Wireless-Optical Network Convergence: Enabling the 5G Architecture to Support Operational and End-User Services

Anna Tzanakaki[(1)(12)], Markos Anastasopoulos[(1)], Dimitra Simeonidou[(1)], Ignacio Berberana[(2)], Dimitris Syrivelis[(3)], Paris Flegkas[(3)], Thanasis Korakis[(3)], Daniel Camps Mur [(4)], Ilker Demirkol[(5)], Jesús Gutiérrez[(6)], Eckhard Grass[(6)], Qing Wei [(7)], Emmanouil Pateromichelakis[(7)], Nikola Vucic[(7)], Albrecht Fehske [(8)], Michael Grieger [(8)], Michael Eiselt[(9)], Jens Bartelt[(10)], Gerhard Fettweis[(10)], George Lyberopoulos[(11)], Eleni Theodoropoulou[(11)]

(1) HPN Group, University of Bristol, UK, (e-mail: anna.tzanakaki@bristol.ac.uk), (2)Telefonica Investigacion Y Desarrollo, Madrid, ESP, (3) University of Thessaly, Volos, GRC, (4) i2CAT Foundation, ESP, (5) Univeristat Politecnica de Catalunya, ESP, (6) IHP GmbH, DE (7) Huawei Technologies Duesseldorf GmbH, DE, (8) Airrays GmbH, Dresden, DE, (9) ADVA Optical Networking SE, (10) Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, DE, (11) COSMOTE Mobile Communications S.A, R & D Dept., Fixed & Mobile, GRC, (12) National and Kapodistrian University of Athens, Department of Physics, GRC

Abstract— **This paper presents a converged 5G network infrastructure and an overarching layered architecture, to jointly support operational network and end-user services, proposed by the EU 5GPPP project 5G-XHaul. The 5G-XHaul infrastructure adopts a common fronthaul/backhaul network solution, deploying a wealth of wireless technologies and a hybrid active/passive optical transport, supporting flexible fronthaul split options. A novel modeling framework has been developed to evaluate the performance of the 5G-XHaul infrastructure. Our modeling results show that the proposed architecture can offer significant energy savings, but there is a trade-off between overall energy consumption and end-user service delay.**

Keywords—5G, backhauling, fronthauling, small cells, C-RAN

I. INTRODUCTION

It has been predicted that global mobile data traffic will increase by a factor of eight between 2015 and 2020. This enormous growth is attributed to the rapidly increasing: a) number of network-connected end devices, b) Internet users with heavy usage patterns, c) broadband access speed, and d) popularity of applications such as cloud computing, video, gaming etc. It is clear that to meet this enormous growth of mobile traffic demands, the traditional single layer macro-cells needs to be transformed to an architecture comprising a large number of smaller cells. Traditional Radio Access Networks (RANs), where Base Band Units (BBUs) and radio units

are co-located suffer several limitations including: i) increased CAPEX and OPEX due to lack of resource sharing, ii) limited scalability and flexibility, iii) lack of modularity and limited density, iv) increased management costs, and v) inefficient energy management.

To address these limitations, Cloud Radio Access Networks (C-RANs) have been recently proposed. In C-RAN distributed Access Points (APs), referred to as remote units (RUs), are connected to a BBU pool, the Central Unit (CU), through high bandwidth transport links known as fronthaul (FH). The interface between RUs and CU is standardized through the Common Public Radio Interface (CPRI), the Open Base Architecture Initiative (OBSAI) and the Open Radio Interface (ORI), with CPRI currently being the most widely used standard. FH is responsible to carry the RU wireless signals, typically over an optical transport network, using either digital transmission (e.g. CPRI), or analog transmission (radio-over-fiber). The adoption of digitized transmission solutions offers reduced signal degradation allowing data transmission over longer distances, offering higher degree of CU consolidation. One of C-RAN's main disadvantages is the very high transport bandwidth required to carry the sampled radio signals as well as strict latency and synchronization constraints. As an example, the CPRI protocol for an LTE system with 20 MHz bandwidth 2x2 MIMO sector, requires 2.46 Gbps capacity for RU-CU interconnection (CPRI line bit rate option 3) and this may further increase to 12.165 Gbps, for CPRI line bit-rate option 9 [1]. Existing mmWave E-Band and optical transport solutions supporting traditional backhaul (BH) requirements are based on Passive Optical Networks (PON), Gigabit-capable PON (GPON) or 10GE technologies offering capacities up to 10 Gbps. Considering that these transport solutions will also need to support FH functions, they can rapidly become the bottleneck. Recognizing the benefits of the C-RAN architecture and the associated challenges, equipment vendors are expanding their mobile FH solutions adopting more effective wireless technologies ((i.e. operating in the Sub-6 GHz and 60GHz frequency bands enhanced with advanced beamtracking and MIMO techniques)), new versatile Wavelength Division Multiplexing (WDM) optical network platforms [2] as well as novel control and management frameworks that allow service driven customization offering increased granularity, end-to-end optimization and guaranteed QoS.

