5G Mobile Transport and Computing Platform for verticals

5G Mobile Transport and Computing Platform for verticals

Paola Iovanna, Teresa Pepe
Ericsson Research
Pisa, Italy

Francesca Moscatelli
Nextworks Srl
Pisa, Italy

Carla Fabiana Chiasserini, Claudio Casetti
Politecnico di Torino
Torino, Italy

Luca Valcarenghi, Barbara Martini
Scuola Superiore Sant'Anna,
Pisa, Italy

Xi Li
NEC laboratories Europe
Germany

Carmen Guerrero
Universidad Carlos III Madrid
Leganes, Spain

Adlen Ksentini
EURECOM
Biot, France

Josep Mangues-Bafalluy
Centre Tecnològic de Telecomunicacions de Catalunya
(CTTC/CERCA)

Giuliana Zennaro
Centro Ricerche Fiat S.C.p.A.
Torino, Italy

***Abstract*— The support of 5G verticals service requires to design an efficient Mobile Transport and Computing Platform where transport, mobile and MEC must interact effectively. In this paper, a novel architecture is proposed providing its mapping on ETSI NFV. Two relevant use cases, such as automotive and cloud robotics are presented to assess the novel architecture.**

Keywords—MTP; Abstraction; Verticals; NFV; 5G

I. Introduction

The fifth generation of mobile technology (5G) will be a key component of the Networked Society. 5G will provide wireless connectivity for a wide range of new applications and use cases. As a result, it will also accelerate the development of the Internet of Everything, that is the connection of anyone and anything, in anyplace at any time.

Networks will experience massive increase in traffic and an ever-expanding number of connected devices will be served meeting stringent requirements for performance characteristics (like reliability and latency).

Mobile and transport networks have to cooperate to provide a final connectivity service with very tight requirements. Several radio split options have been defined to deal with the 5G tight requirements and several transport solutions are under definition to support at the best all that. In most of cases a mix of radio split scenarios where distributed RAN, Cloud/Centralized RAN, with a mix of radio interfaces should concurrent coexist.

In this context, the 5G-TRANSFORMER project aims to drive the transformation of the traditional Mobile Transport Networks from today's rigid interconnection solutions, through an SDN/NFV-based 5G Mobile Transport and Computing Platform (MTP) able of simultaneously supporting an extremely diverse range of networking and computing requirements to meet the specific needs of vertical industries.

In more detail, 5G-TRANSFORMER will deliver a complete scalable MTP by adding the support of: i) integrated MEC services, ii) dynamic placement and migration mechanisms of virtual functions, iii) new mechanisms for sharing of VNFs by multiple tenants and slices, iv) new abstraction models for vertical services, and iv) customized profiles for the C-RAN functional split considering the requirements from verticals.

The rest of the paper is organized as follows. In Section II the 5G-TRANSFORMER framework is presented with the main components VS, SO and MTP. Section III presents the layered architecture for the MTP to manage the mobile and

This work has been partially funded by the EU H2020 5G-Transformer Project (grant no. 761536). * Corresponding author email: paola.iovanna@ericsson.com

transport infrastructures. Section IV provides a more detailed description of the MTP components in relation with the ETSI NFV framework. Section V introduces the design principles of the MTP for the selected uses cases on automotive and robotics.

II. 5G-TRANSFORMER FRAMEWORK

We envision the 5G-Transformer system consisting of 3 major components: *vertical slicer* (VS), *service orchestrator* (SO) and *mobile transport and computing platform* (MTP), as illustrated in Fig. **1**.

The VS is the common entry point for all verticals into the 5G-Transformer system, being part of the operating and business support systems (OSS/BSS) of the administrative domain of a given provider. The VS coordinates and arbitrates the requests for vertical services. Vertical services are offered to the verticals through a high-level interface focusing on the service logic and needs of vertical services. It allows composing vertical services from a set of vertical-oriented service blueprints, which along with instantiation parameters will result in a vertical service instantiation request. Then, the VS maps descriptions of vertical services and requirements at the vertical application level onto a network service descriptor (NSD), which is a service graph composed of a set of V(N)Fs chained with each other and fine-grained instantiation parameters (e.g., deployment flavor) that are sent to the SO. The VS and its functionalities are explained in [1].

