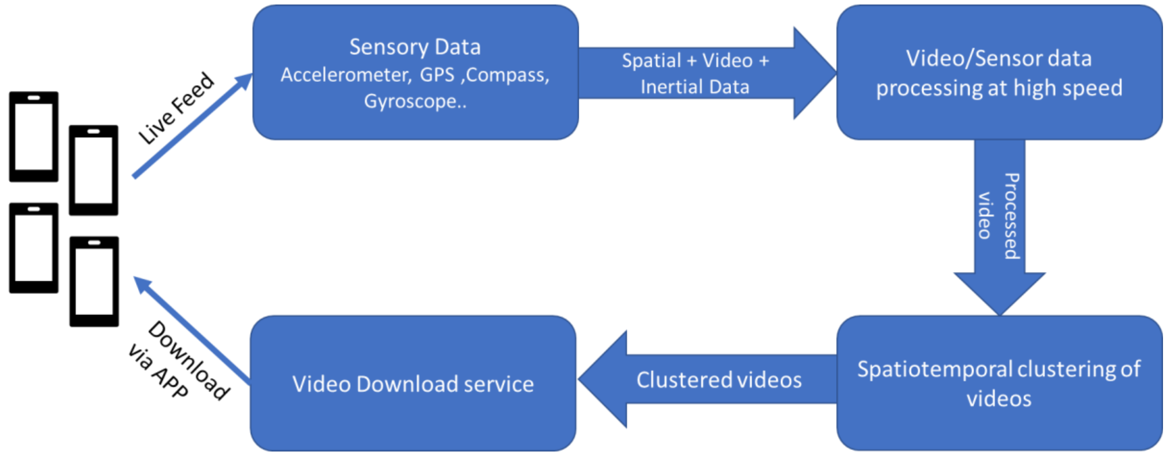


Figure 31 – Crowdsourced video broadcasting workflow.

Figure 32 provides a view of the components of this Use Case mapped on Metro-Haul architecture.

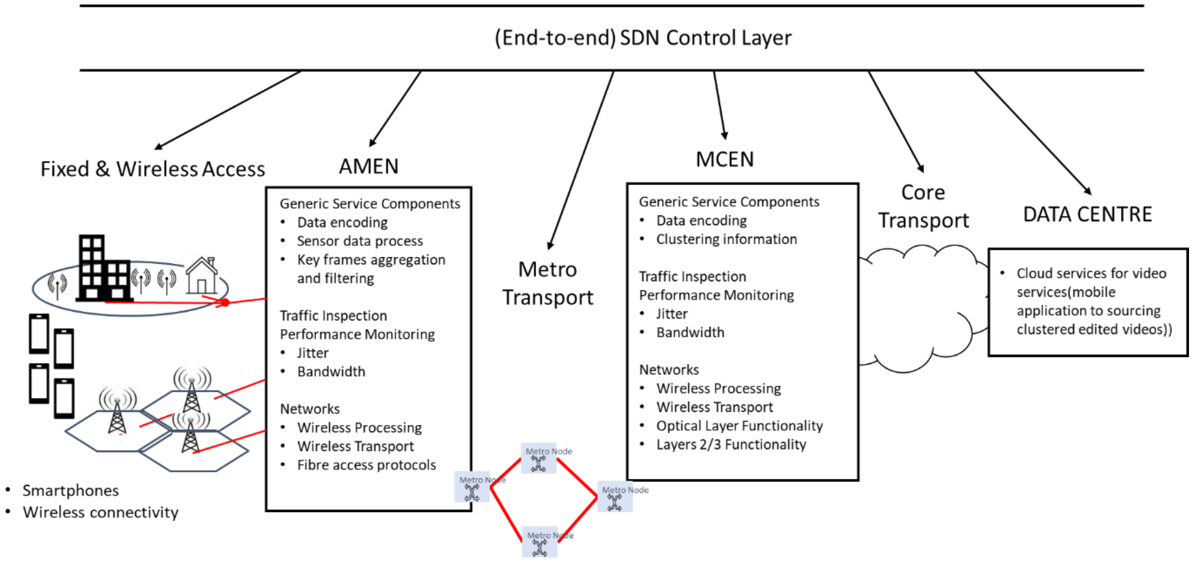


Figure 32 – Mapping of components on Metro-Haul architecture for Crowdsourced video Use Case

At the AMEN/Metro network it will be necessary to have:

- Computing capability at AMEN to process video and sensor data to aggregate and form clusters of crowd videos
- Dynamic high capacity paths to upload clustered video streams. Data rates of >10Gb/s for a 30K capacity stadium
- Monitor and gather video application requirements and provide network services.
- Provide virtual load layer4/7 load balancers, content caching, compression and TCP/IP multiplexing

At the MCEN/Core network it will be necessary to have:

- Identify individual video clusters to provide paths to different datacentre facilities
- Multiple routes (upload/download multi-path TCP) to reach video application data centre
- Monitor and gather video application requirements and provide network services. E.g. Jitter, bandwidth, as the number of video clusters allocate more paths to different cloud locations
- Provide load balancing, between multiple data centre routes, to manage scale

4.10.3 Service requirements

Table 23 shows the KPIs of this Use Case. Bandwidth needs are due to the number of users at a venue trying to use the service. Latency needs are due to the video processing services that need to be deployed on top of the network. Availability of the network is needed to achieve five nine availabilities needed for video reproduction.

Table 26 – KPIs of service required by Crowdsourced Video Use Case.

1	**Bandwidth (data rate) [Mb/s]**	800 Kb/s to 2400 Kb/s per user (240p video, for instance, uses about 1.66 megabytes of data per minute. A 360p video uses 2.66 megabytes per minute, and a 480p uses 4 megabytes per minute.)
2	**Latency [ms]**	Loose (30 ms)
3	**Jitter [ms]**	Stringent
4	**Reliability [%]**	high (five 9)
5	**Availability [%]**	high (five 9)
6	**Mobility - UE speed [km/h]**	UE (client video devices mobility across different access technologies (seconds)
7	**Area Traffic Capacity [Gb/s/km^2]**	High (>10000 users/km^2)
8	**Rely on Sensor Network [Yes/No]**	Yes
9	**Massive Type [Yes/No]**	Yes
10	**Device Direct [Yes/No]**	Yes
11	**Coverage Requirement [Standard/Extended]**	Extended
12	**Energy Consumption critical [Yes/No]**	Medium (save battery on phone)
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Smartphones

4.10.4 Metro-Haul network requirements

To meet the desired network requirements, it is expected that the Metro-Haul network will provide a solution to the following requirements:

- High capacity video upload paths: one of the key requirement for crowdsourced video uploads is scalability. As the number of user's upload increase it quickly consumes metro network bandwidth. Constant high bandwidth bit rate over variable bit rate is better for video uploads. Bandwidth requirements are as follows
    - Multiple wireless technologies to aggregate user traffic at large scale. Typically 1-3Mbps/user to stream compresses 4K video and user density >10000 km2

   - *(Bit rate per second of all combined video streams and the audio stream) x (Number of formats you are broadcasting in) x 2 = Required Upload Speed (>10Gb/s for 27000 capacity stadium)*
   - Sports and video game streams are often encoded at 24 and 60 fps, which require higher bandwidth, so viewers can catch the split-second action that occurs and that takes more bandwidth
   - Most live streams require multiple bit rates for video to be streamed at the same time

  Elastic and reliable optical links can cater to such high bandwidths without over-provisioning inefficiencies. Metro-Haul AMEN nodes are expected to aggregate the video uploads in bandwidth in multiples of 10G. MCEN nodes are used distribute the load towards application data centres which host the video repositories. Dedicated and reliable 100G paths through MCEN are required to datacentres hosting the video application.

- Video streaming delay increases if the network latency increase. For crowdsourced video streaming jitter and minimum packet delay variation (PDV) should be <30ms along with low tolerance for loss <5%. AMEN and MCEN nodes should be able to provide monitoring data to deduce jitter and packet delays for clustered streams. They must be able to measure quality of service of video streams by sending probes to measure path jitters, PDV and loss.

- Video Clustering: Clustering/Grouping of video uploads based on their location and content is one of the key features of crowdsourced video streaming. Based on the sensory data from smartphones and the video quality, a set of streams originating from a geo-location are geo-tagged and clustered together. Few high-quality videos streams are chosen rather than many low/medium quality streams. AMEN nodes with computing capability can help in processing smartphone data along with the video streams. AMEN nodes need to provide high speed, low latency connections to compute nodes to process and tag video clusters typically <5ms latency.

A summary of main KPIs as resulted from the Crowdsourced Video Use Case is reported in Table 27.

Table 27 – Summary of Metro-Haul requirements (KPIs) by Crowdsourced Video Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Bandwidth for user video streams	>10 Gb/s		Multiple moderate pipes				
Transport of clustered videos to data centres		100 Gb/s	Multiple fat pipes with reliability				
VNFs for load balancer		10 Gb/s				2 vCPU per VNF	
Video and Sensor data processing				100 GB		1 vCPU per 100 video streams	

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration

Bandwidth for user video streams	Jitter and Latency monitoring	Scalable traffic paths, Improve quality of service
Transport of clustered videos to data centres	Jitter and Latency monitoring Transmission degradation detection/anticipation	Dynamically allocate paths between domains, Optimal Planning of capacity resources, improve QoS
VNFs for load balancer		Dynamically scale computing resources based on VNFs
Video and Sensor data processing	VM monitoring	Scale computing resources based on streams

More precisely, mapping such requirements on the specific Metro-Haul architecture it results the following tasks and functionalities required in the Metro-Haul nodes:

- Massive users per area (>10000users/sqm2) uploading videos at high definition (1-2Mbps)
- Computing capability at AMEN to process video and sensor data needed to form clusters
- Aggregate and form clusters of crowd videos at AMEN
- Provide dynamic high capacity paths in MCEN to upload clustered video streams
- Platform to monitor and gather video application requirements and provide network services (AMEN/MCEN). E.g. Jitter, bandwidth, as the number of video clusters allocate more paths to different cloud locations
- Platform to provide virtual load balancers in MCEN to provide upload/download paths from data centres. Capabilities include layer4/7 load balancing, content caching and compression, TCP/IP multiplexing.

4.11 Real-Time Low-Latency Object Tracking and Security

This Service Use Case will be demonstrated at the end of the Project in Berlin. Further detailed definition of the demo is part of the Metro-Haul Work package 5.

4.11.1 Use Case Description

Novel 5G Metro-Haul technology developed within the project is showcased in a future smart city scenario, where security is provided by real-time object recognition and tracking. Simultaneous real-time access to the data of several fixed and mobile cameras allows tracking of objects and persons as well as the automatic recognition of critical events which encompass areas larger than the field of view of a single camera (e.g., population density). Mobile cameras may be body cams or mounted on drones connected to nearby base stations depending on the specific application scenario. The number of used cameras and their data rate can be flexibly adapted to the situation and resource availability at hand. Novel low-latency networks and distributed edge computing concepts enable machine-to-machine and human-machine interactions which are not yet possible with current technology. Examples are:

- Automatic camera handover even over different access metro edge nodes,
- Centrally controlled auto-tracking of Pan-Tilt-Zoom (PTZ) cameras
- Event-driven control of stationary cameras by on-site security forces (e.g. by making use of their location data).

While cameras alone only provide tracking of anonymous objects and people, their combination with other data sources enables simultaneous recognition and identification or people counting. Thus, the system may also be used for civil or commercial applications such as public transport optimization, traffic management and customer flow analysis. Using a Video Management System video streams are collected from CCTV cameras inside a local processing node (µ-DC) for initial video analytics (e.g., object tracking, face recognition, people counting), sends the streams towards the core DC for storage. A general diagram for real-time low-latency object tracking and security is seen in Fig. 33.

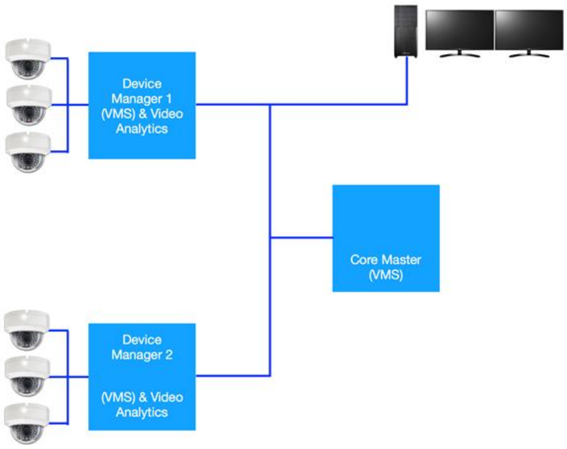


Fig. 33 - Schematic Diagram of real-time low-latency object tracking and security highlighting a Video Management System (VMS).

- **Use Case Demonstration**

This Use Case will demonstrate the embedment and support of a video surveillance application in a 5G Metro-Haul network scenario and its SDN-based orchestration. It shall showcase the solution's capability to exchange real-time data aggregation from multiple sources and to store and compute within appropriate DC resources.