To relax the stringent FH requirements of C-RAN architectures, while taking advantage of its pooling and coordination gains, solutions relying on FH compression as well as on alternative architectures adopting flexible functional splits (Figure 2) have been proposed [3], [4]. In the latter case, the introduction of flexible splits allows dividing the processing functions between the CU and the RU. Based on these solutions, a set of processing functions is performed at the RU and the remaining functions are performed centrally. In the majority of the existing solutions, these functions are implemented via closed and specific purpose hardware,

which introduces significant installation, operational and administrative costs. To address these issues, the concept of *network softwarisation* that enables migration from the traditional closed networking model to an open reference platform able to instantiate a variety of network functions has been recently proposed. A typical example includes the OpenAirInterface (OAI) i.e. an open source 4G/5G radio stack able to be executed on general purpose servers hosted in data centers (DCs) [5]. Such open source frameworks are still in early development stages and do not allow execution of more complex functionalities such as flexible RAN splits. In this study, the concept of flexible functional splits is addressed through a combination of small scale servers (cloudlets) and relatively large-scale DCs placed in the metro areas. However, flexible splits impose the requirement of fine bandwidth granularity and elastic resource allocation at both the wireless and the optical transport network, while the support of remote processing, demands high bandwidth transport connectivity between the RUs and the remote compute resources at the CU.

In response to these observations, a converged optical-wireless 5G network infrastructure interconnecting compute resources with fixed and mobile users is proposed, to support both operational network (C-RAN) and end-user services. In this context, operational network services refer to services required for the operation of the 5G infrastructure with FH services, offered to infrastructure operators/providers, being a representative example. On the other hand, end-user services refer to services provided to end users (e.g. content delivery, gaming etc) that in 5G environments require BH connectivity, referred to in this paper as BH services. This infrastructure is being developed in the framework of the EU-funded HORIZON 2020 5GPPP project 5G-XHaul. The main technical innovations of the proposed solution include i) an overarching architectural framework inspired by the ETSI Network Function Virtualization (NFV) standard and the Software Defined Networking (SDN) open reference architecture [6] that supports jointly BH and FH services and the concept of flexible functional splits. This is a key difference between the proposed architecture and contemporary LTE-A systems, where FH and BH services are supported through separate networks, while network control and management is closed, ii) introduction of a novel data plane architecture converging heterogeneous wireless as well as passive and active optical network technologies to support the overarching architecture and its requirements, iii) development of a novel multi-objective optimization (MOP) modeling framework to evaluate the performance of the proposed approach. This includes a service provisioning model used to study a variety of FH and BH options. The overall objective of the model is twofold: a) to minimize the operational expenditure related with FH services in terms of power consumption, under strict delay constraints and b) to minimize end-to-end service delay of BH services.