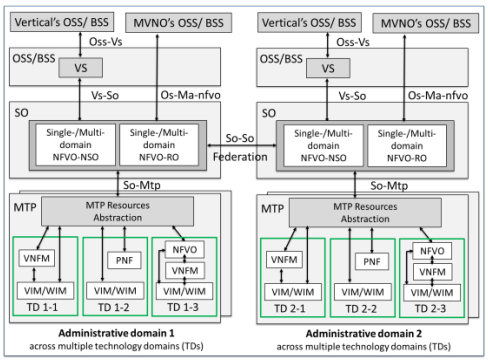

Fig. 1 5G-Transformer basic system architecture

The SO [2] provides end-to-end service orchestration across multiple administrative domains. It receives requests from the VS or directly from the M(V)NO. Depending on the use case, both network service (NFV-NSO) and resource (NFVO-RO) orchestration may be used for both single and multi-domains [3]. In turn, based on the request, the SO may decide to create a new network slice instance or to reuse one previously created by the provider to be shared. Therefore, it manages the monitoring and allocation of virtual resources to network slices, be it for vertical services or for network services of an M(V)NO. If needed (e.g., not enough local resources), the SO interacts with the SOs of other administrative domains (federation) to take decisions on the end-to-end (de)composition of virtual services and their most suitable execution environment. Even if a service is deployed across several administrative domains, e.g., if roaming is required, a vertical still uses one VS to access the system, and so, the SO hides this federation from the vertical. The NFVO-RO functionality of the SO handles resources coming from the local MTP (real or abstracted) and from the SOs of other administrative domains (abstracted). The orchestration decision for creating or updating a network slice includes the placement of V(N)Fs over such virtual networks with virtual nodes and links, as well as the resources to be allocated. The SO will then request the MTP to create the slice instance.

The MTP is responsible for orchestration of resources and the instantiation of V(N)Fs over the infrastructure under its control, as well as managing the underlying physical mobile transport network, computing and storage infrastructure. The computing and storage infrastructure may be deployed in central data centers as well as distributed, as in Multi-Access Edge Computing (MEC) [7]. The MTP provides support for slicing, enforces slice requirements coming from the SO and provides physical infrastructure monitoring and analytics services. Depending on the use case, the MTP may offer different level of resources abstraction to the SO via the MTP resources abstraction component, which in turn forwards the SO requests to the right entity accordingly (as single point of contact): VIM/WIM, VNMF or PNF, or NFVO [3].

1) *Case 1*: the MTP exposes virtual resources and the possibility to instantiate entire VNFs through the VNFM.
2) *Case 2*: the MTP exposes PNFs that can be just configured but not instantiated (e.g. a physical BTS). At the VIM/WIM level the MTP just instantiates virtual resources related to networking.
3) *Case 3*: the MTP abstracts an entire network service to the SO and it takes care internally about how to orchestrate it, through the NFVO – VNFM - VIM/WIM stack.

III. LAYERED ARCHITECTURE FOR RADIO TRANSPORT AND IOT SERVICES

According to the general architecture defined in 5G Transformer, the MTP has the main task to handle the infrastructure resources to manage suitably mobile and transport to provide the connectivity to the computation resources selected for the vertical services request.

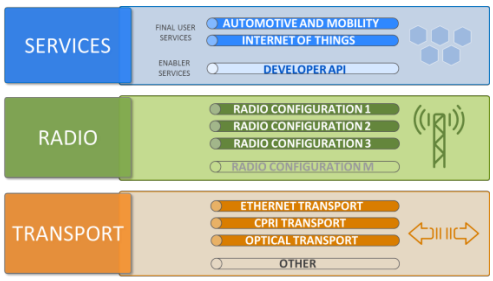

Fig. 2 Layered architecture for radio transport and IoT Services

A layered architecture is defined to regulate the interworking between mobile and transport layers. As shown in **Fig. 2**, a client-server relationship is considered where each layer provides a service defined by detailed service parameters