- **User Community**

Police and public security forces, crisis management and public infrastructure recovery after disaster events, public event management, public and private transport management, flow optimization of all involved traffic participants, smart city and traffic surveillance

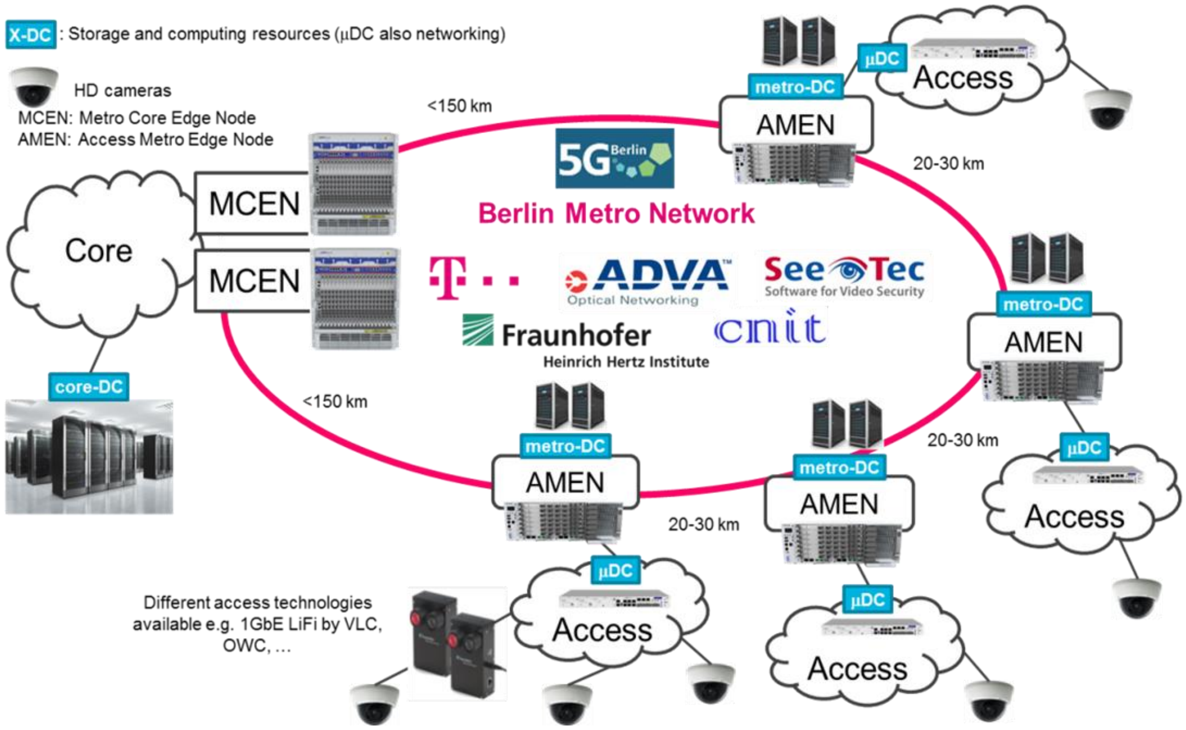


Figure 34 – Scheme of RT-LL-OT&S Demonstration in Berlin.

4.11.2 Components and functionalities and their mapping to Metro-Haul architecture

A geo-distributed architecture is best suited to the low response times and high-bandwidth requirements.

1. RT video surveillance with RT object tracking, automated face recognition and handover between multiple flexible camera devices
2. Low-latency connectivity between fixed and mobile cameras assigned to different AMEN and MCEN with their associated DC storage and computing resources
3. Enable flexible distributed cloud computing capabilities for real-time dynamic object tracking applications
4. Comprehensive SDN-based network and DC control utilizing standardized open APIs.
   - Based on an open platform architecture, the proposed approach combines edge computing, CORD (Central Office Re-Architected as Data Centre) and cloud computing in a three-stage network concept (Figure 1). Virtualization enables multi-tenancy while dynamic provision of location, capacity and latency information allows to optimally map services to the available resources. Standardized interfaces shall provide simple integration of stationary and mobile user equipment.
   - The management of the platform uses SDN-based network control as well as computing and storage resources. An orchestrator allows setup and end-to-end management of distributed software functions. Open APIs enable coordination between the application layer and the infrastructure layer of the platform.

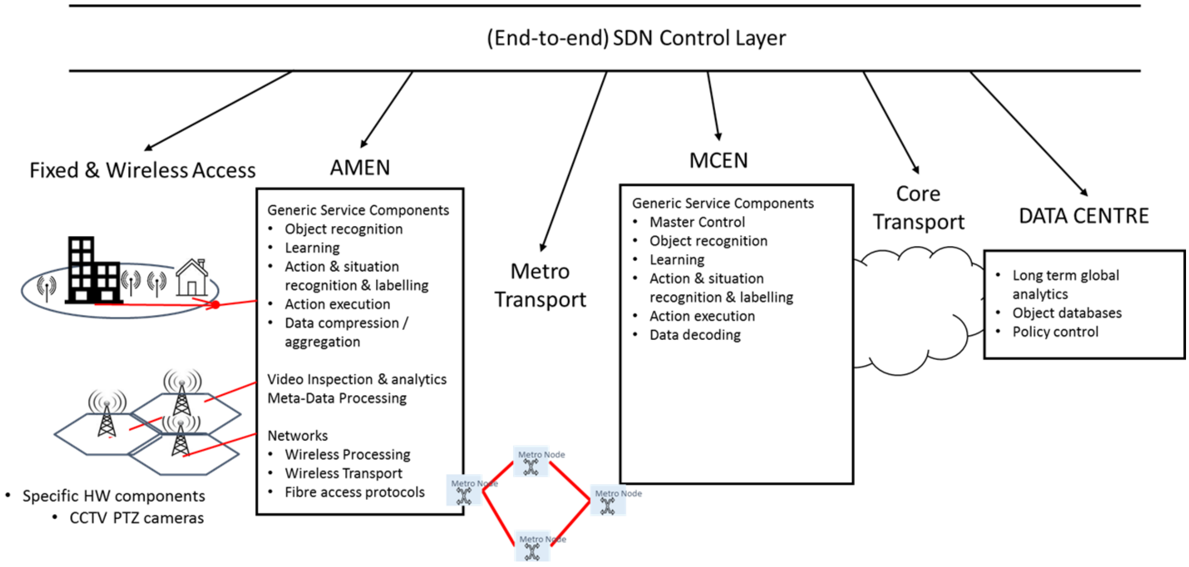


Figure 35 – Mapping of components on Metro-Haul architecture for RT-LL-OT&S Use Case

4.11.3 Service requirements

The transport of surveillance video streams sourced by CCTV cameras can pose a moderate bandwidth request on individual cameras, ranging from 2 up to 25 Mbit/s for SD and UHD, respectively. As an example used as a reference here, in certain areas of London, one of the top cities in the world with the best camera surveillance coverage, like Wandsworth district (34 km^2), government CCTV camera count can exceed 1000 points [48] covering public places, schools, stations, and major retail areas.

According to a 2015 report by the Information Handling Services on the installed base for video surveillance equipment a Compound Annual Growth Rate of 20% is projected for the next 5 years. This leads us to a base of calculation of 2500 cameras in districts like Wandsworth. As full HD is already considered a de facto standard on video quality and 4K is rapidly increasing in adoption, we assume a high video quality stream with a data rate per camera of 25 Mbit/s uplink throughput to the AMEN, which would amount to bandwidths of 62.5 Gb/s. In the case of tracking and motion activation PTZ cameras can double the video throughput. In order to decrease network utilization, learning algorithms can be implemented to variably adjust the streaming quality according to certain areas, events, notifications etc.

Default video storage could be implemented in the AMEN POP for better resource utilization and retained for the duration of up to 30 days according to the EU General Data Protection Regulation [49].

Other limits related to network induced latency range from 10 ms refer to active object recognition and tracking control. Mobility is another aspect that is expected to reach up to 50 km/h for drones or other mobile cameras.

Due to the public safety and security implications, this service falls into the category of Critical Connections with Ultra Reliable Low Latency Connections (CC/URLLC). KPIs of the service are collected in Table 28.

Table 28 – KPIs of service required by RT-LL-OT&S.

1	**Bandwidth (data rate) [Mb/s]**	UE (CCTV Camera) Constant Data Rate: • Low (2.1-4 Mbit/s for SD and HD) • Moderate (10 Mbit/s for full HD) • High (25 Mbit/s for 4K UHD)
2	**Latency [ms]**	Low (< 10 ms for active object tracking control) to Medium (10 - 50 ms)
3	**Jitter [ms]**	un-critical
4	**Reliability [%]**	High (five 9)
5	**Availability [%]**	Moderate (95 – 99.99%)
6	**Mobility - UE speed [km/h]**	0 km/h (static PTZ CCTV cameras) Up to 50 km/h (drones with cameras and other mobile cameras)
7	**Area Traffic Capacity [Gb/s/km^2]**	[Normal] 1 - 10 Gb/s/km^2 (uplink) [Hotspot – Train Stations, Stadiums, Public Places] Up to 50 Gb/s/km^2 (uplink)
8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	Base case: No Extended case: Collaborative Video Surveillance: Yes
11	**Coverage Requirement [Standard/Extended]**	Standard
12	**Energy Consumption critical [Yes/No]**	No
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Special Purpose (CCTV Cameras)

4.11.4 Network requirements

To meet a desired service performance in terms of crucial service KPIs, there are a series of tasks to be deployed in the Metro-Haul infrastructure. Specifically, the most relevant tasks are:

- Video stream distribution: video flows are transported from the UEs (CCTV cameras) to the designated AMEN for initial real-time low-latency object tracking and backhauled towards the core DC for storage. Metadata extracted is sent for further processing towards the MCEN. Client streams can be located in a centralized location (MCEN) or on mobile devices upon request.
- Video analytics: due to low latency requirements for real time object tracking, initial video processing occurs in the AMEN. Functionality for controlling auto-tracking of Pan-Tilt-Zoom (PTZ) cameras is envisioned here as well. Additional processing for object tracking between different AMEN serviced cameras occurs in the MCEN. The MCEN is also responsible for hosting network wide VM deployment and dynamic dimensioning using global reconfiguration techniques. Mobile camera roaming between different AMEN serviced access points is controlled from the MCEN.
- Master Control, action execution on MCEN.
- Analog Video Encoding on AMEN.

Table 29 presents some tentative figures of key Metro-Haul requirements. Values are based on the assumption of 1000 CCTV camera administrated by an AMEN. Note that only relevant cells for this Use Case have been filled

Table 29 – Summary of Metro-Haul requirements (KPIs) by RT-LL-OLT&S Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
CTTV Stream Distribution	25 Gb/s / 1000 channels		25 Gb/s / 1000 channels	4000 TB / 1000 channels			
Video Analytics: Object/Face recognition, Metadata Extraction		1-2 Gb/s / 1000 channels	1-2 Gb/s / 1000 channels			1 vCPU / channel / algorithm	1 vCPU / channel / algorithm
Master Control, action execution							1 vCPU / channel
Analog Video Encoding						1 vCPU / channel	

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
CTTV Stream Distribution		
Video Analytics: Object/Face recognition, Metadata Extraction	CCTV out of service detection Object tracking	Possible recovery mechanism Control auto-tracking of Pan-Tilt zoom (TTZ) cameras
Master Control, action execution	VM monitoring	Global reconfiguration VM dynamic dimensioning (local reconfiguration)
Analog Video Encoding		

4.12 Secure SDN Controlled Video Distribution

4.12.1 Use Case description

Quantum Key Distribution (QKD) is an innovative method for exchanging cryptographic keys by using single photons from a sender (Alice) to a receiver (Bob) over a quantum channel. Fundamental laws of physics prevent an eavesdropper (Eve) from learning the key, as any attempt made to gain information about the photons will irreversibly change them in a manner that can be detected [50].

In this Use Case, the QKD network is comprised of quantum and classical channels [51]. All the channels are based on optical fibre links, with intermediate optical switches at different nodes. The video contribution network is comprised of classical optical channels, end-to-end from the encoder to the decoder [52]. The optical switches are SDN-enabled, with an Agent using extended OpenFlow protocol to establish the communication between the hardware and the SDN Controller The switches can be flexibly (re-)configured by the SDN Controller. The SDN Controller is aware of the network topology, thus making it easy to bypass any amplifier in the path, which is detrimental to QKD networks.

This Use Case proposes the use of SDN to mitigate a Distributed Denial of Service (DDOS) attack over a video contribution network which is secured with QKD resources. Figure 36 (a) shows a preliminary testbed that emulates the DDOS attack. Figure 36 (b) shows also the DDOS mitigation flow chart [52]. When a DOS attack is detected the SDN Controller allocates new secure paths that will perform video transmission and the key exchange. The video is then encoded/decoded in the terminal nodes, using the key generated from the QKD network.

