5G-XHaul: Enabling scalable virtualization for future 5G Transport Networks

Daniel Camps-Mur*, Paris Flegkas†, Dimitris Syrivelis†, Qing Wei‡, Jesús Gutiérrez§
*i2CAT Foundation, Barcelona, Spain
† University of Thessaly, Volos, Greece
‡ Huawei Technologies Duesseldorf, Munich, Germany
§ IHP GmbH, Frankfurt, Germany
daniel.camps@i2cat.net, {pflegkas,dimitris.syrivelis}@gmail.com, qing.wei@huawei.com, teran@ihp-microelectronics.com

***Abstract*—Network slicing is a major trend in the design of future 5G networks that will enable operators to effectively service multiple industry verticals with a single network infrastructure. Thus, network slicing will shape all segments of the future 5G networks, including the radio access, the transport network and the core network. In this paper we introduce the control plane design produced by the 5G-XHaul project. 5G-XHaul envisions a future 5G transport network composed of heterogeneous technology domains, including wireless and optical segments, which will be able to transport end user and operational services. Consequently, 5G-XHaul proposes a hierarchical SDN control plane where each controller is responsible for a limited network domain, and proposes a multi-technology virtualization framework that enables a scalable slicing of the transport network by operating at the edge of the network.**

I. INTRODUCTION

A major trend driving the design of future 5G networks is the desire to open the mobile networks to a plurality of industry verticals, instead of having a network focused only in delivering a mobile broadband service as it has been the case for the past generations. In this regard, the 5G-PPP initiative set up in Europe, has already recognized automotive, energy, health, media and manufacturing as the main verticals to be addressed by future 5G networks [1]. This trend has also been recognized by the NGMN, which has accordingly introduced the concept of *network slicing* [2].

Network slicing is the concept to partition a physical network infrastructure into multiple logical networks and correspondent network resources that can be used to address the requirements from different verticals and services. The alternative, namely deploying purpose built networks for each vertical/service is not economically sustainable. Thus, inspired by the advances witnessed in the past years in the area of cloud computing, network slicing leverages the concepts of *virtualization* and *software defined networking (SDN)*. Virtualization is a technique that allows to provision abstract resources made of the slice or union of underlying physical resources; being the Virtual Machine (VM) abstraction the most common example of virtualization. The 5G community is currently studying how the resources at the various levels of the network, namely access, transport and core, can be virtualized. In the context of 5G, virtual resources are referred to as *Virtual Network Functions* (VNFs). SDN is related to the control of the network resources including the virtualized network resources, which allows operators to efficiently adapt their network services to the requirements of different verticals and in turn reduces the CAPEX/OPEX. In addition, SDN should provide the operator with the ability to easily compose and deploy new network services, which could be for example instantiated through different network slices.

Virtualization and softwarization will shape the architecture of future 5G networks, as recognized by the 5G-PPP Architecture Working Group in [3]. In particular, the virtualization and softwarization trends will also affect the design of the future 5G transport networks, namely the networks connecting the 5G access and core networks. In addition to the general requirement to support slicing, future 5G transport networks also have to address specific requirements such as the cost effective transport of the fronthaul and backhaul interfaces required to support centralized (C-RAN) and distributed RAN deployments. In addition, future 5G transport networks will have to enable novel functional splits, which can provide a better trade-off between the pooling and coordination gains offered by C-RAN architectures, and the light transport requirements imposed by distributed RAN architectures. Examples of these novel functional splits have been discussed in [4] and [15].

5G-XHaul [5] is a European project under the umbrella of the 5G-PPP initiative, aiming to address the previous challenges in the design of the future 5G transport networks. The major contribution of this paper is an overview of the control plane design produced by the 5G-XHaul project, which enables the unified control of a transport network infrastructure composed of heterogeneous technology domains in a scalable way, whilst enabling virtualization and slicing.

This paper is structured as follows. Section II introduces the 5G-XHaul layered architecture. Section III provides a detailed overview of the control plane design focusing on the aspects of virtualization and scalability. Section IV discusses the potential benefits obtained from a tight coupling between the transport and radio access networks in 5G. Finally, section VI summarizes and concludes this paper.

II. 5G-XHaul Layered Architecture

Figure 1 illustrates the layered architecture proposed within the framework of 5G-XHaul. The goals of this architecture are threefold: i) to propose a physical architecture able to simultaneously support end user and operational services, e.g. backhaul and fronthaul interfaces, whilst being operated by a converged SDN control plane, ii) to support the instantiation of virtual transport slices that can be remotely controlled by each tenant, and iii) to provide interfaces to higher level orchestrators, such as those defined within the ETSI MANO reference architecture [6].