and exposes to the upper layer a suitable abstract view. According this architecture well defined demarcation points of the services can be identified between the adjacent layers that allows to define the information to be exchanged in a sort of Service Level Agreement. Hence ingress and egress points of the service is defined with parameters such as latency, resiliency, etc. Technology specific information such as numbers of nodes composing the E2E path in the transport, specific transport channel (e.g. which wavelengths in case of optical technology) are hidden to the exposition view to the upper layer. This approach allows to meet the following needs: i) keep independent the technology of each layer from the model so it is possible to keep the same model when the technology changes; ii) simplify the tasks of each layer with a clear separation of role and responsibility, to facilitate fault locations; iii) decouple the radio and transport solutions that can evolve to different releases independently; iv) associate each layer to different providers that can be combined to each other in N:M relationship. For example, there are situations where same transport provider shares the transport infrastructure to more radio providers, or the case where a radio provider makes use of more transport providers. In this approach, the transport layer works on the physical infrastructure resources and provides the suitable abstract view to the radio. Such abstract view can be organized in slices view according several criteria. In turn, the radio layer works on the abstract view exposed by the transport to manage the radio resources to provide the connectivity services to the SO. Again, also at the radio level a slices view has to be provided according different criteria. In case of Cloud RAN, also the radio functions are virtualized and provided as generic processing on data-centers and server. In this case, the MTP has to manage it and coordinate suitable with the transport to provide the connectivity required to the SO.

IV. MTP HIGH LEVEL ARCHITECTURE (MAPPING IN ETSI NFV)

The design of the MTP architecture draws on the system architectures defined within the H2020 5G-Crosshaul project [3][5], which leverages the standard and reference specifications of the SDN and NFV architectures [5][6]. Specifically, on the data plane, the 5G-Crosshaul architecture includes two types of nodes: the Crosshaul Forwarding Elements (XFEs), handing data traffic, and the Crosshaul Processing Units (XPUs), which are in charge of computing operations. The XFEs can also cope with different link and physical-layer technologies, thanks to the introduction of an innovative common framing to transport both backhaul and fronthaul traffic. XPUs instead can host VNFs and support C-RAN related operations. The main component of the 5G-Crosshaul control plane is instead the Crosshaul Control Infrastructure (XCI), which integrates the SDN control in the ETSI/NFV MANO architecture [3]. The XCI also provides an abstracted view of the available resources, states and functions to the SO through the Northbound Interface (NBI). The Southbound Interface (SBI) instead connects the XCI to the data plane nodes and allows the execution of control and management functions on the hardware elements. Within the XCI structure there is the controller layer, composed of the network, computing, and storage controllers, enabling the allocation and configuration of the different resources composing the NFVI. 5G Transformer project extends the 5G-Crosshaul transport solution with MEC and dynamic creation of slices and placement of VNFs to take into account the needs of vertical industries. Fig. 3 presents the MTP mapping with the ETSI NFV MANO architecture [3], highlighting the three architectural alternatives proposed in Section II. Indipendently from the kind of service exposed by the MTP to the SO, the MTP should contain the following components.

VIMs, each in charge of managing storage, networking and computational resources in its NFVI-PoP administrative domain. A VIM is typically handled by cloud platform, like e.g. OpenStack. In addition, each NFVI-PoP/administrative domain under the VIM responsibility may include one or more SDN Controllers (e.g. OpenDaylight) in charge of establishing the transport connectivity between VNFs deployed within an NFVI-PoP. In case of multi-layer or multi-technology network infrastructures, SDN Controllers can also be deployed in a hierarchical model to handle the heterogeneity of the technological domains through dedicated child controllers.

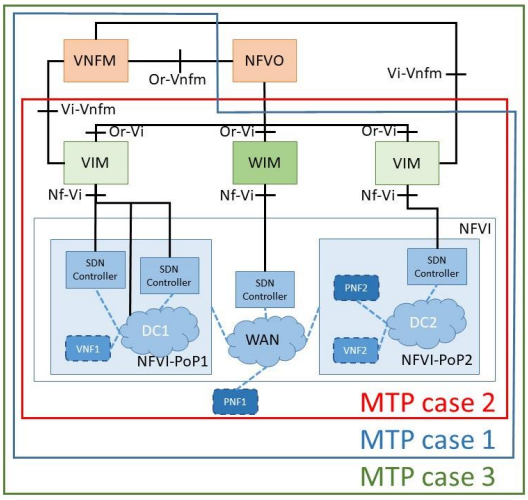

Fig. 3 MTP mapping with ETSI NFV MANO

WIM, in charge of providing inter-domain vLinks, which will be translated in configurations of the transport network between NFVI-PoPs gateways through the proper SDN Controller.