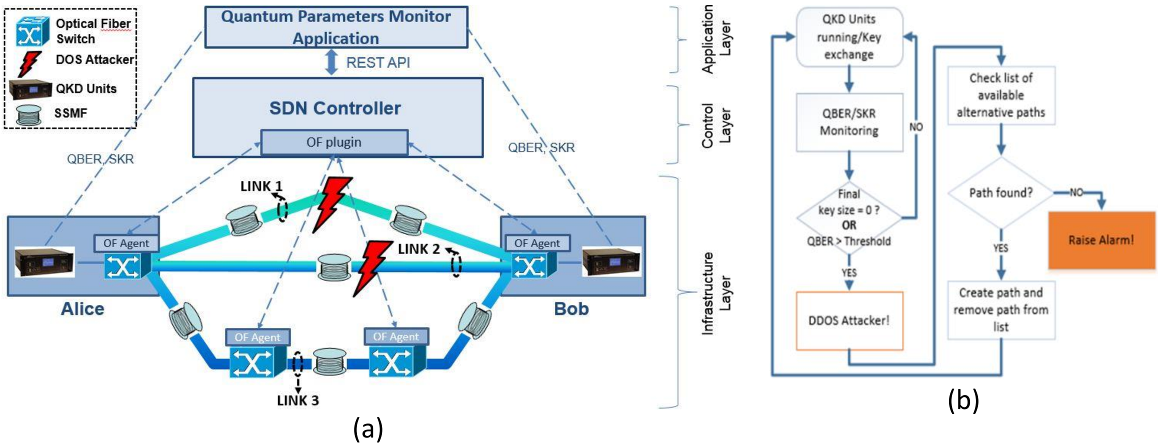


Figure 36 – Mapping of components on Metro-Haul architecture for Crowdsourced video Use Case

4.12.2 Components and functionalities and their mapping to Metro-Haul architecture

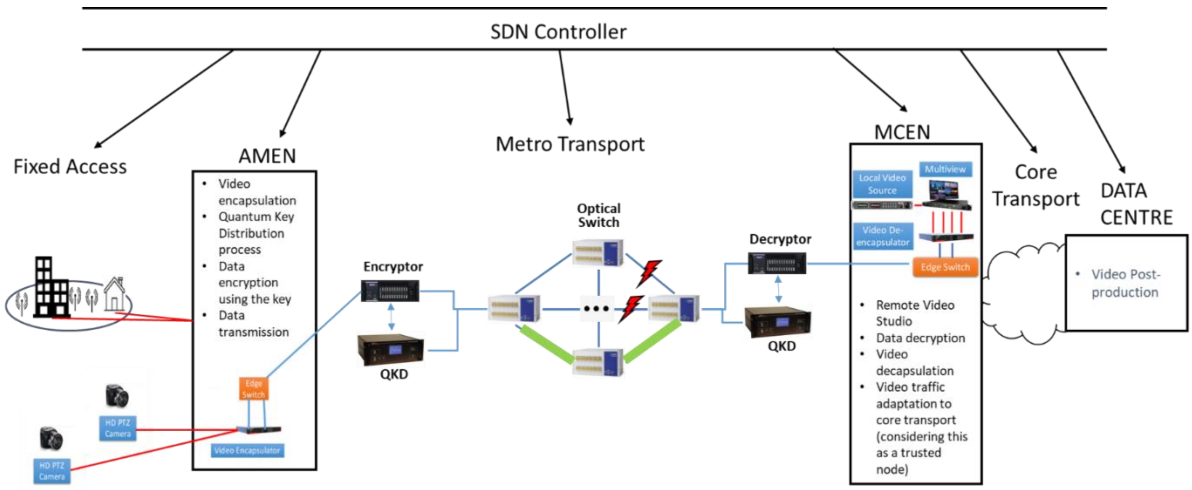


Figure 37 – Components of the secure SDN video distribution and mapping to Metro-Haul architecture

Figure 37 shows the main components of the secure SDN controlled video distribution Use Case. Within the Metro network, optical switches allow the creation of quantum-secured paths. These optical links will connect QKD resources as well as L2 encryptors/decryptors to enable the secure communication. In the AMEN, a HD video camera will be used for collection of video frames, which will subsequently be encapsulated before encryption. At the MCEN, after decryption, the video decapsulation is used to feed the video traffic towards the multiviewer.

As SDN components, SDN agents on the optical switches allow the communication with the SDN controller via using extended OpenFlow protocols. In addition, applications can be used over the SDN controller to enable the monitoring of quantum parameters such as the secret key rate, the QBER or the no generation of keys.

4.12.3 Service requirements

Table 30 shows the KPIs of the secure SDN video distribution Use Case. In this case, the requirements for the video traffic are merged with the requirements for the quantum parameters.

Table 30 – KPIs of Secure SDN Video Distribution Use Case.

1	**Bandwidth (data rate) [Mb/s]**	Very High (1500 Mbps per video feed at least) Very Low (0.0005 Mbps over 25km for Secret Key Rate in QKD)
2	**Latency [ms]**	Flexible (broadcaster defined for video) Unknown for QKD
3	**Jitter [ms]**	Unknown for QKD Stringent (broadcaster defined for Video)
4	**Reliability [%]**	high (five 9) for video Up to 11% QBER for QKD
5	**Availability [%]**	high (five 9)
6	**Mobility - UE speed [km/h]**	Fixed
7	**Area Traffic Capacity [Gb/s/km^2]**	High
8	**Rely on Sensor Network [Yes/No]**	No
9	**Massive Type [Yes/No]**	No
10	**Device Direct [Yes/No]**	No
11	**Coverage Requirement [Standard/Extended]**	Standard
12	**Energy Consumption critical [Yes/No]**	No
13	**Type of User Equipment (UE) [Conventional/Special purpose]**	Special purpose, SDI to L2/L3 Encoders/Decoders, ATEM controllers, QKD devices, Encryptors/Decryptors

4.12.4 Metro-Haul network requirements

A more detailed architecture of the optical network used in specific Use Case is described in Figure 35 (AMEN and MCEN remain the same).

For the Quantum Key Distribution (QKD) scheme, 3 optical links are suggested to create and distribute a link. One link will be the quantum channel, which will be a single fibre connecting the

QKD units (called Alice and Bob). This link cannot include any amplifiers in between[14], as this will highly affect the photon. Also the distance of this fibre should not exceed 50km and most importantly the losses of this link cannot be higher than 10dBs. The remaining two links comprise the classical channel (Tx/Rx) and connect the servers that the QKD units are attached to and are used for the communication of the QKD units after the measurement of the photons. After the key exchange between the two nodes, the transmission of the encrypted data happens through the encryptors and the data channels as depicted in the picture.

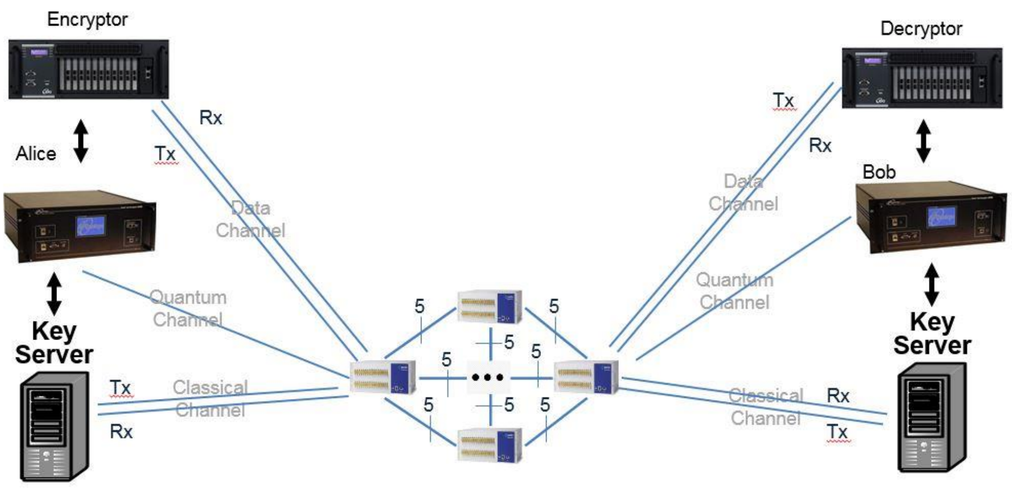


Table 31 – Impact of crucial tasks on Service KPIs in Secure SDN Video Distribution Use Case.

Service KPI	Value	Crucial Tasks requiring most stringent KPI value
Data rate	Secret Key Rate of 0.0005Mbps for over 25 km	**Quantum Key Distribution/Data encryption-decryption**
Latency	unknown	
Jitter	unknown	
Reliability	Up to 11% QBER	**Quantum Key Distribution**
Mobility (speed)	NA	
Area traffic capacity	NA	
Sensor density	NA	

[14] A quantum encoded photon will be affected by the Raman noise generated by the high power of optical signals used in non-quantum encoded signals. Similarly, the ASE noise generated by any EDFA will also affect the detection of the single encoded photon which is used for the key extraction. Therefore a specific link can not include any EDFA and that is one of the challenges in QKD networks. However, there are QKD units that allow transmissions and detections of encoded photons of up to 50km or 10dBs of link power budget (which are the ones that we use) without any amplification.

Table 32 – Summary of Metro-Haul requirements (KPIs) by Secure SDN Video Distribution Use Case.

Task	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
Quantum Key Distribution	0.001 Mbps		0.001 Mbps				Quantum Key Distribution
Data en/de-cryption	6 Gb/s per video feed	6 Gb/s per video feed				Dedicated hardware	Data en/de-cryption

Task	Monitoring and Data analytics	Management, Control and e2e Orchestration
Quantum Key Distribution	Monitoring of Secret Key Rate(SKR) and Quantum Bit Error Rate(QBER)	SDN Controller/Recovery Mechanisms
Data en/de-cryption	Monitoring of the Encryption/Decryption interfaces	

5 5G RN Transport Architectures

In traditional 4G (LTE) architectures, the Base-Band Unit (BBU) is co-located with the Remote Radio Head (RRH) in a Distributed Radio Access Network (D-RAN) architecture. A transport network would physically connect the BBU with the Evolved Packet Core (EPC) server in the Point-Of-Presence (PoP) where it would be connected to all other services. Although initially the BBU was located at the top of the RRH, next to the RF antenna, it was eventually moved to the Radio Cabinet at ground level (Figure 38 a). The connection between the RF antenna and the BBU was known as *front-haul* and a metal cable was used. CPRI is a protocol developed by an industrial group to carry traffic over the front-haul.

A centralised RAN architecture was later applied where the BBU was pulled back to a Central Office (or Local Exchange) and connected via point-to-point (PtP) fibre, in order to share the expensive BBU among several RRHs (Figure 38 b). Some of the functionalities in the BBU have very stringent latency requirements and hence the distance limit for the front-haul application is around 20km. The most latency-restrictive BBU functionality is known as HARQ loop (Hybrid Automatic Repeat Request) and schedules the transmissions from the End Users to the radio antenna through the air interface.

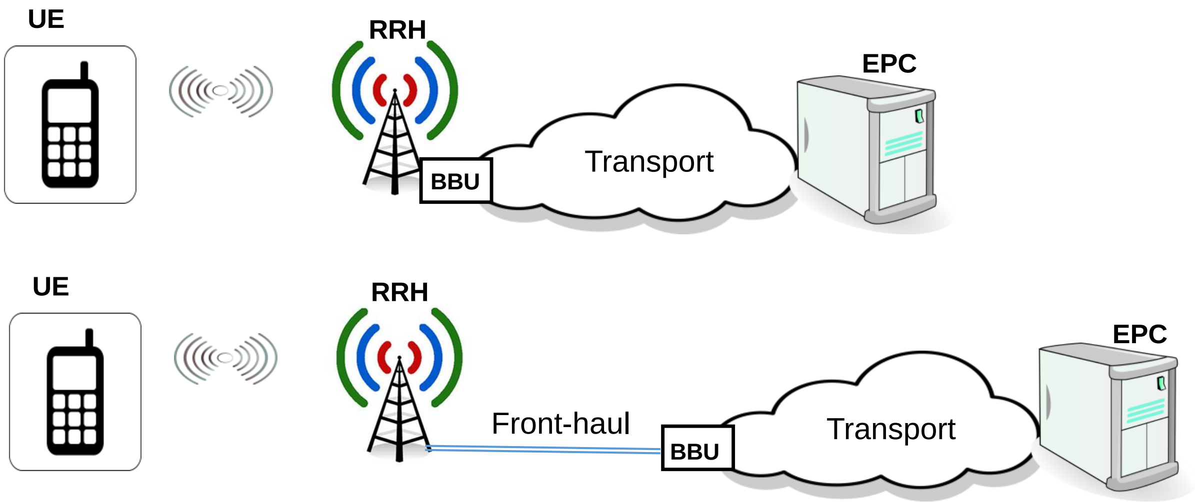


Figure 38 – RAN architectures: a) distributed (top), b) centralised (bottom)

The capacity needed to carry traffic through front-haul (using a CPRI) is relatively high and constant [53] because it is based on transport of digitized time domain IQ data. Such capacity requirement will grow even further, as 5G Radio Networks are expected to offer much higher user bandwidths by increasing the width of the radio channel, using higher order modulation, and increasing the number of antenna ports.

Table 33 shows the fronthaul capacities (CPRI rates without line coding) needed to support various radio channel bandwidths and numbers of antenna ports in a 5G wireless network using parameter ranges defined by 3GPP [55]. We see that the fronthaul capacities required to carry CPRI traffic become very high very fast, which makes it non-economical.

Table 33 – Required fronthaul bandwidth in 5G wireless network (CPRI rates without line coding) [55]

Number of Antenna Ports	Radio Channel Bandwidth			
	10 MHz	**20 MHz**	**200 MHz**	**1GHz**
2	1 Gb/s	2 Gb/s	20 Gb/s	100 Gb/s
8	4 Gb/s	8 Gb/s	80 Gb/s	400 Gb/s
64	32 Gb/s	64 Gb/s	640 Gb/s	3,200 Gb/s
256	128 Gb/s	256 Gb/s	2,560 Gb/s	12,800 Gb/s

5.1.1 Base Band Unit splitting options

As seen in Table 33, the choice of CPRI leads to high capacity and strict latency requirements for the fronthaul. More flexible distributions of the functions in the 3GPP LTE RAN protocol stack between RUs and DUs are being investigated. The assorted options of this flexible distribution are typically referred as RAN functional splits.