HeNB: Home eNodeB   vBBU: Virtual Base Band Unit   EPC: Evolved Packet Core

VM: Virtual Machine   RU: Remote Unit   S GW: Serving GateWay
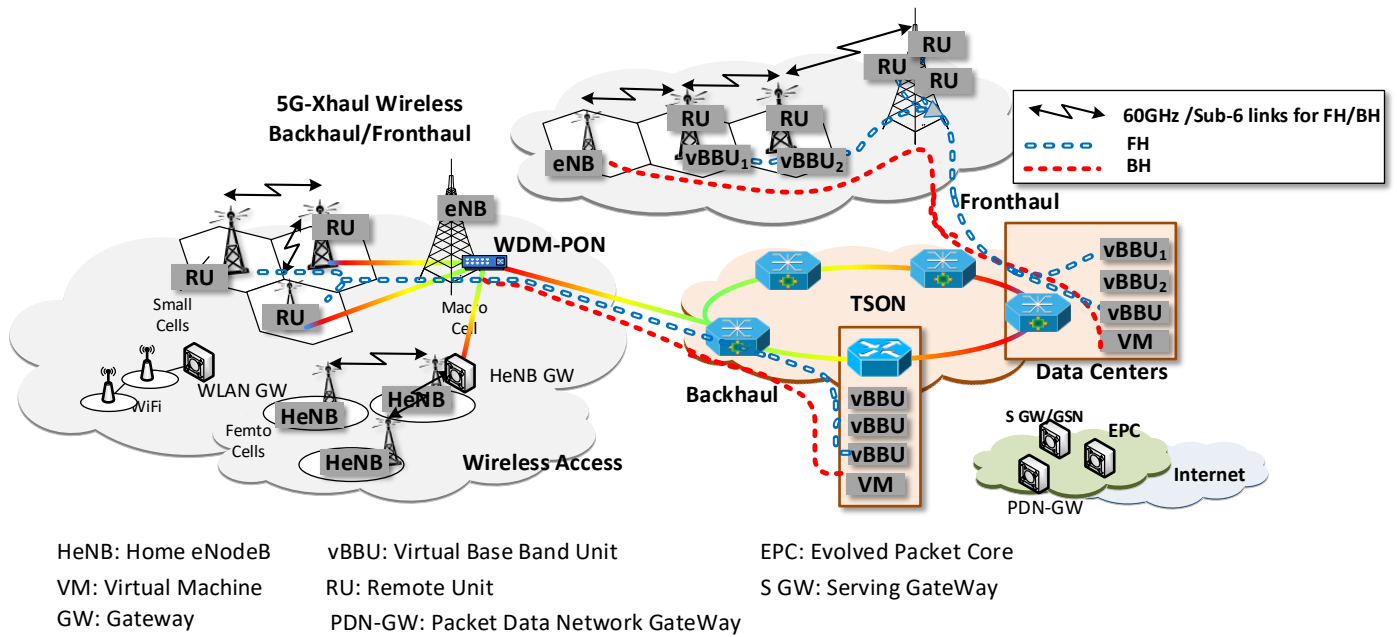
GW: Gateway   PDN-GW: Packet Data Network GateWay

Figure 1: The 5G-XHaul Physical Infrastructure: FH and BH services are provided over a common wired/wireless network infrastructure. In the FH case, parts of the BBU processing can be performed locally and some parts remotely at the DCs enabling the C-RAN flexible split paradigm. BBUs are executed in general purpose servers in the form of virtual entities. BH services interconnect end-users with Virtual Machines hosted in the DCs.

II. OVERVIEW OF THE 5G-XHAUL ARCHITECTURE

A. Data Plane Architecture

The 5G-XHaul data-plane considers an integrated optical and wireless network infrastructure for transport and access. The wireless domain comprises a dense layer of small cells that are located 50-200 m apart. This small cell layer is complemented by a macro cell layer to ensure ubiquitous coverage. Macro-cell sites are about 500m apart. Small cells can be wirelessly backhauled to the macro-cell site using a combination of mm-Wave and sub-6 wireless technologies. Alternatively, the 5G-XHaul architecture allows small cells to be directly connected to a CU using a hybrid optical network platform. This adopts a dynamic and flexible/elastic frame based optical network solution combined with enhanced capacity WDM PONs [7]. This platform can support demanding capacity and flexibility requirements for traffic aggregation and transport. Through this architecture, that has been incorporated in the overall data plane architecture of the 5G PPP architectural vision [ref white paper]. 5G-XHaul aims to efficiently support a large variety of services envisaged for the 5G era.

A key architectural issue associated with this type of infrastructure, is the placement of the Base Band (BB) processing with respect to the RUs. In 5G-XHaul, the concept of C-RAN, where RUs are connected to remote BB processing pools through high bandwidth transport links, is proposed as one solution that can be adopted