In order to achieve the previous goals, 5G-XHaul splits the required functionality in four majors layers illustrated in Figure 1. The bottom layer is the *Managed Physical Infrastructure* layer, consisting of the physical wireless and optical network elements, as well as the compute resources, namely edge and regional data centers, connected by the network. On the wireless domain, 5G-XHaul considers street level mesh networks of mmWave and Sub6 wireless devices used to connect densely deployed small cells. On the optical domain, 5G-XHaul considers WDM-PON in the access and a dynamic frame based optical network in the metro segment, implemented by means of TSON [7]. In addition, the proposed architecture also allows to accommodate other packet based technologies in the access and metro domains such as Ethernet. The managed physical infrastructure layer contains a hierarchical control plane that allows the 5G-XHaul infrastructure provider to control in an unified way a physical infrastructure composed of heterogeneous technology domains. The 5G-XHaul hierarchical control plane will be discussed in detail in section III-D. Operating on top of the managed physical infrastructure, lies the *Infrastructure Management Layer*, which is in charge of virtualizing compute, storage and network resources, and to offer virtual slices to the tenants operating over the 5G-XHaul infrastructure. Consequently, the *Tenant Control Layer* is hosted by each tenant and contains the control logic, i.e. tenant specific SDN controllers, which operate over the virtual tenant slices; Section III-E provides an overview of the abstraction that the 5G-XHaul system offers to the tenant control layer. Finally, the *Management and Service Orchestration* layer contains the higher level orchestrators, VNF and virtual infrastructure managers that provide each tenant with holistic control over its compute and network resources.

In the next sections we will describe in more detail how 5G-XHaul proposes to address the control plane aspects of the managed physical infrastructure layer, as well as the network virtualization aspects of the infrastructure management layer. The interested reader is referred to [8] for an more detailed description and an initial performance evaluation of the 5G-XHaul layered architecture.

III. 5G-XHaul control plane design

A. Overall control plane architecture

The 5G-XHaul control plane architecture that is based on the following design principles:

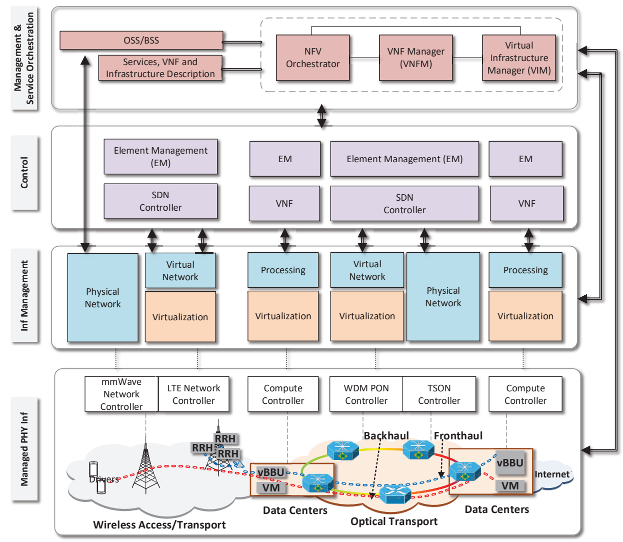

Fig. 1. 5G-XHaul layered architecture from [8]

i) Full address space virtualization is offered through an overlay, implemented using encapsulation at the edge of the transport network. This means that different tenants can use overlapping L2 or L3 address spaces.
ii) Data plane scalability is achieved by isolating the forwarding tables of the transport network elements inside the 5G-XHaul infrastructure from any tenant related state (overlay). This is again achieved by encapsulating tenant frames at the edge of the network into transport specific tunnels.
iii) Scalability of the SDN control plane is achieved introducing the concept of *areas*. An area defines a set of transport network elements that are under the control of a logically centralized SDN controller[1]. A control plane hierarchy is introduced whereby higher level controllers are used to coordinate the actions of area level controllers.
iv) Finally, the vision of converged heterogeneous technology domains, e.g. wireless and optical segments in the transport network, is enabled by: i) the previously introduced areas, which embody a single type of transport technology (e.g. wireless mesh, optical or Ethernet), and ii) a transport adaptation function that maps the per tenant traffic at the edge nodes to the transport specific tunnels of a given area.

In order to support the previous principles, three types of transport nodes are defined in 5G-XHaul. These are depicted in Figure 2. First, *Edge Transport Nodes* (ETNs), connect the tenant VNFs to the 5G-XHaul transport network, maintain the corresponding per-tenant state, and encapsulate tenant traffic into transport specific tunnels. Second, *Inter-Area Transport Nodes* (IATNs), support the necessary functions to connect different areas, which may be implemented using different

[1] In practice we could have a controller cluster with a mechanism to synchronize state between instances.

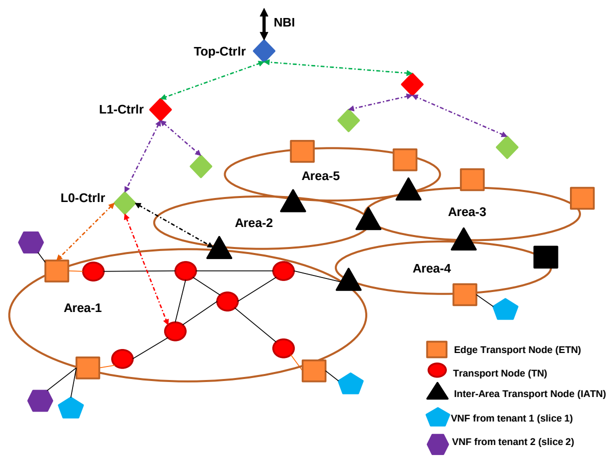

Fig. 2. 5G-XHaul control plane architecture

transport technologies. Finally, regular *Transport Nodes* (TNs), support an area specific transport technology, and provide forwarding services between the ETNs and IATNs of that area. The interested reader is referred to [16] for an in-depth introduction to the 5G-XHaul control plane architecture.