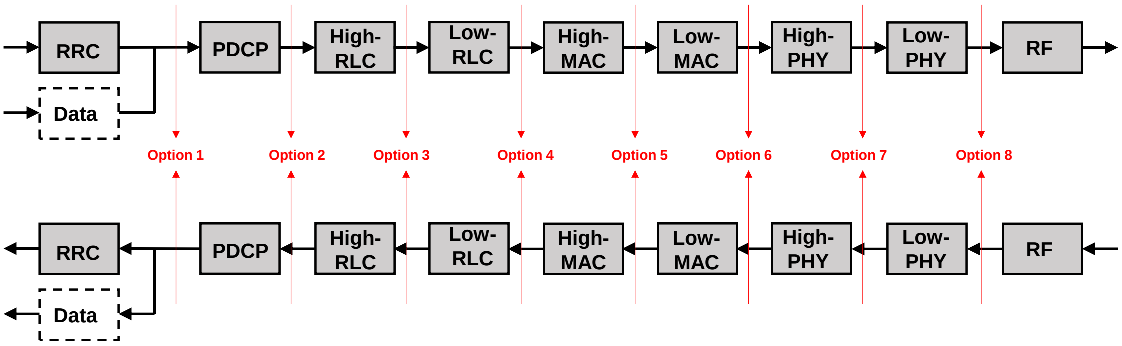


Figure 39 – BBU Split Options.

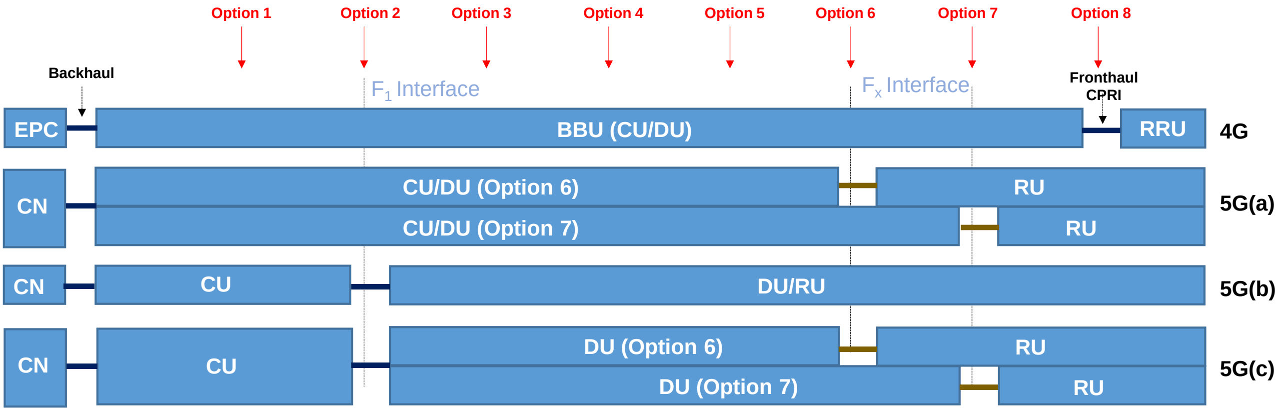


Figure 40 – Mapping of CU and DU functions according to the split points. 5G(a) low layer split (F1); 5G(b) high layer split (FX); 5G(c) cascaded split.

In principle, splits can be applied on any protocol layer or on the interfaces between layers. Hence, before introducing the different RAN functional splits, we overview the layers (and the main functions in each layer) of the 3GPP LTE RAN protocol stack.

- The physical (PHY) layer is responsible for preparing the bit stream for transmission by executing some baseband functionalities, such as, OFDM demodulation, MIMO precoding, Channel estimation, Fast Fourier Transform.
- Medium Access Control (MAC) layer is responsible for lower layer 2 functions as multiplexing, error correction through hybrid automatic repeat request (HARQ) and other scheduling and prioritization functions.
- Radio Link Control (RLC) layer implements upper layer 2 functions as concatenation, segmentation and reassembly of data units. Also, it implements the time-domain estimation/compensation of non-idealities which occurs due to carrier frequency offset (CFO) and sampling frequency offset (SFO).
- Packet Data Convergence Protocol (PDCP) layer performs ciphering, integrity protection, IP header compression.
- Radio Resource Control (RRC) protocol is used in LTE on the air interface. It is a layer that exists between UE and eNB, working at the IP level, and managing, among others mobility QoS management functions;

As indicated in Figure 39, multiple options for splitting processing functions between the DU and RU are possible:

Option 1 (RRC-PDCP split): Centralization of RRC provides important benefits in terms of allowing central execution of QoS and mobility management.

Option 2 (PDCP-RLC split): In this option, PDCP functions such as data packets header compression, ciphering, integrity remain centralized. The centralized functions are not sensitive to latency (30ms) since all the scheduling functions are distributed at the cell site.

Option 3 (RLC split): This split option divides the RLC into two parts and centralize the upper RLC, PDCP and RRC layers.

Option 4 (RLC-MAC split): Here the RLC layer functions are moved to the DUs. This increases the functionalities which can be centralized including load balancing, but it does introduce complexity as the downlink RLC layer is tightly coupled to both the MAC and scheduler.

Option 5 (Split MAC): This split option separates between the higher and lower MAC layer. Here the majority of the MAC layer is centralized at the DU, but the hybrid automatic repeat request (HARQ) scheduling remains at the RU**.**

Option 6 (MAC-PHY split): This RAN architecture splits between the physical and the MAC layer. PHY layer functions are distributed, while MAC, RLC and PDCP functions remain centralized at the DUs. It imposes a 2ms latency requirement.

Option 7 (PHY split): Here the physical layer functions are split into two parts, lower and upper physical layer. For this split, the latency requirement is 2ms. By applying this split option, we enable the centralization of the upper PHY, MAC, RLC and PDCP functions. While the lower PHY functions such as filtering, sampling, FFT/IFFT, resource mapping/de-mapping, and channel estimation as well as RF processing, A/D conversions and power pre-processing are distributed.

Option 8 (CPRI split): All baseband functionalities (L1, L2, L3) are located at the BBU pool, meanwhile only power amplification and radio frequency processing remain decentralized at the cell site. This architecture has all the advantages of C-RAN as it enables the maximum multiplexing gain and reduce the complexity at RRH, in terms of digital signal processing hardware. On the other hand, this architecture should meet strict latency requirements (in the order of 0.25ms) for physical layer processing.

5.1.2 Traffic capacity & latency depending on splitting options

This section summarizes the latency and bandwidth requirements for the various split options, providing an overall comparison between them.
Table 34 shows the different downlink/uplink bandwidth requirements for each split option. It shows that the splits up to the MAC-PHY (i.e., options 1 to 6) have similar bandwidth requirements, while split options at the physical layer (i.e., from option 7 onwards) the required bandwidth increases significantly with functions centralization.

Table 34 – Transport bit rates required at the interfaces for different functional splits [55].

Protocol Split option	Required downlink bandwidth	Required uplink bandwidth	Comment
Option 1	4 Gb/s	3 Gb/s	Service (L3) required bandwidth

Option 2	***4016 Mb/s***	***3024 Mb/s***	[16 Mb/s for DL and 24 Mb/s for UL is assumed as signalling]
Option 3	[lower than Option 2 for UL/DL]		
Option 4	4000 Mb/s	3000 Mb/s	
Option 5	4000 Mb/s	3000 Mb/s	
Option 6	***4133 Mb/s***	***5640 Mb/s***	[133 Mb/s for DL is assumed as scheduling/ control signalling. 2640 Mb/s for UL is assumed as UL-PHY response to schedule]
Option 7a	***10.1~22.2 Gb/s***	***16.6~21.6 Gb/s***	[713.9 Mb/s for DL and 120 Mb/s for UL is assumed as MAC information]
Option 7b	***37.8~86.1 Gb/s***	***53.8~86.1 Gb/s***	[121 Mb/s for DL and 80 Mb/s for UL is assumed as MAC information]
Option 7c	***10.1~22.2 Gb/s***	***53.8~86.1 Gb/s***	
Option 8	157.3 Gb/s	157.3 Gb/s	

Table 35 – Assumptions for required bandwidth in Table 34.

Items	Assumption	Applicability
Radio Channel Bandwidth	[100MHz(DL/UL)]	All options
Modulation	[256QAM(DL/UL)]	
Number of MIMO layers	[8(DL/UL)]	
IQ bit-width	[2*(7~16)bit(DL), 2*(10~16)bit(UL)]	Options 7a, 7b, 7c
	[2*16bit(DL/UL)]	Option 8
Number of antenna port	[32(DL/UL)]	Option 7b Option 7c(UL) Option 8

The latency requirements vary for each split option, and whether it includes the HARQ loop function: While options 4 – 8 centralise this function and therefore require latency values in the order of a few hundred microseconds, latency requirements for splitting options 1 -3 are relaxed to a range of several milliseconds.

From these latency requirements follows the adequacy to either “non-real-time” or “real-time” transport type of services. It also follows the possible 5G architecture as shown below.

5.1.3 Potential network architectures as a function of splitting options

The use of different BBU splitting options following 5G(a), 5G(b), and 5G(c) as shown in Figure 40 renders itself to different network architectures depending on the location of the different boxes, i.e. the RU, DU and the CU.

a) Network architecture using low-layer split option

When using a low-layer split option (5G(a) in Figure 40), the stringent latency requirements forces the DU/CU to be located not further than 20km, which normally corresponds to the local exchange (AMEN in Metro-Haul). In this case the DU/CU box will be co-located with the access head-end, i.e. the OLT. A backhaul will be used through the Metro network to connect the BBU to the Network Core.

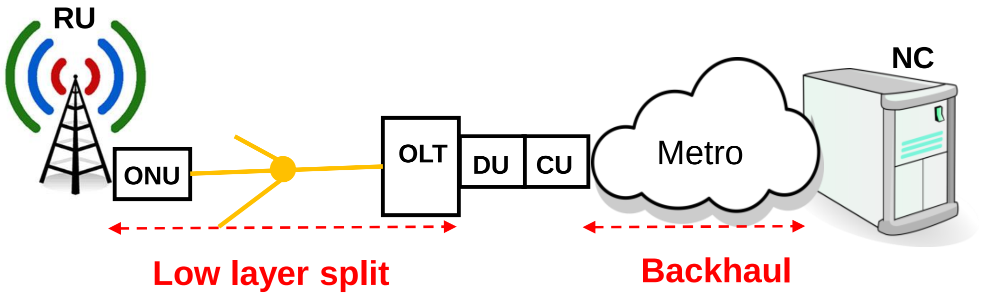


Figure 41 – Network architecture using a Low Layer Split option (Scenario 5G(a)).

b) Network architecture using high-layer split option

When using a high layer split option (5G(b) in Figure 40) the latency requirement is now in the order of a few milliseconds and therefore the CU box can be collocated in a more central location. Depending on the exact latency requirement, and also on the fibre distances involved in the metro network, the CU box can be located either in the Local Exchange (AMEN in Metro-Haul, see Figure 42, top) or somewhere deeper in the Metro Network (e.g. a MCEN, see Figure 42, bottom).

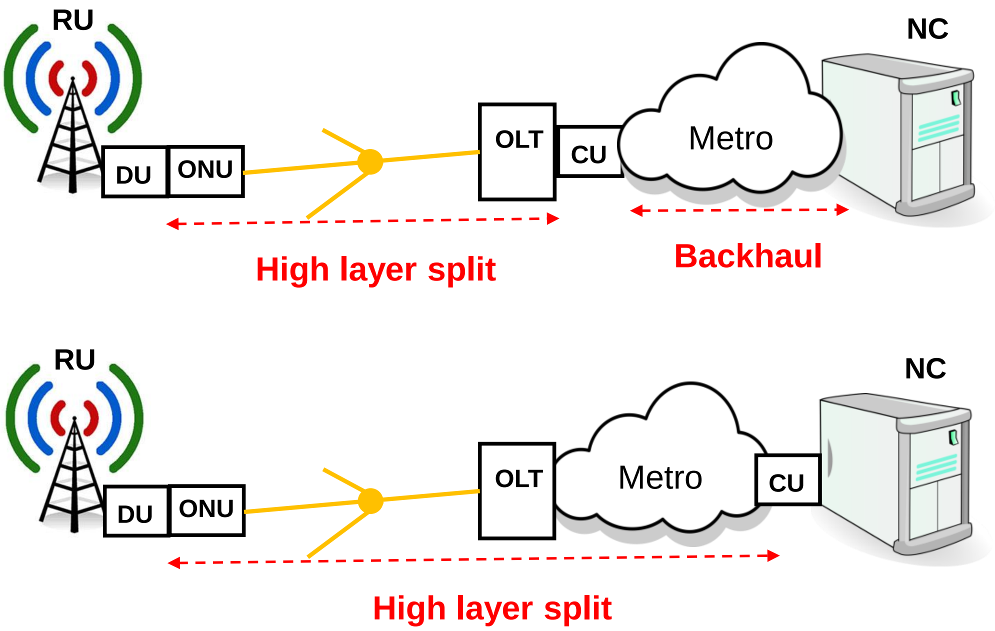


Figure 42 – Network architecture using a High Layer Split option (Scenario 5G(b)); two possibilities exist: CU collocated with OLT (top figure), or CU deeper into the network (bottom figure).

c) Network architecture using cascaded split option

When using a cascaded split option (5G(c) in Figure 40), the stringent latency requirement still applies between RU and DU, and therefore this latter box must be located not further than 20km, which is consistent with network architecture 5G(a). The location of the CU box depends on the exact latency requirement, but similar to 5G(b) it could be located in the Metro Network (e.g. a MCEN, see Fig. 23).