NFVI, which provides all the hardware (e.g. compute, storage and networking) and software (e.g. hypervisor) components that constitute the infrastructure where VNFs are deployed. Eventually, also PNFs to be shared between different NSs can be taken into consideration for the virtualisation infrastracture.

A VIM or either a WIM can interface with the underlying SDN Controllers to request virtual connectivity services through the Nf-Vi reference point or establish directly the connectivity services by configuring the network nodes. In this case, SDN Controllers become part of the VIM itself, controlling directly virtual entities such as virtual switches or

network functions within the related NFVI PoP. This kind of hierarchy in management and orchestration of heterogeneous resources provided at the NFVI brings the benefit of different layers of abstraction, where, from the bottom to the upper layer of the MTP inner architecture, each component provides the proper NBI to request services. With the aim of offering NFV MANO services across multiple administrative domains, the NFVI pool of resources can be provided as a service [7]. In the NFVIaaS paradigm, we can identify the consumer as a service provider which wants to run VNF instances inside an NFVI provided as a service by a different administrative entity: the NFVIaaS provider. This means that the NFVIaaS consumer has the control of the VNF instances, but it does not control the underlying infrastructure. In particular, since the provider's NFVI is structured in several VIMs, the provider can offer the access to the service following two different types of interaction between the two administrative entities:

- Multiple Logical Point of Contact (MLPoC), where the consumer has the visibility of the different VIMs within the provider's administrative domain and communicate directly with each of them.
- Single Logical Point of Contact (SLPoC) (see **Fig. 4**), where the VIMs are hidden to the consumer and the provider's administrative domain contains a SLPoC function in charge of acting as a single unified interface offered to the consumer.

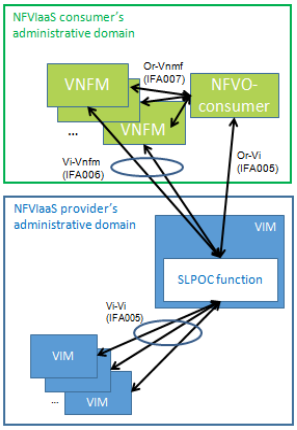

Fig. 4 SLPoC function

To enable the deployment of vertical use cases with mobile applications that require very low latency, the MTP architecture shown before should be extended to deal with the Multi-access Edge Computing (MEC) technology. In fact, the possibility of reducing the latency, bringing IT and cloud computing capabilities near to mobile access side, allows the deployment of use cases in different industry's branches, such as the automotive and the cloud robotics, where the "instantaneous" processing of the data is a key factor. An example of possible integration between MEC and NFV MANO architecture is provided in [8]. On the MEC side, we *can identify the following components for the MTP MEC extension.*

Mobile Edge Platform (MEP) a VNF deployed at the MTP NFVI-PoP or NFVI edge. It offers services, such as Radio Network Information Service (RNIS), and location API for ME VNF applications. The latter are deployed in the same NFVI-PoP. The MEC applications is using MEC service to adapt the application to user context or run low-latency applications at the edge.

Mobile Edge Platform Manager – NFV (MEPM-V) it corresponds to the MEP Element Manager. It is in charge of managing the application rules and requirements. The lifecycle management in the context of MTP is delegated to the VNFM-MEC.

VNFM-MEC – It is in charge of Life Cycle Management of the MEC application VNF as well as the Mobile Edge Platform. It is connected to the MTP NFVO via the well-defined Or-Vnfm interface, while it uses the Ve-Vnfm-em and Ve-Vnfm-vnf interfaces to communicate with the MEPM and MEC application VNF, respectively. At the MTP level, the VNFM-MEC communicates with the VIM in order to manage the needed resources for the deployment of the MEC Apps, where the VIM uses the Nf-Vi interface to manage the NFVI Edge resources, e.g. supporting containers.