B. Overlay: ETNs

ETNs maintain per-slice state providing 5G-XHaul tenants the required abstraction to operate on their slices. In particular, ETNs host the VNFs and logical datapaths defined in each tenant's slice, as depicted in Figure 4. In this regard, we introduce the notions of *tenant ID* (T-Id) and *slice ID* (S-Id), which are globally unique identifiers for the end-to-end slices instantiated by a tenant, as later discussed in Section III-D; e.g. *opA.slice1*, where *opA* is the tenant ID for operator A and *slice1* is the slice ID of a slice that operator A wants to deploy for a specific service. There is thus a 1:N relationship between tenant and slice IDs. While the above identifiers need to be globally unique in the control plane, in the data plane different local transport slice IDs may be used in each 5G-XHaul area. For example a *Transport Slice ID* (TrSlice-Id) is used to represent in the data plane a given *T-Id.S-Id*, where *TrSlice-Id* may be different in each 5G-XHaul area (c.f. Figure 2). Notice that having a notion of slice ID at the data plane is useful in order to easily deploy policies at the tenant or slice level, e.g. deploying an Access Control List (ACL) that binds all the traffic for a tenant to a given QoS class, or that drops all the traffic of a malfunctioning slice. Thus, a function is embedded in the ETNs that performs the mapping between the global slice IDs of the control plane and the local Transport Slice ID in the data plane. In particular, an ETN embeds three major functions that are discussed next, namely: i) per-tenant Logical Datapaths, ii) a Forwarding Information Base (FIB), and iii) a Transport Adaptation Function (TAF). These components are illustrated in Figure 3.

As depicted in Figure 4 an ETN may host logical datapaths for a set of tenants. Logical datapaths receive high level control policies from the tenant's own control plane, and push those policies to a local SDN controller in the ETN. As mentioned in Section III-E, logical datapaths represent an implementation of the tenant control layer depicted in Figure 1. The local controller in the ETN obtains the rules from each logical datapath, adds appropriate context, and pushes the rules to the ETN forwarding information base (FIB). Introducing multi-tenancy support at this last controller layer, which runs on the actual forwarding element, leaves intact the underlying performance-centric multi-stage FIB architecture design that performs the actual packet forwarding [11]. Hence, a critical aspect in the design of the ETN is the datapath delay introduced by the ETN FIB. In order to minimize this delay, tenant-specific rules are often evicted, resulting in a small number of simultaneous rules hosted in the FIB. Consequently, a hierarchical structure of rule caches is used to scale up to a large number of rules coming from different tenants. The design of appropriate caching strategies is an open research area [12]. Consequently, the FIB matches tenant-specific rules and inserts packets into transport-specific tunnels, which are pre-instantiated and it is expected that traffic from multiple slices can be combined into a single transport tunnel.

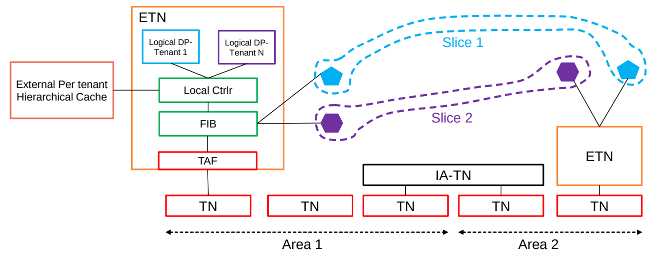

Fig. 3. Detailed overview of an Edge Transport Node (ETN)

As mentioned in section III-A an ETN uses encapsulation to isolate transport network elements from per-tenant related state. A Transport Network adaptation Function (TAF) is included in the ETN that pushes the corresponding transport header before injecting the packets into the transport network. The transport header signals three major pieces of information: i) the path to be followed by a transport tunnel in the area, ii) the local Transport Slice ID, and iii) the QoS allocated to that tunnel. Each ETN features a TAF corresponding to the transport technology used in the 5G-XHaul area where the ETN is located. As a matter of example, a transport adaptation function based on Ethernet is discussed next.

The Ethernet TAF provides MAC-in-MAC encapsulation as defined in Provider Backbone Bridging (PBB) [14]. In the Ethernet TAF the path to be followed by a packet is signaled using the MAC address of the final destination ETN, which may be located in the same or a different area. In the latter case, the packet should be first delivered to an IATN in the same area, which is signaled using a segregated space from the outer VLAN tag. The outer VLAN field is also used for load balancing when multiple paths are available to the destination ETN/IATN in the same area. The Transport Slice Id is signaled in the Ethernet TAF using the 24 bit I-SID field in the outer

MAC header. QoS classes can be signaled using the priority bits in the outer VLAN tag. Notice that in the Ethernet TAF the QoS parameters from a slice can be mapped to the outer Ethernet header, thus enabling transport nodes to provide an appropriate treatment. In addition, as in any tunneling solution, VNFs should adjust path MTUs to avoid packet fragmentation.