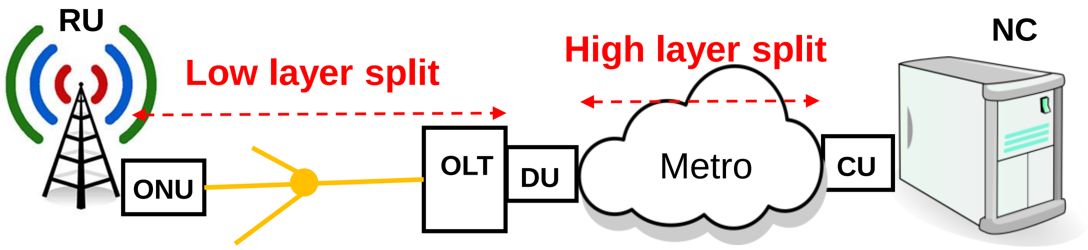


Figure 43 – Network architecture using a Cascaded Split option with DU collocated with Head-End and CU deeper in the network (Scenario 5G(c)).

5.1.4 Impact of 5G network architectures on Metro-Haul

To derive the impact of the 5G network architectures onto the Metro-Haul architecture we need to make some assumptions. These assumptions are collected in Table 36.

Table 36 – Assumed max latency values (one-way).

	Max. (one-way) latency	**Max. Fibre Distance**	**Max. Metro Fibre Distance**
RT services	100 µs	20 km	n/a
NRT services	1.5 ms – 10 ms	300 km – 2,000 km	280 km – 1,980 km

Table 37 – Location of functionality for the 5G network architectures.

Split Option	**DU Functionality**	**CU Functionality**	**NC Functionality**
Low Layer	RU	AMEN	MCEN
High Layer	RU	AMEN or MCEN	MCEN
Cascaded	AMEN	MCEN	MCEN

The conclusions that can be drawn from these assumed latency values, concerning the fibre distances in the Metro-Haul network, are:

- In the case of using low layer split options (suitable for RT services), the Metro-Haul network cannot be used for fronthaul, but only for backhauling traffic between the DU/CU in AMEN, and the NC deeper in the network such as the MCEN. The maximum fibre distance between the DU/CU and the NC will depend on the maximum service latency requirement.
- In the case of using a high layer split option, we have the option of locating the CU functionality either in the AMEN, or in the MCEN. If in the MCEN, the maximum fibre distance between the DU and the CU, and between the CU and the NC, will depend on the maximum service latency requirement.

- In the case of a cascaded split option, the maximum fibre distance between the DU and the CU, and between the CU and the NC, will depend on the maximum service latency requirement.

5.1.5 Evaluating processing requirements for different 5G RAN functional splits

Among the objectives of WP2 in Metro-Haul, we plan to provide input to other WPs regarding the amount of processing capacity needed to support various split options in the AMENs and/or MCENs. This subsection introduces the approach used to estimate the amount of processing operations in the CU, DU and RU required to serve a given area considering the four following split options: CPRI split, PHY split, PHY-MAC split and RLC-PDCP Split, corresponding to BBU Split Options 8, 7, 6 and 2 of Figure 39, respectively. To estimate the CU/DUs computational effort, we first classify the digital baseband functionalities in the CU/DU and the RU for each spliting option (see Figure 44).

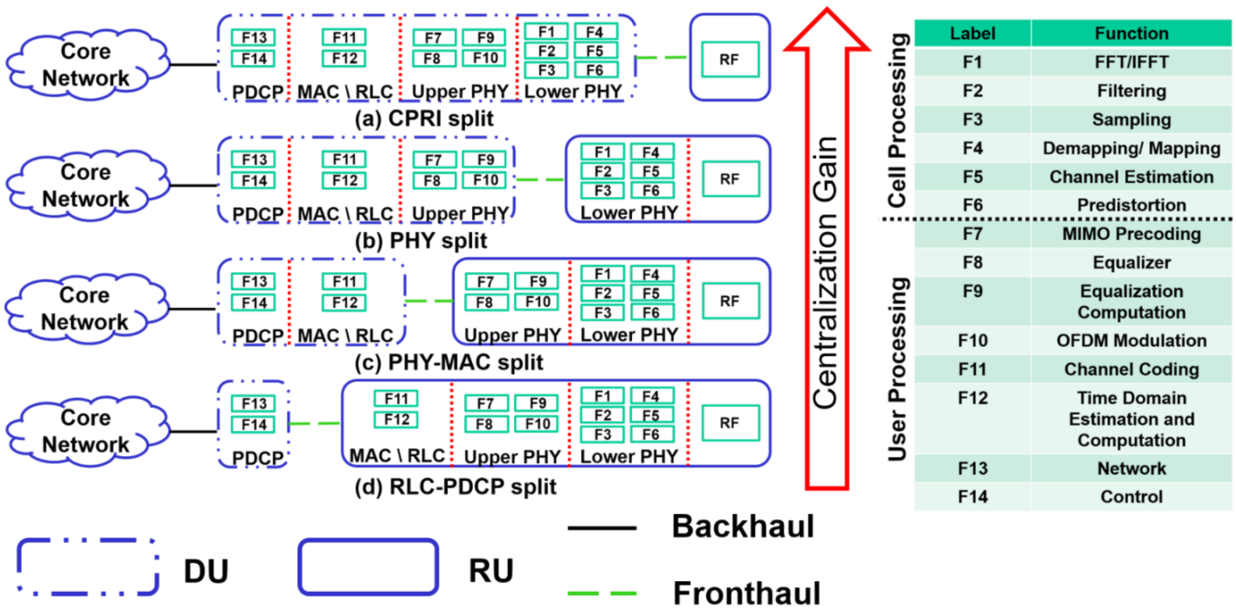


Figure 44 –RAN Split options and corresponding mapping of network functions.

We employed an analytical model which takes into consideration different tunable input parameters, such as the spatial distribution of the users in serving area, the association strategy between the users and eNBs and physical resource blocks (PRBs), the scheduling algorithm among the active users, as we explain more in detail below. A synthetic flow-chart of the process carried out to perform the evaluation is shown in Figure 45. First, we choose a statistical distribution (namely, normal or uniform) to model the spatial distribution of mobile users in a given serving area. Then we allocate the users to their serving eNBs according to a realistic user-eNB association strategy, i.e. maximum power association strategy. After that, we apply a scheduling algorithm to distribute the physical resource blocks of the eNB among its users. Finally, we calculate the computational effort per user for a given channel condition, used resources, modulation, code rate and MIMO mode. The computational effort in a given time interval t for user i can be defined [58] as:

$$CE(i,t) = \left(3A_{i,t} + A_{i,t}^2 + \frac{1}{3}M_{i,t}C_{i,t}L_{i,t}\right)\frac{R_{i,t}}{10} \quad (1)$$

where A is the number of used antennas, M is the modulation bits, C is the code rate, L is the number of spatial MIMO-layers, and R is the number of PRBs. *CE* is expressed in Giga Operations per Second (GOPS) unit.

The data processing rate of semiconductor chips is represented in Millions of Instructions Per Second (MIPS). Thus, mapping the obtained GPOS to several CPUs according to the CPU processing capabilities, requires converting GOPS to MIPS as follows [59].

$$MIPS = GOPS \times \frac{10^3}{64} \quad (2)$$

We consider Intel E7-X870 CPU with 2.4 GHz and 10 cores per CPU and has processing capabilities 96.90k MIPS [60], to estimate the amount of processors needed to serve a given serving area

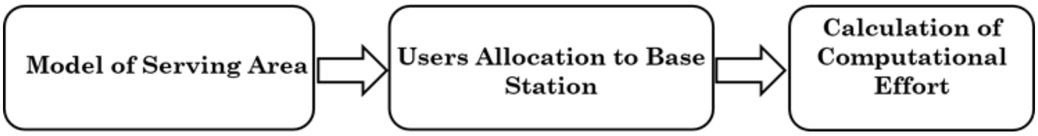


Figure 45 –Model steps to calculate the computational effort.

To obtain numerical results, we consider two different geographical type areas (geotypes), namely, Dense-urban and Urban. We consider a serving area with 12 cell sites where the number of cell sites per unit area, and number of users per unit area are counted as in [61] and [62], respectively, and reported in Table 38. We compare the GOPS needed in the four considered split options to the GOPS needed in the D-RAN to estimate the processing reduction in each case.

Table 38 – Features of the considered geotypes.

	Dense Urban	Urban
Number of sites per Km^2	4	1.5
Total area to accommodate 12 site [Km^2]	3	8
Number of users per Km^2	3000	1000
Total number of users in 12 site	9000	8000

The results are shown in Table 39. As expected CPRI split option could achieve the highest reduction in processing requirements with respect to D-RAN for all different geotypes, while it is reduced gradually when centralizing fewer functions. In split PHY, the needed functionalities for the user processing are all implemented in the CU/DUs. The results confirm the fact that as we centralize less, the processing requirements in terms of GOPS increase.

Table 39 – GOPS for different split options

	Dense Urban					**Urban**				
	RU side		**DU side**		**Total GOPS**	**RU side**		**DU side**		**Total GOPS**
	GOPS	**# of Cores**	**GOPS**	**# of Cores**		**GOPS**	**# of Cores**	**GOPS**	**# of Cores**	
CPRI	0	0	28,272	45	28,272	0	0	28,668	46	28,668
PHY	12,936	21	17,076	28	30,012	13,452	22	17,160	28	30,612
PHY-MAC	22,092	36	9,048	15	31,140	22,260	36	9,180	15	31,440
RLC-PDCP	28,680	47	3,108	6	31,788	28,932	47	3,144	5	32,076
D-RAN	32,484	53	0	0	32,484	32,916	53	0	0	32,916

From Table 38 it is possible to derive the computational effort for each different architecture possibility that has been described in Section 5.1.3. This is shown in the following tables.

Table 40 – Processing requirements when using low-layer split option 7 (Scenario 5G(a))

	RU		CU/DU [AMEN]	
	GOPS	#Cores	GOPS	#Cores
Dense Urban	12,936	21	17,076	28
Urban	13,452	22	17,160	28

Table 41 – Processing requirements when using low-layer split option 6 (Scenario 5G(a))

	RU		CU/DU [AMEN]	
	GOPS	#Cores	GOPS	#Cores
Dense Urban	22,092	36	9,048	15
Urban	22,260	36	9,180	15

Table 42 – Processing requirements when using high-layer split option 2 (Scenario 5G(b))

	DU/RU		CU [AMEN or MCEN]	
	GOPS	#Cores	GOPS	#Cores
Dense Urban	28,680	47	3,108	6
Urban	28,932	47	3,144	5

Table 43 – Processing requirements when using cascaded split options 2 & 7 (Scenario 5G(c))

	RU		DU [AMEN]		CU [MCEN]	
	GOPS	#Cores	GOPS	#Cores	GOPS	#Cores
Dense Urban	12,936	21	15,744	26	3,108	6
Urban	13,452	22	15,228	25	3,144	5

Table 44 – Processing requirements when using cascaded split options 2 & 6 (Scenario 5G(c))

	RU		DU [AMEN]		CU [MCEN]	
	GOPS	#Cores	GOPS	#Cores	GOPS	#Cores
Dense Urban	22,092	36	6,588	11	3,108	6
Urban	22,260	36	6,672	11	3,144	5

6 Summary and Conclusions

Three complementary approaches have been used to understand the requirements that the Metro-Haul architecture needs to meet in order to support future vertical services over a 5G RN: a) a high level approach, b) an analysis of end user (vertical) Use Cases, and c) the analysis of impact of functional split option of 5G RN transport over the Metro-Haul architecture.

6.1 High level approach.

The data sources considered and the analysis performed in the high level (top-down) approach shows that the AMEN nodes are expected to be served by pipes of 10-100 Gb/s, e.g. 10 Gb/s on the access interfacing side, and aggregated pipes of up to 100 Gb/s on the Metro interface side. The AMEN DC would have storage capability measured in TB of RAM and HD capacity, with processing power estimated of the order of 1 TFLOPS (e.g. we reference the Intel Core i9 Extreme Edition chip, which consists of 18 vCPUs or multi-cores on a single chip, and exhibits teraflop capability).

On the MCEN side, interfaces with the photonic core can be expected to require a factor x10 increase as compared with the AMEN. MCEN will be served by 10-100 Gb/s pipes from other AMEN and intermediate Metro-Haul nodes, and interface via 400G/s and 1 Tb/s pipes into the photonic core network. The storage capacity of the MCEN will be measured in the 10s of TB of RAM and HD, with associated compute resource in the order of 10 TFLOPS.

Concerning the average node degree sizes for the MCEN and AMEN, these can be expected to remain similar to the one characterizing current network topologies, since intrinsic connectivity and topological (geographic) parameters are not expected to appreciably change in the considered timeframe (i.e. the first period of roll out of the 5G network, which is the target timeframe of Metro-Haul).

6.2 Summary of results of the approach based on Use Cases

The results of the approach based on Use Cases to derive Metro-Haul requirements are summarized in the final table of every subsection 4.X.4, (X=2…12). Table 45 combines the input of all those tables to give an overall view of network KPIs required by Metro Haul. As in corresponding tables of section 4, network KPIs assessed in Table 45 are the Throughput, the Storage capacity and the Computing capacity in the Metro-Haul Nodes. Concerning the Throughput, the expected size and number of optical interfaces are also considered. Please note that, in compiling Table 45, for each Use Case, the KPI value takes into account all the tasks that characterize the Use Case, whose details are provided in the tables of section 4. From the requirements in Table 39, a quite heterogeneous set emerges, with some Use Cases that dominate the demand of bandwidth (the ones belonging to Media & Entertainment which require hundreds of Gb/s in the AMEN or MCEN) while others dominate the compute capacity (hundreds of vCPUs, under our service-penetration hypotheses), mainly, the Smart factory, the Automotive and the Service robotics. 6DoF Virtual reality, Content delivery (these two again in the Media & Entertainment vertical) and Smart factory are the ones that potentially require big Storage memory capacity (around 100 TB) while Use Case Real-Time Low-Latency Object Tracking and Security overcomes all other Use Cases considered of more than one order of magnitude as storage capacity requirement in the AMEN (4000 TB).