to overcome the limitations of the traditional RAN approach (Figure 1). This inclusion of FH requirements in the 5G-XHaul infrastructure introduces new operational network services that need to be supported over the transport network. More specifically, the densely distributed RUs need to be connected to compute resources responsible for BB processing, with very stringent delay and ==synchronization== requirements. 5G-XHaul proposes to support BH and FH jointly in a common infrastructure, ==maximizing== the associated sharing gains, improving efficiency in resource utilization and providing measurable benefits in terms of cost, scalability, sustainability and management simplification. In addition, 5G-XHaul proposes to split processing flexibly, with the aim to relax the stringent requirements transport capacity, delay and ==synchronization==. As illustrated in Figure 2, the range of "optimal split" options, spans between the "traditional distributed RAN" case where "all processing is performed locally at the AP" to the "fully-centralized C-RAN" case where "all processing is allocated to a CU". All other options allow allocating some processing functions at the RU, while the remaining processing functions are performed remotely at the CU. The optimal allocation of processing functions to be executed locally or remotely i.e. the optimal "split", can be decided dynamically based on a number of factors such as transport network characteristics, network topology and scale as well as type and volume of services that need to be supported.

The joint FH and BH requirements described above are supported through the adoption of the 5G-XHaul architecture as well as the advanced wireless and optical network technologies developed by the project. A key enabler of the proposed approach is a high capacity, flexible optical transport comprising both passive and active solutions that plays a central role in the 5G-XHaul infrastructure. The passive solution employs WDM-PONs, while the active solution adopts the Time-Shared Optical Network (TSON) [7] enhanced with novel features for improved granularity and elasticity. These can provide the required connectivity, capacity and flexibility to offer jointly FH and BH functions and support a large variety of end-user and operational services. A high level view of the 5G- XHaul infrastructure is provided in Figure 1.

Given the technology heterogeneity supported by the 5G-XHaul data plane, a critical function of the converged infrastructure is interfacing between technology domains. The required interfaces are responsible for handling protocol adaptation as well as mapping and aggregation/de-aggregation of the traffic across different domains.

It should be noted that different domains (wireless/optical) may adopt different protocol implementations and provide very diverse levels of overall capacity (varying between Mbps for the wireless domain up to tens of Gbps for TSON), granularity (varying between Kbps for the wireless domain and 100 Mbps for TSON) etc. A key challenge also addressed by these interfaces, ==is the mapping of different Quality of Service (QoS) classes==

across different domains as well as the development of flexible scheduling schemes supporting QoS differentiation mechanisms. More specifically, at the optical network ingress point (e.g. TSON edge node) the interfaces receive traffic frames generated by fixed and mobile users and arrange them to different buffers (that the TSON edge node comprises). The incoming traffic is aggregated into optical frames, which are then assigned to suitable time-slots and wavelengths for further transmission according to the adopted queuing policy. For FH traffic a modified version of the CPRI protocol supporting the concept of functional split (eCPRI) has been adopted. It should be noted that due to the large variety of technologies involved in 5G, these interfaces need to support a wide range of protocols and technology solutions and execute traffic forwarding decisions at wire-speed. This requires the development of programmable network interfaces combining hardware level performance with software flexibility. At the egress point, the reverse function takes place. More details on the interfaces integrating wireless and optical domains can be found in [7], [13].

B. Overarching Layered Architecture

As shown in Figure 1 the 5G-XHaul infrastructure exhibits a great degree of heterogeneity in terms of technologies. To address the challenge of managing and operating this type of complex heterogeneous infrastructure efficiently, 5G-XHaul proposes the integration of the SDN and NFV approaches. In SDN, the control plane is decoupled from the data plane and is managed by a logically centralized controller that has a holistic view of the network [6]. At the same time, NFV enables the execution of network functions on commodity hardware (general-purpose servers) by leveraging software virtualization techniques [8]. Through joint SDN and NFV consideration, significant benefits can be achieved, associated with flexible, dynamic and efficient use of the infrastructure resources, simplification of the infrastructure and its management, increased scalability and sustainability as well as provisioning of orchestrated end-to-end services.