Finally, regarding implementation, ETNs are assumed to be closely located to the IT infrastructure hosting the operator VNFs, e.g. edge or regional data centers. Thus, one can envision ETNs implemented as a software agent inside a server hypervisor hosting VNFs, or as an external gateway device if a hardware based implementation is preferred.

C. Underlay: TNs and IATNs

Transport Nodes (TNs) connect ETNs and IATNs within a given 5G-XHaul area (c.f. Figure 2). The concept of a TN is technology agnostic, thus a TN could be represented by a mmWave wireless node at the street level, by an Ethernet switch at the access or metro segments, or by an active optical node at the metro network (e.g. TSON [7]). Regardless of the actual technology, in 5G-XHaul a TN offers a dataplane abstraction where forwarding, along with some other primitives like bandwidth provisioning or reliability, can be programmed by a logically centralized control plane.

Since the set of ETNs and IATNs available in a given area are fairly static, 5G-XHaul assumes that transport tunnels between ETNs/IATNs in an area are pre-provisioned. A pre-provisioned transport tunnel means that the ETNs and IATNs in that area have an interface representing such tunnel, and that the required TNs have the corresponding entries in their FIBs. Notice however that pre-provisioned transport tunnels do not need to be static, as tunnels can be reconfigured by the control plane in case of network situation changes. For example, the control plane may switch down a set of TNs for energy saving, while relocating all the affected transport tunnels to other TNs; the ETNs would be agnostic to such relocation. In addition, pre-provisioned tunnels may be point to point tunnels, or multicast trees connecting a set of ETNs/IATNs in a given area. In the case of a multicast tree, the transport technology in the 5G-XHaul area needs to support packet replication along the interfaces participating in each multicast group. In 5G-XHaul multicast group membership is managed by the logically centralized control plane. In the case of a 5G Mobile Network a multicast tree can be useful for example to connect base stations implementing a cooperative transmission scheme. Transport tunnels are associated to a set of transport classes. In particular, 5G-XHaul has proposed a set of four transport classes, described in [15], dimensioned to transport fronthaul traffic, backhaul traffic, as well as traffic resulting from other functional splits. Thus, multiple transport tunnels to a given ETN/IATN may be pre-provisioned representing the different transport classes.

Inter-Area Transport Nodes (IATN) provide connectivity between neighboring 5G-XHaul areas (c.f. Figure 2). As illustrated in Figure 3, an IATN can be understood as an interconnection function sitting above one TN for each area being connected by the IATN. The different areas can use the same or different transport technologies. The IATN interconnection function contains a control plane function and a data plane function that are described next.

In the control plane, an IATN needs to discover the areas that it has access to, and convey the identifier of these areas to the 5G-XHaul control plane. In addition, a unique identifier is required for an IATN that also needs to be conveyed to the control plane. This information is required by the control plane to be able to allocate paths at the area level. More detail on control plane functions are discussed in section III-D

In the data plane, an IATN needs to implement the forwarding principle used in each of its connected areas. Thus, an IATN includes the corresponding TAF (described in section III-B) for each area it connects. IATNs maintain their own FIB function that maps tunnels from one area to tunnels of another area. In case, an IATN interconnects areas belonging to the same technology, technology specific optimizations are possible to accelerate the datapath that are currently being investigated.

D. Hierarchical Controller design

In 5G-XHaul the control plane is composed of a hierarchy of controllers as illustrated in Figure 2. The top level controller, hereafter referred to as the *Top* controller is responsible for provisioning per tenant slices, and orchestrating the required connectivity across different 5G-XHaul areas and domains (e.g., optical transport domain, wireless transport domain). At the lowest level of the hierarchy we find the *Level*-0 controller that is responsible for the provisioning and maintenance of transport tunnels between ETNs and IATNs of a given 5G-XHaul area; a Level-0 controller operates at the level of individual network elements. A set of Level-0 controllers are logically organized under a Level-1 controller, which is in charge of maintaining connectivity between the corresponding Level-0 areas, and operates with a higher level of abstraction, namely maintains state at the area level instead of maintaining state for each network element as Level-0 controllers do. Notice that the proposed architecture is recursive in the sense that a Level-i controller can be defined to coordinate a set of Level-$(i-1)$ controllers, where coordination may include among others adapting QoS parameters between different areas, or breaking end to end QoS requirements into specific requirements for each area. Hereafter, and without loss of generality, we assume a three level hierarchy consisting of Level-0, Level-1 and Top controllers.

Dimensioning the number of network elements under a Level-0 controller, or the number of Level-0 controllers under a Level-1 controller depends on many factors and is an area of active research. A Level-0 controller is assumed to be in charge of an area instantiating a single type of transport technology, i.e. a mmWave area, an Ethernet area, or an active optical area. Thus, the number of elements under a Level-0 controller will be very dependent on the particular technology. For example, for scalability reasons a large number of mmWave transport nodes deployed at the street level can be partitioned into a plurality of areas and Level-0 controllers, whereas the optical switches composing the metro network and the corresponding Ethernet clients could be controlled respectively by a different Level-0 controller. Notice that having technology specific controllers allows to develop solutions tailored to the control plane of each technology. In addition, Level-0 controllers enforce QoS in each area according to the transport technology being used. Upper level controllers, i.e. Level-1 and Top controllers, do not need to be technology specific since they operate at a higher abstraction level (the area level). In practice for scalability reasons controllers at each level will be deployed in clusters of synchronized controllers.