V. MTP ABSTRACTION AND INTERWORKING WITH SO: A USE CASES PROSPECTIVE

Resources abstraction is a key element for the interworking of the MTP and the SO. To allow a suitable deployment of the resources, the MTP have to provide an abstract view of the available resources with an adequate level of detail. For this reason, one of the objective of the 5G Transformer is to provide new abstraction models for vertical services. In the context of 5G-Transformer, the MTP is made up of Radio Access Network (RAN), transport network, and core network functions, mobile edge computing (MEC) infrastructure and computing resources. Thus, the MTP will provide a scalable and efficient abstraction that takes into account all these aspects. In particular, to allow a correct selection of the resources for a specific service, the MPT will expose (with the suitable level of abstraction) information about:

- availability of datacenter resources, identifying also the geographical location of the servers for a correct placement of the V(N)F
- type of available connectivity.

As far as the connectivity is concerned, two possible approaches may be followed. In the first one, the MTP exposes the connectivity hiding the complexity of the radio and transport interaction. It just provides as a logical link characterized with performance parameters e.g., latency, delay, bandwidth. The network functions to implement (e.g., eNB, EPC) as either physical or virtual network functions and the type of functional split to be implemented is responsibility of the MTP that will orchestrate them not only on the bases of the service requirements but also taking into account the constraints of the underlying transport network. Using the second approach instead the MTP exposes also the network functionality (e.g., eNB, EPC) giving the possibility at the SO level to choose whether to implement some network functions

(both PNF and VNF). In the following, to better understand the main functionalities and requirements of the MTP and highlight the importance of the resource abstraction, two relevant use cases (*Automotive* and *Cloud Robotics*) are described.

A. Automotive

Future connected vehicles pave the way to the development of several services where the automotive industry and mobile networks play a fundamental role. Among such services, Vehicle-to-Everything (V2X) safety applications have a prominent social and economic impact. The exchange of information on the vehicle dynamics, their processing as well as the delivery of warnings, can greatly reduce the risk of accidents involving vehicles as well as vulnerable roads users (pedestrians, cyclers). While the traditional way to deploy such services foresees vehicle-to-vehicle communication and safety applications implemented at the vehicle, within the 5G-TRANSFORMER project, an infrastructure-based deployment has been considered, taking the collision avoidance between vehicles as a use case. In this scenario, an example of infrastructure-based service deployment is presented below. According to the ETSI specifications in [9], each vehicle periodically (e.g., every 0.1 s) generates Cooperative Awareness Messages (CAMs), including the vehicle position, speed, heading, among others. As depicted in Fig. 1, in our example CAMs are transmitted as V2I unicast messages to the (v)eNB covering the area of interest (e.g., urban intersection). Messages are then forwarded towards the Packet Gateway (P-GW) within the (v)EPC, which then hands them to a Collision Avoidance Server (CAS). The CAS should store the most recent CAM sent by the vehicles travelling over the geographical area of interest. Within the CAS, a collision detection algorithms then processes the information and detects the risk of collisions between the vehicles, if any. To this end, an algorithm following the approach presented in [10] can be adopted. Upon detecting a possible collision, the application generates a warning message, following the Decentralized Environmental Notification Message (DENM) format specified by ETSI [11], which is delivered to the involved vehicles through the (v)EPC and the (v)eNB. Vehicles receiving the warning can then display it to the driver, or actually execute a proper action (e.g., braking) if they are automated vehicles. In this context, one main feature of the MTP is the provisioning of connectivity service (CS), i.e., the MTP should make sure that radio connectivity is provided in the areas of interest and the information included in the CAM and DENM messages is properly handled at the data plane. Additionally, control traffic should be handled by the (v)EPC, and the connection between P-GW and CAS should be provided. To this end, different technologies can be used such as Metro Ethernet Forum (MEF) carrier Ethernet services like E-line, E-LAN and E-tree, or, alternatively, optical connections, and Label Switched Paths (LSP). Moreover, in the considered example, the MTP must choose also whether to implement some network functions (e.g., eNB, EPC) as either physical or virtual network functions and the type of functional split to be implemented based not only on the service requirements but also on the constraints of the underlying transport network.