Network functionalities are summarized in Table 46, which shows that Monitoring and Data Analytics are common requirements across Use Cases for the monitoring of transmission for out of service detection or degradation anticipation, and VM monitoring. On the Management, Control and e2e Orchestration functionalities, Use Cases show the need for dynamic dimensioning of local and global resources, and reconfiguration involving both connectivity and memory.

Table 45 – Summary of Metro-Haul capacity KPIs from Use Case services requirements

Use Case	Throughput			Storage		Computing capacity	
	AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
mIoT Utility metering	10/20 Gb/s		10 Gb/s	2 GB		2 vCPU	
Content Delivery Network	72 Gb/s	72 Gb/s[15]	Nx10 Gb/s	22.5 TB	11.5 TB	4 vCPU	5 vCPU
Live TV distribution	600 Gb/s	600 Gb/s	Nx100 Gb/s (N>=6)			400 vCPU (50 VM)	
Service Robotics	10-100 Gb/s		10-100 Gb/s	10 TB		200 vCPU	
Smart Factories	50 Gb/s		Nx10 Gb or 100 Gb/s	50 TB		750 vCPU	
6DoF Virtual Reality	70-250 Gb/s	100 Gb/s[16]	100 Gb/s	10 TB	100 TB	20 vCPU	50 vCPU
Intelligent Transport System and Autonomous Driving	10 Gb/s		100 Gb/s	10 TB		100 vCPu[17]	
Enterprise Access with NG Ethernet	100 Gb/s	100 Gb/s	100 Gb/s				
Crowdsourced video broadcasts	10 Gb/s	100 Gb/s	Nx100 Gb/s	100 GB		20 vCPU[18]	

[15] CDN Use Case - According to the two levels hierarchical model used for content delivery. Per 10,000 users per AMEN and 6 AMENs per MCEN.
[16] In case MCEN is the host of original source.
[17] Under the hypothesis of a total of 1,000 sensors under the control of a single AMEN node.
[18] Crowdsourced video broadcasts/Bristol Demo - 1 vCPU per 100 video streams: 1000 video stream supposed plus some VNFs.

Real-Time Low-Latency Object Tracking and Security	25 Gb/s.	2 Gb/s	25 Gb/s	4,000 TB		200 vCPU[19]	200 vCPU
Secure SDN Controlled Video Distribution	150 Gb/s [20]	150 Gb/s	Nx100 Gb/s			n/a[21]	n/a[22]

Table 46 – Summary of Metro-Haul functional KPIs from Use Case services requirements.

Use Case	Monitoring and Data analytics	Management, Control and e2e Orchestration
mIoT Utility metering	Transmission monitoring (for out of service or degradation detection/anticipation) VM monitoring	Fast recovery mechanism Optimal Planning of capacity resources VM dynamic dimensioning (local reconfiguration)
Content Delivery Network	Traffic monitoring for early detection of new popular videos Traffic monitoring for fast detection of a flash crowd phenomenon	Fast recovery mechanisms (protection and restoration) Local and Global reconfiguration of virtual cache Local and global reconfiguration of virtual cache
Live TV distribution	Quality of Transmission monitoring and degradation detection/anticipation Applications monitoring (user activity, video quality) VM monitoring	Recovery mechanism Local and Global reconfiguration of virtual cache VM dynamic dimensioning (local)
Service Robotics	Transmission monitoring for degradation detection/anticipation VM monitoring	VM dynamic dimensioning (local) Global connectivity Reconfiguration

[19] Crowdsourced video broadcasts/Bristol Demo - 100 channels and one algorithm per channel are assumed.
[20] Secure SDN Controlled Video Distribution -25 video feeds assumed.
[21] Dedicated Hardware, not quantifiable in terms of vCPU.
[22] Data encryption and decryption, not quantifiable in terms of vCPU.

Smart Factories	Transmission monitoring (for out of service or degradation detection/anticipation) VM monitoring	Very fast recovery mechanism Fast restoration VM dynamic dimensioning (local reconfiguration) Global connectivity reconfiguration
6DoF Virtual Reality	Transmission monitoring (for out of service or degradation detection/anticipation) VM monitoring	Fast restoration VM dynamic dimensioning (local reconfiguration) Global connectivity reconfiguration
Intelligent Transport System and Autonomous Driving	Transmission out of service detection and degradation detection/anticipation	Fast-restoration mechanisms
Enterprise Access with NG Ethernet	Active measurements: Capacity, delay, packet loss (reliability) and availability. Passive measurements: Time-series and Netflow records	Recovery mechanism Global reconfiguration
Crowdsourced video broadcasts	Jitter and Latency monitoring Transmission degradation detection/anticipation VM monitoring	Scalable traffic paths Improve quality of service Dynamically allocate paths between domains Optimal Planning of capacity resources Dynamically scale computing resources based on VNFs Scale computing resources based on streams
Real-Time Low-Latency Object Tracking and security	CCTV out of service detection Object tracking VM monitoring	Possible recovery mechanism Control auto-tracking of PTZ cameras VM dynamic dimensioning (local reconfiguration)
Secure SDN Controlled Video Distribution	Monitoring of Secret Key Rate(SKR) and Quantum Bit Error Rate(QBER) Monitoring of the Encryption/Decryption interfaces	SDN Controller/Recovery Mechanisms

The next two tables (Table 47 and Table 48) report the Metro-Haul KPIs derived from the analysis of the Service Use Cases. It is assumed that all service Use Case are active simultaneously.

Table 47 – Metro-Haul requirements resulting from all Use Cases considered.

Throughput			Storage		Computing capacity	
AMEN	*MCEN*	*Optical*	*AMEN*	*MCEN*	*AMEN*	*MCEN*
1.1 / 1.4 Tb/s	1.2 Tb/s	Nx100 Gb/s +Mx10 Gb/s	4,200 TB	100 TB	1700 vCPU	250 vCPU

Table 47 shows that an AMEN would need a switching throughput of around 1 Terabit per second (only for supporting the service Use Cases, and not considering the need of switching capacity for transmission purposes between AMEN), storage capacity of 4.2 Peta bytes, and a computing power of 1,700 vCPU. It is important to highlight that these requirements come from very specific service Use Cases. For example, the storage & computing capacities are requirements from the "Real-Time Low-Latency Object Tracking and Security" service Use Case when the AMEN is controlling 1,000 CCTV cameras. If the AMEN does not need to support this service Use Case, the necessary amount of storage and computing capacities are drastically reduced. The requirement values reported for the MCEN in this table are much lower than for the AMEN. The reason for this is because the MCEN does not actively participate in supporting the service as much as the AMEN. Also, some analysis have not considered the number of AMEN that need to be supported for each MCEN. Finally, the MCEN plays a role also in other network related services that have not been analysed in this work.

In terms of the optical metro network, we see that while it will be necessary to use 100 Gb/s connections between AMEN and MCEN. The number of AMEN times the necessary throughput yields a respectable high capacity figure in the order of ~15 Tb/s if 10 AMEN are assumed to be connected to an MCEN, with video services requiring 10 Gb/s pipes or more. This leads to the requirement of using 100 Gb/s pipes between the MCEN and the AMENs. We can deduct from this requirement that having a multicast functionality in the optical domain can substantially lower the need of packet switching capacity in the higher layers.

Table 48 – Metro-Haul functionalities resulting from all Use Cases considered.

Monitoring and Data analytics	Management, Control and e2e Orchestration
Transmission monitoring: for out of service for degradation detection/anticipation Packet level measurements: Active measurements: capacity, delay, latency, jitter, packet loss (reliability) and availability. Passive measurements: Timeseries and Netflow records Detection of new popular videos (for CDN) Applications monitoring (user activity, video quality for Live TV) VM monitoring Object tracking (for)	Optimal capacity planning Global connectivity reconfiguration *(including dynamic allocation of paths between domains)* Fast recovery mechanisms *(protection and restoration)* Local and Global reconfiguration of virtual cache VM dynamic dimensioning (local reconfiguration) *(also called: dynamic scale computing resources based on VNFs)*

According to Table 48 in all the Use Cases the basic function required is the collection of data from data-plane components (both telecom equipment and computing elements). The collected data is then fed to the control plane (SDN controllers, service orchestrator, OSS, planning tools, etc.), where they it is used as input to various decision-taking algorithms. Monitoring data-sets collected from the data plane by the monitoring system, are not necessarily delivered to the control plane as they are. Indeed, in most cases the monitoring and data analytics system will process the data sets and deliver them in form of data modes, according to requests received by the control-plane components. It will be however possible, in some specific cases, and for specific observation time-windows, to request raw monitoring data sets.

The four services identified in the table differ in terms of layer of the data plane that generates the data and in terms of performance required to the monitoring and data-analytics system.

More specifically:

- Transmission monitoring data are generated at the physical (and in particular, optical) layer. Sampling rate need to be high and thus storage space necessary to save the data sets will be large, too. Most probably, a high compression factor is needed to transform these data sets before feeding them to the control plane
- Packet level data can be generated at various layers (e.g. IP, MPLS, Ethernet, OTN): these datasets are related to monitoring and measurement of traffic. The requirements in terms of sampling rate are more relaxed than for transmission data
- The CDN monitoring data comes from the application layer. The time-scale of this monitoring information is much larger than the previous cases. Given the very particular nature of this service, data sets in this use-case may come directly from the video servers: the precise implementation will be defined in the course of the project
- Data sets regarding virtual-machine status are collected from the computing elements. Most probably, this monitoring information will be consumed by the service orchestrator.

Regarding the second column in Table 48 (“Management, Control and e2e Orchestration”), it shows a summary of the types of network functionality requested by the service Use Cases to the control and the management planes, and also to the planning tools (see Section 1.1 for a presentation of these components).

We can identify the following network functions, listed in order of requested rapidity of intervention (from the fastest to the slowest):

- Fast recovery mechanisms (protection and restoration): a network function requested upon detection of a failure. The monitoring and data-analytics system, in some cases, can predict a failure is approaching given certain trends detected in monitoring data, thus anticipating the failure and making the recovery system autonomic. The speed of intervention is critical to ensure a high availability
- Optimal capacity planning: this network function request is generated whenever a new service is requested by a user or when the network operator needs to instantiate new traffic relations. For most use-cases reported in this deliverable, there are no severe constraints on the planning time. According to the type of service requested by the user, planning can involve the allocation of several types of resources.

- Global connectivity reconfiguration (including dynamic allocation of paths between domains): the network function requests are sequentially satisfied in the order they arrive to the system, each time triggering the optimal capacity planning for each individual request.
- Local and global reconfiguration of virtual cache: this network function request can be provided by adopting the same approaches as described in the two previous points and applying them to the specific context of CDN and cache management

The values shown in Table 47 are the most demanding in terms of performance across all analysed service Use Cases. However, there are other features that need to be specified for each use-case, that do not necessarily determine a demanding performance target to achieve, but that are nevertheless necessary for the control / management plane and the planning tools in order to be able to support each Use Case.

Some of these other features, assuming that they characterize the slices, and that a specific network slice is defined for each use-case instance:

- Typical duration of the requested service
- Typical rate of arrival of service requests
- Typical topology of the slice (e.g. point-to-point, full mesh, hub and spoke, etc.)
- Typical service chain requested

Especially for the global reconfiguration actions, it is important that the planning tool of MH is able to map all the service requirements onto a set of features that is common to all the use-cases. In this way, the slices can be optimized on the ground of comparable requirements.

Finally, it is very difficult at this stage to comment about the reduction on energy consumption required by general Metro-Haul KPIs (ten times less energy consumption). From Use Cases involving a massive number of devices or sensors (e.g. mIoT of Subsection 4.2) the energy saving must be pursued by the efficiency of such devices and sensors. In other words the solution must be sought in the technology underpinning the service. Regarding the metro network infrastructure, which is the focus of Metro-Haul, the energy saving is entrusted on one side in the development of lower consuming equipment (this is a technological challenge) and on the other side fostering the optical bypass everywhere where it is possible thus eliminating unnecessary switching and processing of traffic at electrical level (while this is a networking issue that can be better supported using and SDN/orchestrator). Next steps of the project will consider energy saving KPIs when developing solutions for the network architecture, i.e. the routing and resource allocation algorithms (including resilience aspects) and the planning strategies.

6.3 Summary of results of RU split option analysis

As reported in subsection 5.1.4, the following man indications emerge from the analysis of the impact of 5G-network functional-split options on Metro-Haul.

In case of low layer split (CPRI) the Metro-Haul network is not intended to be used for fronthaul, but only for backhauling traffic between the DU/CU placed in AMEN and the NC possibly placed in MCEN or outside of the metro network in the core backbone. Instead, in the case of using higher-layer split options, the CU functionality can be either located in the AMEN or in the MCEN (NC is as in the low-layer split option). In case of cascade split option DU is located in AMEN, while both CU and NC are located in MCEN (or outside the Metro).

Fibre maximum allowed distances on various segments depend on the latency requirement one want to guarantee to the services.

Concerning the evaluation of processing requirements at RU and DU sides for different split option, the results of an accurate dimensioning study are reported in Table 39. As a general indications it emerges that, as expected, low-layer splits achieve the higher reductions in processing requirements with respect to higher layer splits, while it is reduced gradually when centralizing fewer functions in DUs. The results confirm the fact that as we centralize less, the processing requirements in terms of processing increase and provide useful input for dimensioning the compute capacity of AMEN and MCEN in support to different functional splits.