The major functionality carried out at each controller level is illustrated through an example. Consider a tenant defining a slice according to the abstraction described in Section III-E. The tenant indicates the ETN where each VNF included in the slice is connected, where the selected ETNs may be located in different 5G-XHaul areas (c.f. Figure 2). Thus, once defined, the slice is submitted to the 5G-XHaul Top controller through a north bound interface (NBI). The responsability of the 5G-XHaul control plane is then to wire the transport tunnels connecting the ETNs involved in the tenant's slice. The first task of the Top controller is to look up the Level-1 controllers in charge of the ETNs included in the layer two segments $s \in S$ defined in the received slice. Once the Level-1 controllers are identified, the Top controller runs a path allocation algorithm to establish a path between all the ETNs participating in the same layer two segment. The path determined by the Top controller is expressed as a set of Level-1 controller areas, whereby a Level-1 controller area is composed of all the 5G-XHaul areas (c.f. Figure 2), where the corresponding Level-0 controller is controlled by the Level-1 controller. For each determined path, the Top controller requests the involved Level-1 controllers to allocate a transport connection between ETNs under their control, or between an ETN and a neighboring Level-1 area. In order to determine these paths, the Level-1 controllers run a path allocation algorithm that returns the set of 5G-XHaul areas belonging to the path, along with the corresponding Level-0 controllers in charge of each area. Consequently, for each path, the Level-1 controller submits a request to the corresponding Level-0 controller, which runs a path allocation algorithm to identify the transport tunnels and paths connecting each involved ETN and IATN within the 5G-XHaul area under its control. Once the process completes, the tenant slice is fully connected and communication may begin. Table I illustrates the high level functionality included at each controller level.

E. Tenant Transport Network Abstraction

A network virtualization technology allows a tenant to instantiate a virtual network connecting its distributed virtual network functions (VNFs). Notice that in the case of 5G-XHaul a VNF could as well represent a (subset of a) base station. In the domain of cloud computing, a typical virtual network abstraction is that of a layer two switch directly connecting the tenant's VNFs, e.g. [10]. This abstraction

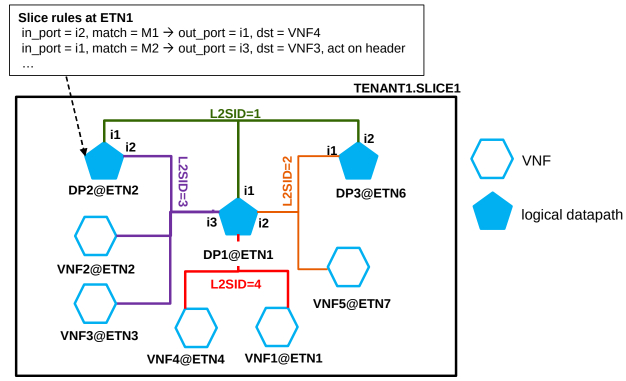

Fig. 4. Tenant Transport Network Abstraction in 5G-XHaul

although simple for the tenant, comes at a cost for the infrastructure owner, who needs to accommodate for each tenant the unconstrained *any to any* communication patterns enabled by this abstraction, hence consuming significant resources on the physical substrate. In [13] it is shown that a more efficient embedding is possible if the tenant abstraction declares information about the expected communication patterns between its VNFs.

The goal of 5G-XHaul is to define a transport network architecture connecting the VNFs of tenants offering 4G or 5G connectivity services. To effectively support multi-tenancy in this environment, the communication patterns imposed by the mobile network should be exploited. For example, in 4G, most of the traffic generated by base stations is addressed to the packet gateways in the core network, however low delay direct connections between neighboring base stations may be beneficial for handover or interference coordination signaling. In 5G we expect the following changes in the communication patterns: i) the amount of local traffic between neighboring base stations will increase to enable more demanding interference coordination techniques, ii) native support for multicast to a group of base stations may be beneficial to support co-operative transmission schemes, and iii) core packet gateways will be virtualized and possibly distributed to regional data centers, which will result in a more distributed traffic matrix over the transport network [9].

In order to address the previous communication patterns 5G-XHaul proposes the tenant abstraction, or slice definition, depicted in Figure 4. A tenant defines a set of layer two segments, $S = \{s_1, ..., s_N\}$, where each segment in S is meant to directly connect a subset of the tenant's VNFs. Figure 4 illustrates layer two segments in different colors, and assigns to each segment a unique identifier referred to as layer two segment ID (L2SID). Each segment $s_i \in S$ is associated with QoS parameters, such as a peak bandwidth B_i and maximum latency L_i, defining the constraints of that layer two segment. The underlaying Level-0 area controllers are responsible for enforcing the appropriate QoS for each slice on the physical transport network. Each VNF in the tenant slice, depicted with hexagon shapes in Figure 4, is associated to a single L2 segment $s_i \in S$. In addition, the