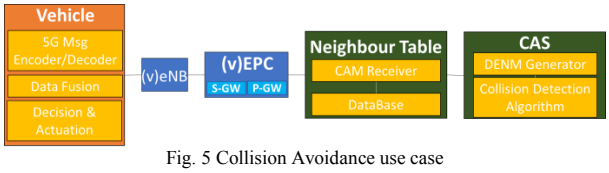

Fig. 5 Collision Avoidance use case

Importantly, as mentioned, collision avoidance is characterized by strict latency requirements, which make it a candidate for a Multi-access Edge Computing (MEC)-based implementation.

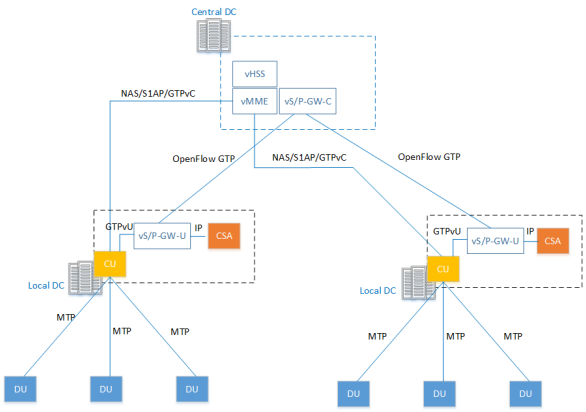

Fig. 6 A detailed description of the splitting of functions EPC and C-RAN

Fig. 3 illustrates a detailed view of the EPC and RAN function split and their placement for the automotive use-case. All these elements are deployed by the MTP NFVO, as per the request of the VS/SO. As explained before, the MTP manages a central DC and several Area DC (i.e. NFVI-PoP and NFVI-Edge). At the RAN level, the 4.5 split is used, which means that the PDCP layer and upper control plan functions are located at the CU, while the lower layers (RLC, MAC, PHY) are run at the DU. The CU is run as a software entity, and hosted at the local DC. The MTP transport layer is in charge of carrying the traffic from the DU to CU. The CSA function is in charge of communicating with vehicle application, it is deployed at the local DC to reduce the communication latency. It is worth noting that the CSA is expecting to receive IP packets. However, the traffic going out from the CU is encapsulated in GTP packets, which introduces the need to have a GTP encapsulation and decapsulation process at the edge. In this case, we consider the 3GPP CUPS proposition [20] to split the S/P-GW into two entities: one managing the control plan (S/P-GW-C) and one managing the data plan (S/P-GW-U). Therefore, the S/P-GW-U will be hosted at the local DC. The S/P-GW-C, on the other hand, is located with the other EPC element in the central DC. The S/P-GW-U is managed by the S/P-GW-C using OpenFlow protocol (GTP patched) [12] to install rules to encapsulate and decapsulate GTP traffic. Besides managing several S/P-GW-U deployed at the edge, having a centralized S/P-GW-C allows to handle mobility by redirecting the traffic from/to the S/P-GW-U using the flexibility of OpenFlow. When the vehicle is

connected to the network, the MME/SP-GW-C create a bearer connecting the CU with the local S/P-GW-U. Then the /P-GW-C pushed rules to the S/P-GW-U to encapsulate and decapsulate GTP packet, allowing to host the CSA at the MEC. It is worth noting that the proposed architecture may be extended to any service, by using appropriate function split and their placement in the MTP.

B. Cloud Robotics

Cloud Robotics (CR) is a paradigm that leverages on cloud technologies and mobile communication to enhance the capabilities of robots. Control services are moved into the cloud running on dedicated hosts or datacenters allowing the development of smart robotic systems with unlimited computing capacity. Offloading computation-intensive tasks to the cloud, only the necessary sensors, actuators, and basic processing power are kept on the robots. The communication with the remote control is made possible by a connection through the mobile network. The mobile communication must satisfy specific requirements in terms of data rates, latency, reliability, density of connections, coverage, etc.