6.4 Overall conclusions

In addition to the definition and description of a set of relevant set of 5G service Use Cases, which is extensively covered by Section 4, this Deliverable gives the first indications on how to build a metro network, and in particular a network adopting the Metro-Haul architecture.

It is important to note that results obtained from the analysis reported in this document give only a rough and not comprehensive indication of the number and level of resources that would be necessary on Central Offices of the metro network, and on links between them. A More accurate quantitative evaluation on network resources is the objective of planning, dimensioning and techno economic activities that will start in WP2 in the second year of the project (task T2.3).

The analyses of Use cases has also led to identify, at a high level, the main functionalities required by the network in terms of Monitoring, Control and Orchestration, and also for specific Data Analytics applications.

A committed task of WP2 (T2.1) is already started and it is working on functional network architecture definition, assuming as the reference input both the computation-storage-networking resources and the set of functionalities identified in this deliverable.

Many use cases analysed in this work are mainly served from the AMEN and this is due to latency requirements (Smart factories or 6DoF VR) or to bring as close as possible the contents to the end users (CDN) to reduce the traffic to be carried in the network.

For this reason the AMEN results, in the performed analysis, in number provided and features collected, with requirements higher than the ones of MCEN node. This, in a more accurate analysis, is not completely correct for many reasons.

Firstly, in some Use Cases, the impact on MCEN is evaluated taking into account, in the service completion, the requirement coming from a single AMEN node. Indeed, in metro network, many AMENs are associated to a single MCEN (a ratio of 6:1 to 10:1 could be realistic) and this multiplies the load to be carried by a MCEN.

Secondly, a Central office where an MCEN is present hosts, with a high probability, also the functionalities of an AMEN. This the requirements of the two nodes, overall, should be added, at least if we consider the requirement of the site as a whole.

Finally MCEN has to assure the interworking within the metro (which depends on the topology), and the traffic exchange of the Metro area with the Core. A significant amount of traffic, not considered in Use Cases presented in this deliverable, transits through the MCEN nodes, and this require additional resources.

If requirements of an AMEN, in terms of throughput and number of optical higher rate interfaces have been evaluated in 1.1-1.4 Tb/s and 15x100 Gb/s respectively, the requirements for an MCEN, which include AMEN capabilities (i.e., connection with the access), could be significantly higher and estimated in the order of 5 Tb/s and 60x100Gb/s. This numbers hold with all the Use case loading an AMEN or an MCEN, under the hypothesis done on parameters and end user basin coverage. They don't take into consideration traffic from mass market, but Video components considered in media and entertainment vertical (CDN, IP-TV and 6DoF VR) should cover most of the bandwidth requirements of fixed residential and mobile user devices.

Multicast capacity in the optical layer helps in limiting the total capacity needed in the metro network. This should be further investigated by the project.

Concerning the datacentre facilities, AMEN needs to be provided with Storage and Computing capabilities. The evaluation on the of Use Cases considered gives 200TB to 4PB (PetaBytes), the last value only if the Use Case on Real-Time Low-Latency Object Tracking and Security, which is extremely demanding, is considered, for storage capacity.

For computing capacity. This almost in line with the high level evaluation described in Section 2 and summarized above in Section 6.1, with the important difference that computing capacity results significantly higher with the Use Case approach compared to the High-Level one. This is probably due to the fact that the Verticals considered are quite demanding in terms of computation capacity, while the high level approach does not capture specific requirements and simply scale requirements with CAGR factors.

The optical layer needs to be flexible in order to meet the service requirements, i.e. a control plane needs to be built that can set up, tear down, or reconfigure the services that need to be supported.

Resilience is also a necessary element, which needs the SDN/controller in order to link the optical layer in the metro network with the different access technologies including 5G RN, and the Data Centre facilities in the AMEN.

7 List of acronyms

Acronym	Description
3DoF	Three(3) Degrees of Freedom
3GPP	Third Generation Partnership Project
5G PPP	5G Infrastructure Public Private Partnership
6DoF	Six(6) Degrees of Freedom
ABNO	Application Based Network Operations
ADAS	Advanced Driver Assistance Systems
AMEN	Access Metro Edge Node
API	Application Programming Interface
AR	Augmented Reality
BBU	Base Band Unit
BER	Bit Error Rate
CAGR	Compound Annual Growth Rate
CAGR	Compound Annual Growth Rate
CCTV	Closed Circuit Television
CN	Core Network
CORD	Central Office Re-architected as a Datacentre
CoS	Class of Service
CoS	Class of Service
CPRI	Common Public Radio Interface
CPS	Cyber Physical System
CriC	Critical Connections
CU	Centralized Unit
DA	Data Analytics
DASH	Dynamic Adaptive Streaming over HTTP
DB	Data Base
DC	Data Centre
DM	Data Mining
DU	Distributed Unit
EB	Exabyte
eMBB	Enhanced Mobile Broad Band
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FLOPS	FLoating point Operations Per Second
GMPLS	Generalized Multi-Protocol Label Switching
GPON	Gigabit Passive Optical Network
GPU	Graphic Processing Unit

GSA	Global mobile Suppliers Association
GSMA	Global Mobile Supplier Association (Industrial Association)
HD	High Definition
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure-as-a-Service
IoE	Internet of Everything
IoT	Internet of Things
IT	Information Technologies
ITS	Intelligent Transportation System
KB	Kilobyte
KDD	Knowledge Discovery on Database
KPI	Key Performance Index
MB	Megabyte
MCEN	Metro Core Edge Node
mIoT	Massive Internet of Things
MPD	Media Presentation Description
MPEG	Moving Picture Experts Group
MR	Mixed Reality
NBI	North-Bound Interface
NFV	Network Functions Virtualization
NFV-I	Network Function Virtualization Infrastructure
NFV-O	Network Function Virtualization Orchestrator
NGMN	Next Generation Mobile Networks (Industrial Association)
OA	Obstacle Avoidance
OEM	Original Equipment Manufacturers
OLT	Optical Line Terminal
OSS	Operations Support System
PaaS	Platform-as-a-Service
PLC	Programmable Logic Controller
PP	Path Planning
PTZ	Pan Tilt Zoom
QBER	Quantum Bit Error Rate
QKD	Quantum Key Distribution
RN	Radio Network
RU	Radio Unit
SaaS	Software-as-as-Service
SD	Standard Definition
SDN	Software-Defined Networking

SKR	Secret Key Rate
SLAM	Simultaneous Localization and Mapping
T-SDN	Transport-SDN
UHD	Ultra High Definition
URLLC	Ultra-Reliable Low Latency Connections
V2I	Vehicle to Infrastructure
V2N	Vehicle to Network (Internet servers)
V2P	Vehicle to Pedestrian
V2V	Vehicles to Vehicles
V2X	Vehicle to Everything
VR	Virtual Reality
WAN	Wide Area Network
XR	Extended Reality
ZB	Zetabyte

8 References

[1] Cisco Global Cloud Index:Forecast and Methodology,2015–2020 Online: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf

[2] M. Smolkas, Modular data center market to triple its size by 2020 (2015). Online: http://www.datacenterdynamics.com/contenttracks/power-cooling/modular-data-center-marketto-triple-its-size-by-2020/94114.fullarticle/

[3] Christy Pettey, "Top 10 Technology Trends Impacting Infrastructure & Operations", *Gartner*, Online: http://www.gartner.com/smarterwithgartner/top-10-technology-trends-impacting-infrastructure-operations/

[4] AT&T to Build Edge Data Center Network for Self-Driving Cars, VR/AR. Online: http://www.datacenterknowledge.com/archives/2017/07/24/att-to-build-edge-data-center-network-for-self-driving-cars-vrar

[5] D. Sabella, et al., "Energy Efficiency Benefits of RAN-as-a-Service Concept for a Cloud-Based 5G Mobile Network Infrastructure", IEEE Access, 2, p1586, 2014

[6] Cisco's Internet of Things reference model. Online: cdn.iotwf.com/resources/71/IoT_Reference_Model_White_Paper_June_4_2014.pdf

[7] 3G PP TS # 22.261 V16.1.0 (2017-09) - Service requirements for the 5G system; Stage 1. Online: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3107

[8] Perspectives on Vertical Industries and Implications for 5G. NGMN white paper, 19-09-2016. Online: https://docbox.etsi.org/Workshop/2016/201611_MANAGING_RAIL_MOBILE_COMMS/S04_SEEKING_SYNERGIES/5G_VISION_VERTICALS_NGMN_DEIBERT.pdf

[9] 5G PPP Use Cases and performance evaluation models, (living document, Version: 1.0, 2016-04-25). Online: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-use-cases-and-performance-evaluation-modeling_v1.0.pdf

[10] 5G White Paper, NGMN 17-02-2015. Online: http://www.ngmn.de/fileadmin/ngmn/content/downloads/Technical/2015/NGMN_5G_White_Paper_V1_0.pdf

[11] How 5G changes perceptions of applications, coverage and architecture, GSA white paper, July 2017.

[12] 5G PPP 5G vision brochure. Online: https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[13] N. Nikean et al: Simple Traffic Modeling Framework for Machine Type Communication, In Proc. of ISWCS 2013

[14] "RACH intensity of Time Controlled Devices", 3GPP report R2-102296. Online: http://www.3gpp.org/ftp/tsg_ran/wg2_rl2/TSGR2_69bis/Docs/

[15] Deliverable D3.3 of Combo project "Analysis of transport network architectures for structural convergence". Online: http://www.ict-combo.eu/data/uploads/deliverables/combo_d3.3_pu.pdf

[16] Cisco Visual Networking Index Global Mobile Traffic Forecast Update, 2016-2021 White paper. Online: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[17] ENEA product datasheet. Online: http://www.qosmos.com/wp-content/uploads/2017/10/Enea_Qosmos_ixEngine_Datasheet_170608.pdf

[18] ST 2036-1 "Ultra High Definition Television — Image Parameter Values for Program Production," SMPTE, 2009
[19] S. Namiki et al., "UHD video transmission and extremely Green Optical Networks for Future", IEEE Sel. Top. in Q. Elect. 2011
[20] Cisco, "The Zettabyte Era: Trends and Analysis," White Paper, id 1465272001812119, 2017.
[21] A. Asensio, L. M. Contreras, M. Ruiz, V. Lopez, L. Velasco, "Scalability of Telecom Cloud Architectures for Live-TV Distribution," in Proc. IEEE/OSA Optical Fiber Communication Conference (OFC), 2015.
[22] ISO Standard, "Information technology -- Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media presentation description and segment formats," ISO/IEC 23009-1, 2014.
[23] DASH Industry Forum. Software. Online: http://dashif.org/software/
[24] M. Ruiz, M. Germán, L. M. Contreras, and L. Velasco, "Big Data-backed Video Distribution in the Telecom Cloud," Elsevier Computer Communications, vol. 84, pp. 1-11, 2016.
[25] F. Cugini, M. Ruiz, T. Foggi, L. Velasco, N. Sambo, and P. Castoldi, "Dedicated Protection with Signal Overlap in Elastic Optical Networks (EON)," IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 9, pp. 1074-1084, 2017.
[26] International Federation of Robotics. Definition of Service Robots. Online: https://ifr.org/service-robots
[27] DFG – Schwerpunktprogramm SPP 1835, April 2014, 5G Automated Vision 2015
[28] The automotive revolution is speeding up – Mckinsey Center for Future Mobility. Online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-automotive-revolution-is-speeding-up .
[29] 5G Automotive Association. Online: http://www.5gaa.org
[30] Robot Operating System. Online: http://www.ros.org/
[31] RoboEarth project. Online: http://roboearth.ethz.ch/
[32] Factory of the future, IEC white paper (2015). Online: http://www.iec.ch/whitepaper/pdf/iecWP-futurefactory-LR-en.pdf
[33] P. Feuilherade, "Connecting machines, IoT and the Cloud", IEC e-tech review. May 2015. Online: https://iecetech.org/issue/2015-04
[34] 5G and the Factories of the Future-5G PPP white paper (2015). On line: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf
[35] M. Wollsh Laeger, T. Sauter,and J. Jasperneite, "The Future of Industrial Communication - Automation Networks in the Era of the Internet of Things and Industry 4.0", IEEE Industrial Electronics Magazine, March 2017.
[36] J. Hibbard, "5 Real-Time, Ethernet-Based Fieldbuses Compared", Robotics online, Posted 03/18/2016. Online: https://www.robotics.org/content-detail.cfm/Industrial-Robotics-Tech-Papers/5-Real-Time-Ethernet-Based-Fieldbuses-Compared/content_id/6003
[37] Internet Speed Requirements for Video Streaming, Lifewire. online: https://www.lifewire.com/internet-speed-requirements-for-movie-viewing-1847401
[38] Virtualization for process automation systems, Rockwell Automation Publication: PROCES-WP007A-EN-P January 2013, online: http://literature.rockwellautomation.com/idc/groups/literature/documents/wp/proces-wp007_-en-p.pdf
[39] D. Chatzopoulos, C. Bermejo, Z. Huang, and P. Hui, "Mobile Augmented Reality Survey: From Where We Are to Where We Go," IEEE Access, Vol 5, April 2017.
[40] "Augmented and Virtual Reality: the First Wave of 5G Killer Apps," ABI research and Qualcomm white paper, March 2017, online:

https://www.qualcomm.com/documents/augmented-and-virtual-reality-first-wave-5g-killer-apps

[41] "On-device motion tracking for immersive mobile VR," Qualcomm presentation, August 2017. Online: https://www.slideshare.net/qualcommwirelessevolution/ondevice-motion-tracking-for-immersive-vr

[42] Optical Internetworking Forum, "Flex Ethernet Implementation Agreement," March 2016, Online: http://www.oiforum.com/wp-content/uploads/OIF-FLEXE-01.0.pdf.