TABLE I
MAIN FUNCTION OF 5G-XHAUL CONTROLLERS AT DIFFERENT LEVELS

Level	Function	Description
Level-0	Path Allocation	Allocate tunnel path inside the 5G-XHaul area
	Topology Management	Maintain topology within 5G-XHaul area
	QoS & OAM	Maintain per tunnel statistics and OAM metrics
	ETN end point discovery	Discover per tenant end points connected to an ETN in the area
	IATN Discovery	Discover IATNs in the 5G-XHaul area, and the areas they connect
	NBI to Level-1	North Bound Interface to Level 1 controller
Level-1	Inter-Area Path Allocation	Allocate paths between ETNs at area level
	Inter-Area Topology Management	Maintain connectivity graph between 5G-XHaul areas
	Area level QoS & OAM	Maintain area level QoS and OAM metrics
	NBI to Top controller	North Bound Interface to Top controller
Top	Inter-Level 1 Path Allocation	End to End path allocation at Level-1 controller level
	NBI to service	North Bound Interface to service / VIM / orchestrator
	Tenant and Slice management	Generation and assignment of unique slice IDs and Tenant IDs

tenant abstraction allows to define *logical datapaths* (DPs), illustrated with solid pentagon shapes in Figure 4, which may have multiple interfaces, each interface connecting to a different L2 segment $s_i \in S$. Logical datapaths are used to control the forwarding state between VNFs in the tenant slice, according to the tenant's own control logic. In particular, logical datapaths host the custom control state defined by each tenant's control plane, thus implementing the *Tenant Control Layer* depicted in Figure 1[2]. The tenant control plane interacts with its logical datapaths through the 5G-XHaul northbound interface described in Section III-D. Figure 4 provides an example of the control rules that can be pushed by the tenant's control plane into the logical datapaths. In the next section we will describe how the state required to maintain the described per tenant slices, is embedded into the 5G-XHaul physical infrastructure in order to provide scalability.

IV. INTERACTIONS BETWEEN THE TRANSPORT NETWORK AND THE MOBILE NETWORK

The goal of 5G-XHaul is to create a transport infrastructure destined to serve 5G mobile network operators. Therefore, in most cases the slices instantiated by the 5G-XHaul tenants will be connecting mobile related VNFs, for example base stations with elements of the mobile packet core. In this context, we envision that an important service provided by 5G-XHaul to a mobile network slice is the ability to interact with the transport network in a more tightly coupled way than what is possible in 4G, in order to improve Quality of Experience (QoE).

In 4G, and previous generations, the mobile network is composed of the Radio Access Network (RAN) and the Core Network (CN). In the case of 4G, RAN and CN communicate over a transport network, but the only assumption about the transport is that it is an IP network. Thus, the transport is unaware that is carrying packets between base stations in the RAN and packet gateways in the CN. In particular, base stations and packet gateways set up IP tunnels over the transport in order to communicate with each other. Such tunnels need to be updated for example when a mobile device hands over between base stations. For security reasons the IP tunnels used in the mobile network are often encrypted, which refrains the transport nodes from becoming aware of the type of radio traffic being carried in the tunnel in order for example to deliver a tailored treatment. Instead, 4G defaults to standard IP QoS mechanisms such as DSCP markings. In 5G-XHaul we advocate that more open interfaces between the mobile and transport network can be beneficial for a number of reasons, which we discuss next.

[2]In Figure 4 tenant rules are represented with a generic match M.

A. Use cases for information exchange between the Mobile and the Transport Networks

Unlike in 4G, where it is assumed that the transport network can be overprovisioned, in 5G the mobile network design, and in particular the RAN design, needs to consider the performance of the transport network. In particular, the characteristics of the transport network will dictate the optimal allocation of RAN signal processing functions between a Remote Unit (RU) and Centralized Unit (CU) [15], which is critical to implement cooperative and interference mitigation techniques that increase spectral efficiency. Thus, parameters like the transport network available rate and latency are key in the selection of the RAN configuration to be used.

We next present a set of use cases that motivate the need for a mobile network and transport information exchange:

a) Proactive congestion avoidance: Lack of coordination between the mobile network and the transport may result in the mobile network triggering a handover to a target base station that is then connected to a congested link in the transport network. In order to avoid these situations, the mobile network must be aware of the transport network congestion levels when triggering handovers between cells.

b) Load balancing: Information about the RAN enables the transport network to more effectively balance the traffic load between the antenna sites and mobile network functions across different paths. This results in a better utilization of resources within the transport, as well as in an overall improvement of QoE.

c) Fairness: Currently, transport networks do not offer the same granularity in QoS profiles as the mobile network provides. Transport networks are therefore unable to distinguish among different types of traffic, which may result in unfairness, or policy violations, upon congestion. A mobile network-transport information exchange would for example allow to appropriately re-classify mobile traffic into transport QoS classes in order to comply with the policies defined by the mobile network operator at all times.

d) Self-backhauling: Early 5G deployments require means for incremental deployment as initially the density of 5G base stations with dedicated backhauling would be limited. A useful technique which can be beneficial in future systems is self-backhauling. The support of wireless self-backhauling is a technique studied by some 5G RAN proposals. If such capabilities are available, RAN and transport should coordinate to decide when it is best to make use of self-backhauling.

e) Energy Saving: If operating in isolation, the mobile network and the transport may take conflicting decisions when trying to minimize energy consumption by independently switching off RAN and transport nodes. Energy efficiency is a clear example where RAN-transport coordination is required for a global system optimization.

f) Cell-less RAN architecture: Some 5G RAN proposals operating at very high frequencies, where blockage and path loss due to NLoS are very significant, are studying the possibility of not having a mobile device exclusively attached to a single base station, but rather be able to receive/transmit data from/to different base stations according to channel measurements performed by the mobile device [9]. Changing points of attachment at such short time scales requires a very tight coordination between mobile network and transport in order to quickly reconfigure the downlink and uplink paths.