Within the 5G-TRANSFORMER project, it was decided to analyze the CR service for industrial automation because considered particularly challenging for the project. In particular, we consider an exemplary use case of a warehouse automation where the control of the production processes and of the robot's functionalities is moved into the cloud, exploiting wireless connectivity to minimize infrastructure complexity, optimize processes, and thus implement lean manufacturing. To make possible a real-time interaction among robots and the external environment, a huge amount of information will have to be transferred instantaneously. Thus, very tight requirements are imposed in terms of latency, reliability, and availability to guarantee that:

- data generated at a sensor are correctly received by the actuator, whatever is in between
- messages are successfully transmitted within a defined latency budget or delay (in industrial automation, only one message in one billion data transfers may be lost or delayed more than the given latency budget)
- the system will work as required in a given period (this typically imposes a "5-nines" availability on wireless links)

Consequently, the VNFs composing the CR service have to be conveniently selected and instantiated to satisfy the service requirements; for instance, non-time-critical services, like the system in charge to manage and control the plant, can be located remotely. Time-critical services, instead, like navigation or data processing have been kept close to the radio base station (preferring a MEC-based implementation) to guarantee low latency, stability and a proper reaction time. In 5G-transformer, data collected by sensors must be transmitted to a remote control (the virtual PLC - vPLC), located in the cloud, through a low latency network [13] (e.g., e-Xhaul [14]) as shown in **Fig. 7**. In this specific use case, messages enter the 5G Network from the Distributed Unit, reach the Centralized Unit through the e-Xhaul network by means of a suitable wavelength, then they are forwarded towards the Packet Gateway (P-GW) within the (v)EPC to reach the vPLC.

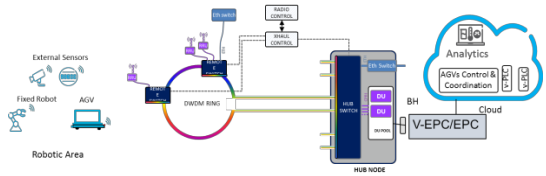

Fig. 7 Cloud Robotics use case

This means that the MTP on the base of the SO request, must manage the DC resources selected for the specific service and to orchestrate transport and radio resources to provide the required connectivity. On the other hand, to allow a cross optimization of both computing-storage and radio-transport resources, the MTP must provide the suitable resource abstraction.

REFERENCES

[1] Casetti et.al, Network Slices for Vertical Industries, submitted to ComVert18

[2] Xi Li et.al., Service Orchestration and Federation for Verticals, submitted to ComVert18

[3] ETSI GS NFV-MAN 001 V1.1.1 (2014-12), Network Functions Virtualisation (NFV); Management and Orchestration.

[4] 5GPPP 5G-Croshaul project, http://5g-crosshaul.eu

[5] X. Costa-Perez et al., "5G-Crosshaul: An SDN/NFV Integrated Fronthaul/Backhaul Transport Network Architecture," in IEEE Wireless Communications, vol. 24, no. 1, pp. 38-45, February 2017.

[6] X. Li et al., "Novel Resource and Energy Management for 5G integrated backhaul/fronthaul (5G-Crosshaul)," 2017 IEEE International Conference on Communications Workshops (ICC Workshops), Paris, 2017, pp. 778-784.

[7] ETSI GR NFV-IFA 028 V0.11.1 (2017-09), Network Functions Virtualisation (NFV); Management and Orchestration; Architecture Options to Support the Offering of NFVO Mano Services Across Multiple Administrative Domain.

[8] B. Li et al., "An MEC and NFV Integrated Network Architecture", April 2017, http://wwwen.zte.com.cn/endata/magazine/ztecommunications/2017/2/articles/201706/P020170613285738133203.pdf

[9] ETSI TS 102 637-2 V1.2.1 (2011-03) - Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications Part 2: Specification of Cooperative Awareness Basic Service.

[10] G. Avino et al., "A Simulation-based Testbed for Vehicular Collision Detection," IEEE VNC, Turin, Nov. 2017.

[11] ETSI TS 102 637-3 V1.1.1 (2010-09), Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 3: Specifications of Decentralized Environmental Notification Basic Service.

[12] I. Alawe et al., "On Evaluating Different Trends for Virtualized and SDN-ready Mobile Network", in Proc. of IEEE CloudNet 2017, Prague.

[13] A. De La Oliva et al., "Xhaul: toward an integrated fronthaul/backhaul architecture in 5G networks" IEEE Wireless Communications, Volume: 22 Issue: 5

[14] P. Iovanna et al., "Future Proof Optical Network Infrastructure for 5G Transport" Journal of Optical Communications and Networking Vol. 8, Issue 12, pp. B80-B92 (2016)