[43] Ponemon Institute, "2016 Cost of Data Center Outages," January 2016. Online: https://www.ponemon.org/blog/2016-cost-of-data-center-outages

[44] Uptime Institute, "Uptime Institute Tier Certifications - All Certifications," December 2017. Online: https://uptimeinstitute.com/TierCertification/allCertifications.php?page=1&ipp=All&clientId=&countryName=&tierLevel=4

[45] Oracle, "Synchronously Replicating Databases Across Data Centers – Are you Insane?," The Oracle MySQL Blog, October 2011. Online: https://blogs.oracle.com/mysql/synchronously-replicating-databases-across-data-centers-are-you-insane

[46] Dell EMC, "Distance for sucessful replication," EMC Community Network, May 2010. Online: https://community.emc.com/thread/104126?start=0&tstart=0

[47] Watchity, video on Collaborative Streaming. Online: https://www.youtube.com/watch?v=LK6aKb3VnEg&feature=youtu.be

[48] BBR report "The Statistics of CCTV", Online: http://news.bbc.co.uk/2/hi/8159141.stm

[49] EU General Data Protection Regulation. Online: https://gdpr-info.eu

[50] The UK Quantum Technology Hub for Quantum Communications, online: https://www.quantumcommshub.net/

[51] E. Hugues-Salas, F. Ntavou, Y. Ou, J. E. Kennard, C. White, D. Gkounis, K. Nikolovgenis, G. Kanellos, C. Erven, A. Lord, R. Nejabati and D. Simeonidou, "Experimental Demonstration of DDoS Mitigation over a Quantum Key Distribution (QKD) Network Using Software Defined Networking (SDN)", Paper accepted to be presented at OFC 2018.

[52] K. Antoniou, P. Wright, K. Farrow, A. Lord, R. Nejabati and D. Simeonidou, "Demonstration of SDN Application for Multilayer Video Contribution Network Service", Paper accepted to be presented at OFC 2018.

[53] Common Public Radio Interface (CPRI) Specification V7.0 (2015-10-09)

[54] Uwe Doetsch, et al., "Quantitative analysis of split base station processing and determination of advantageous architecture for LTE,' Bell Labs Technical Journal 18(1), 105–128 (2013)

[55] 3GPP TR 38.801 V2.0.0 (R14), "Technical Specification Group Radio Access Network; Study on New Radio Access Technology; Radio Access Architecture and Interfaces," March 2017

[56] eCPRI Transport Network D0.1 (2017-08-30)

[57] IEEE P1914.1/D0.1 Draft Standard for Packet-based Fronthaul Transport Networks

[58] Werthmann, Thomas, Heidrun Grob-Lipski,"Multiplexing gains achieved in pools of baseband computation units in 4G cellular networks." Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on. IEEE, 2013.

[59] Ge, Xiaohu, et al. "Energy Efficiency Challenges of 5G Small Cell Networks." IEEE Communications Magazine 55.5 (2017): 184-191

[60] Sorensen, N. "Industry Benchmarks Performance." (2011).

[61] COMBO Deliverable 3.3, Analysis of transport network architectures for structural convergence, V2.0, July 2015.

[62] Auer, Gunther, et al. "How much energy is needed to run a wireless network?." IEEE Wireless Communications 18.5 (2011).

Appendix A1 - Dimensioning exercise for the Smart Factories Use Case

This appendix reports a simple exercise made with the aim to evaluate the requirements in terms of aggregated bandwidth and computational capacity needed by a plant of a smart factory of medium size. The model is very generic, and the factory could be for instance a plant for assembling automobiles.

The Hypothesis made for dimensioning exercise are the following ones. The factory is assumed to be organized in islands, each of them made of the components listed in section 4.6.2. In a real plant, islands and automated transportation vehicles are differentiated for specialized functions. For model simplicity the factory is assumed to have a given number of identical cyber physical islands (CPI), each of them including the same system components, and a number of identical automated vehicles for material transportation. Size and quantities of the Smart Factory taken into consideration for the dimensioning purpose are the following ones:

20 cyber physical islands take part of the plant, each composed by:

10	Robots
5	PLC lines (or their evolution)
100	Sensors for process monitoring
2	Video cameras (very high definition) for quality check
1	Video camera (high definition) for plant surveillance

50 automated guided vehicles used inside the plant, to transfer materials among the islands, and to load raw material from the factory entrance and to move finished goods to the exit.

Other components and parts which can be also present in a real and comprehensive smart factory scenario are not considered in this simplified model. The factory model is depicted in Figure 46.

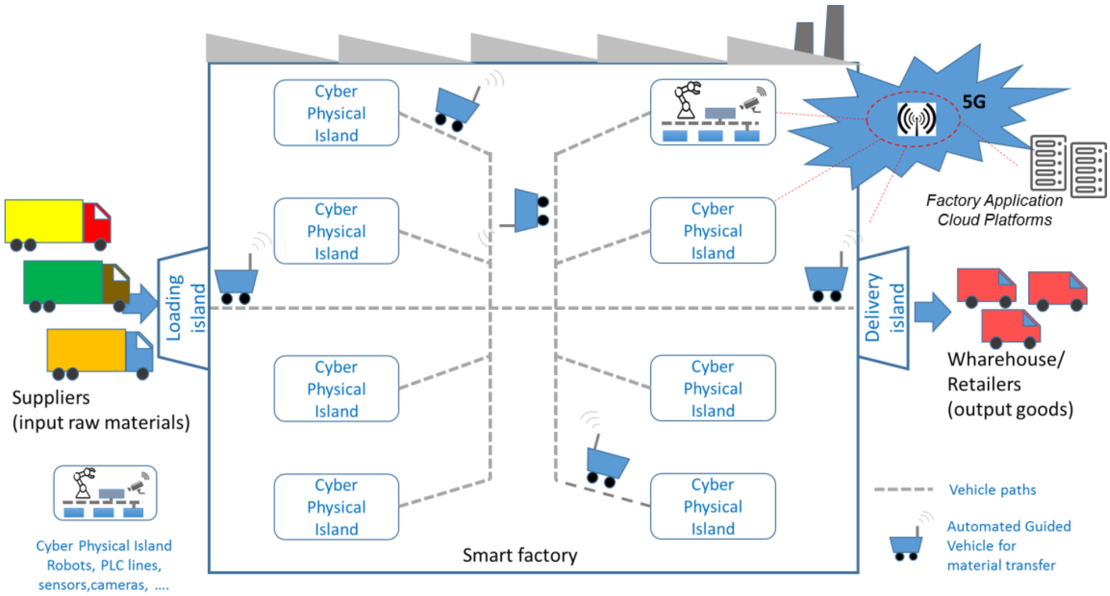


Figure 46 Factory model composed of cyber physical islands communicating by the 5G infrastructure.

Each component (robot, PLC, etc.) is supposed to require computation resources and a mass memory for information storage. Such resources, in traditional factory components like PLCs or robots, are hosted in a local dedicated PC or microprocessor ad hoc system. In smart factories most (even if not all) of the computational and the memory storage needed by a single component can be moved to the cloud as virtual machine (VM). The adopted model assumes two types of VM resources: units of virtual CPU (vCPU or Core), including a standard RAM provision, and a quantity of mass memory storage (in byte).

Some basic concept on virtual computation architecture and resource instantiation can be found in [38]. Figure 47 shows the concept of machine virtualization, the related server architecture and the resource classification.

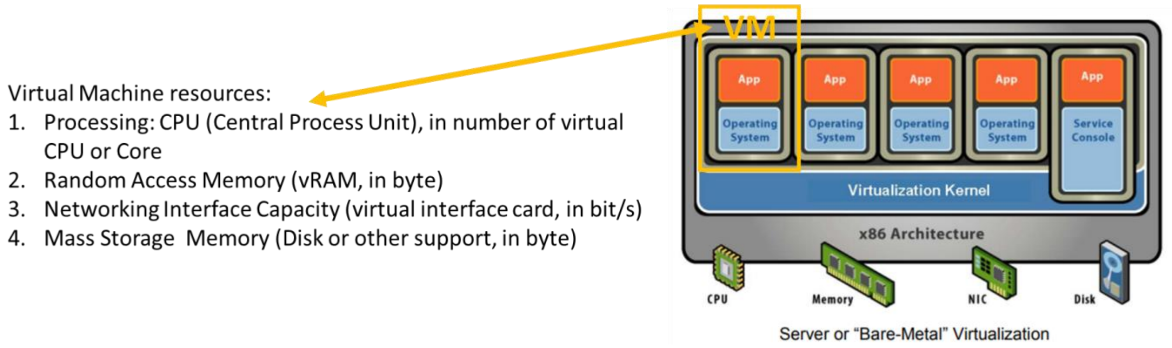


Figure 47 – Model for virtualization of processes on a server. [38]

Robots, PLC lines, transport automated vehicles and video cameras are assumed having their own dedicated VM. In case of sensors, a number of items share the same Virtual machine. Communication needs of a VM associated to a given component is given by the correspondent KPIs of the service (i. e. bandwidth, latency and jitter). Assumptions about requirements in terms of Memory storage and Computation power of each smart factory sub Use Case service is given in Table 49. This requirements extend to memory and computation the connectivity requirements (KPIs) reported in Table 15.

Table 49 – Additional requirements of service components.

	Requrements	*Robot*	*PLC*	*Sensors (Req. for 100 sensors)*	*Video Surveillance*	*Video Quality Check*	*Automated Vehicle*
1	Mass Memory Storage Capacity (MMSC) (SW + temporary data storage)	1 Gbyte	10 Mbyte	1 Gbyte	1 Tbyte	100 Gbyte	100 Mbyte
2	Computation in terms of Virtual CPU units (vCPU) (including standard RAM provision)	2 vCPU	1 vCPU	1 vCPU	1 vCPU	2 vCPU	1 vCPU

The requirements in terms of communication, computing and mass memory storage according to the above assumptions and supposing also that the factory is installed in an area of 1 km^2, are the following ones:

- 200 robots requiring 200 connections at 1 Mb/s URLLC each (200 Mbit/s in total, Area Traffic of 200 $(Mb/s)/km^2$) and 200 Gbyte of MMSC and 400 vCPU in total

- 100 PLC lines requiring 100 connections at 100 Kb/s URLLC each (10 Mbit/s in total, Area Traffic of 10 (Mb/s)/km^2) and 10 Gbyte of MMSC and 100 vCPU in total
- 2000 sensors requiring 2000 connections at 10 kb/s mIoT each (20 Mbit/s in total, Area Traffic of 20 (Mb/s)/km^2) and 20 Gbyte of MMSC and 20 vCPU (please note that density is 2000 sensors / 1 km^2= 2000 sensors per km^2)
- 40 UHD video cameras requiring 40 connections at 100 Mbit/s eMBB each (4 Gb/s in total, Area Traffic of 4 (Gb/s)/km^2) and 40 Tbyte of MMSC and 80 vCPU in total
- 20 HD video cameras requiring 20 connections at 5 Mb/s eMBB each (100 Mb/s in total, Area Traffic of 100 (Mb/s)/km^2) and 2 Tbyte of MMSC and 20 vCPU
- 50 Automated vehicles requiring 50 connections at 1 Mbit/s URLLC each (5 Mbit/s in total, Area Traffic of 5 (Mb/s)/km^2) and 5 Gbyte of MSC and 50 vCPU in total

Such requirements the whole factory are presented in an aggregate form and also specified per CoS in Table 50.

Table 50 – Capacity requirements of service components at factory level. (Total and differentiated by Class of Service)

Resource	**Total**	*URLLC*	*eMBB*	*MIoT*
Aggr. bandwidth	**4.38 Gb/s**	*260 Mb/s*	*4.1 Gb/s*	*20 Mb/s*
Area traffic	**4.38 (Gb/s)/Km^2**	*260* (Mb/s)/Km^2	*4.1* (Gb/s)/Km^2	*20* (Mb/s)/Km^2
MMSC (memory)	**42.2 Tbyte**	*215 Gbyte*	*42 Tbyte*	*20 Gbyte*
vCPU	**670 vCPU**	*550 vCPU*	*100 vCPU*	*20 vCPU*

The dimensioning example reported in this appendix can be representative of a Factory of medium size and for manufacturing goods of big size (for instance automobiles). It has 200 robots and it is supposed to occupy an area of 1 Km^2. To give some example, the automotive FCA factory in Melfi, Italy, currently has 320 operating robots and it occupies a surface of about 1.7 Km^2 (800 employees and 400,000 car/year). Another example is the huge AUDI factory located in Ingolstadt, Germany, where 800 robots are installed on a surface of 2.7 Km^2 (44,000 employees and 600,000 car/year).

Other types of manufacturers could require different kind of plants with different size, mix and density of machinery (robot or PLC per Km^2), and this could change significantly the Capacity requirements for the 5G network for providing the communication and computation services to the business. The plants which produce goods of big dimension like cars, trucks, ships or airplanes require huge space for storing, moving and processing the semi-finished parts and then for that types of manufacture the density of communication requirements is not high. For other kind of manufacturing like the consumer electronics (PC, smartphones, TV, etc.) the density of machinery and sensors required in the smart factory can be much higher and values of two to three orders of magnitude greater than the one reported in Table 50. In addition, scenarios could be very different in terms of number of plants in a given area. One big factory or many small factories (or a mix of both) can be present in an area of 1 Km^2 which is the reference area surface assumed in the above example.

< END OF THE DOCUMENT >