B. Envisioned types of interfaces between the Mobile and Transport networks

The proposed 5G-XHaul transport relies on a mix of legacy and new technologies. The information regarding the type of resources and how to share this information between the mobile and transport networks depends on the specific transport service, the deployment scenario, the radio deployment architecture, and on the choice of the transport technology. Hence, it would be desirable to achieve an information-sharing model, where each domain (mobile network and transport) manages the information to be shared with the other, and prescribes how to use that information.

The trend towards adopting a logically centralized control plane in 5G, both for the transport and the mobile network, is a step towards enabling this information exchange, because an SDN controller naturally collects information about the network state that can be shared with other domains. Consequently, effective mechanisms for inter-domain policy negotiation need to be established to decide which domain takes precedence upon a certain network condition. An example is the following. In case of congestion, two actions are possible: redistribution in the RAN and rerouting in the transport. In the case of redistribution, the RAN circulates traffic around, effectively load-balancing in an optimal way across the transport network. In the case of rerouting, the transport network uses a number of techniques, like SDN-based traffic engineering, to make better use of available alternate paths.

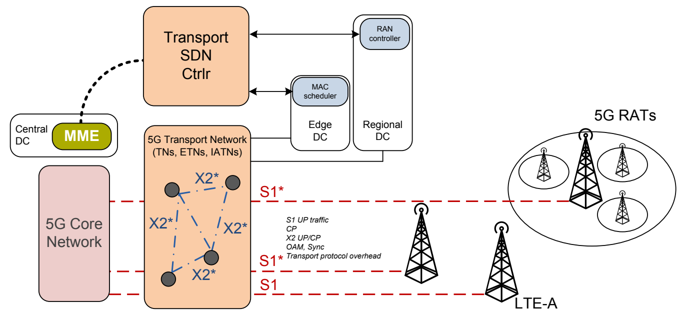

Fig. 5. Envisioned interfaces between mobile network and transport.

Three main types of mobile network-transport interfaces depicted in Figure 5 are identified to support the use cases introduced in section IV-A:

i. An interface between the transport controller and the Mobility Management Entitiy (MME[3]). The MME is an entity in the CN that is aware of the current position of mobile devices at the cell level, for RRC connected devices, and at the tracking area level, for RRC idle devices. The MME is also involved in handover preparation. Hence, the MME could provide the transport controller with statistics about aggregate mobility behaviors in the RAN, or about the cell level trajectory of a given mobile device.
ii. An interface between the transport controller and a RAN controller. Unlike the MME, a RAN controller is aware of radio conditions and is in charge of managing cooperative and interference coordination techniques, for which the support of the transport network may be required. In principle, the information exchanged through this interface has a shorter lifespan than the information exchanged with the MME.
iii. An interface between a centralized MAC scheduler[4] of a given (group of) base station(s), and the transport controller. The information exchanged through this interface has the shortest life span. This interface would be for example required to support a cell-less RAN architecture.

V. State of the Art

The authors in [17] introduce an architecture for optical networks that allows a transport service provider to offer a range of network abstractions to its tenants. In particular the paper discusses the management and implementation trade-offs of a *big switch*, an *abstract link* and a *direct* network abstraction models. Unlike [17] the virtualization solution proposed in 5G-XHaul is an overlay that pushes per-tenant state to the edge, while isolating transport nodes from any tenant related state (e.g. tenant rules). The overlay solution is

[3]Or equivalent entitiy in 5G

[4]In 4G, the MAC scheduler is a function inside the RAN protocol stack in charge of scheduling packets from different radio bearers on the available radio resources. An equivalent meaning is assumed for 5G.

deemed more scalable for heterogeneous transport networks, which might comprise transport nodes with limited resources. The work in [18] proposes a control plane architecture for an integrated C-RAN and DWDM transport network, where an overall orchestrator sits on top of a radio and a transport controller. The architecture in [18] does not directly address multi-tenancy, but studies the impact of different optical network abstraction models on the resource allocation decisions taken by the orchestrator. In particular, the paper evaluates the trade-off between blocking probability and control overhead for transport network abstraction models with increasing levels of detail. The network abstraction models proposed in [18] can be used in 5G-XHaul by a Level-0 controller to export its area topology towards a Level-1 controller. In 5G-XHaul though network abstraction models for different wireless and optical technologies need to be studied. Finally, [19] introduces a hierarchical SDN based architecture for C-RAN, cloud and a DWDM transport, while discussing several deployment models. The 5G-XHaul control plane architecture introduced in this paper shares the same design principles as the one defined in [19], but differs in that it considers a transport network composed of several wireless and optical technologies, thus enabling centralized and distributed RAN deployments, instead of a DWDM based transport for C-RAN.

VI. Conclusions

Network slicing is a major trend driving the design of 5G networks. Slicing is enabled by the concepts of virtualization and softwarization, which have revolutionized the IT and cloud computing domains in the past years. In this paper, we have reviewed how these trends will affect the design of future 5G transport networks. In particular, we have presented the initial control plane design of the 5G-XHaul project, a collaborative project within the umbrella of the 5G-PPP initiative focusing on the design on transport networks for 5G. The presented control plane allows an operator to control in an unified manner a transport network infrastructure composed of heterogeneous technology domains, including wireless and optical network segments. In addition, the 5G-XHaul control plane enables virtualization and slicing of heterogeneous transport resources in a scalable way, by maintaining per-tenant state only at the network edge. Finally, the proposed control plane and virtualization framework allows tenants to externally control the virtual resources in their slices.

VII. Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 671551. The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the view of its authors only.

References

[1] 5GPPP, 5G empowering vertical industries. Online. Available here: https://5g-ppp.eu/wp-content/uploads/2016/02/BROCHURE_5PPP_BAT2_PL.pdf

[2] NGMN Alliance, NGMN 5G White Paper, White paper, Feb. 2015.

[3] 5GPPP Architecture Working Group, View on 5G Architecture, June 2016. Online. Available here: https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-For-public-consultation.pdf

[4] China Mobile NGFI White paper: China Mobile, ALU, Nokia, ZTE, Broadcom, Intel, White Paper of NGFI (Next Generation Fronthaul Interface) Version 1.0, October 4, 2015

[5] 5G-XHaul project. Online. Available here: http://www.5g-xhaul-project.eu/

[6] ETSI NFV ISG, ETSI Network Functions Virtualisation (NFV) Industry Standards (ISG) Group Draft Specifications, http://docbox.etsi.org/ISG/NFV/Open, December 2014.

[7] Zervas, G. S., Triay, J., Amaya, N., Qin, Y., Cervello-Pastor, C., Simeonidou, D. (2011). Time Shared Optical Network (TSON): a novel metro architecture for flexible multi-granular services. Optics express, 19(26), B509-B514.

[8] A. Tzanakaki et al., 5G Infrastructures Supporting End-User and Operational Services: The 5G-XHaul Architectural Perspective, Accepted to IEEE ICC 2016, Workshop on 5G Architecture, Kuala Lumpur, Malaysia, May 2016.

[9] March. P., et.al., Preliminary Views and Initial Considerations on 5G RAN Architecture and Functional Design, March 2016. Online. Available here: https://metis-ii.5g-ppp.eu/documents/white-papers/

[10] Mudigonda, Jayaram, et al. "NetLord: a scalable multi-tenant network architecture for virtualized datacenters." ACM SIGCOMM Computer Communication Review. Vol. 41. No. 4. ACM, 2011.

[11] Pfaff B, Pettit J, Koponen T, Jackson E, Zhou A, Rajahalme J, Gross J, Wang A, Stringer J, Shelar P, Amidon K. The design and implementation of open vswitch. In12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15) 2015 (pp. 117-130).

[12] Katta, Naga, et al. "Infinite cacheflow in software-defined networks." Proceedings of the third workshop on Hot topics in software defined networking. ACM, 2014.

[13] Ballani, H., Costa, P., Karagiannis, T., and Rowstron, A. (2011, August). Towards predictable datacenter networks. InACM SIGCOMM Computer Communication Review(Vol. 41, No. 4, pp. 242-253). ACM.

[14] Salam, Samer, and Ali Sajassi. "Provider backbone bridging and MPLS: complementary technologies for next-generation carrier ethernet transport." Communications Magazine, IEEE 46.3 (2008): 77-83.

[15] Deliverable D2.1 Requirements Specification and KPIs, 5G-XHaul Project, March 2016.

[16] Deliverable D3.1 Analysis of state of the art on scalable control plane design and techniques for user mobility awareness. Definition of 5G-XHaul control plane requirements, 5G-XHaul Project, March 2016.

[17] Autenrieth, A., Szyrkowiec, T., Grobe, K., Elbers, J. P., Kaczmarek, P., Kostecki, P., Kellerer, W. (2014, May). Evaluation of virtualization models for optical connectivity service providers. In Optical Network Design and Modeling, 2014 International Conference on (pp. 264-268). IEEE.

[18] Fiorani, Matteo, Ahmad Rostami, Lena Wosinska, and Paolo Monti. "Transport abstraction models for an SDN-controlled centralized RAN." IEEE Communications Letters 19, no. 8 (2015): 1406-1409.

[19] hln, Peter, Bjrn Skubic, Ahmad Rostami, Matteo Fiorani, Paolo Monti, Zere Ghebretensa, Jonas Mrtensson, Kun Wang, and Lena Wosinska. "Data plane and control architectures for 5G transport networks." Journal of Lightwave Technology 34, no. 6 (2016): 1501-1508.