Examples of features that enable these benefits include the option to virtualize the separate control plane, using NFV and deploy Virtual Network Functions (VNFs) [11]. VNFs are controlled by the SDN controller, to allow on-demand resource allocation and dynamically changing workloads [6]. SDN network elements can be either Physical Network Functions (PNFs) or VNFs, since they can be implemented as software running on general-purpose platforms in virtualized environments [6]. The virtualization of network elements enables flexible allocation of data plane resources according to network applications requirements. On the other hand, network applications can include SDN controller functions, or interact with SDN controllers and can themselves provide VNFs. Service Chaining (SC), supporting orchestrated service provisioning over heterogeneous environments, is considered to be one possible network application.
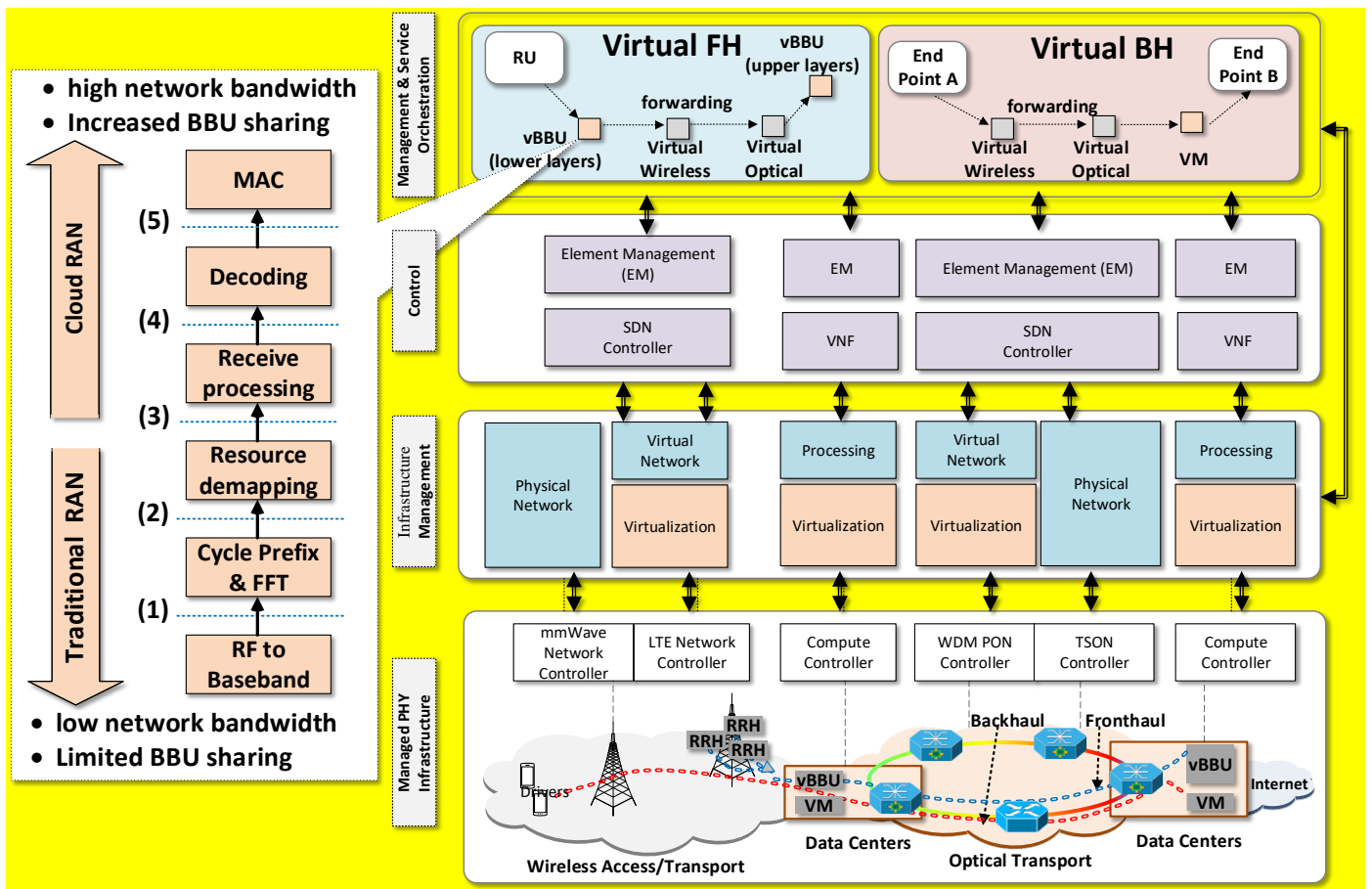
Figure 2: The overall overarching architecture supporting functional split processing [3]-[4].

Taking advantage of the SDN concept and the benefits of cross-technology virtualization, 5G-XHaul proposes an overarching layered architecture able to efficiently and effectively support 5G services presented in Figure 2. The managed Physical Infrastructure (PI) layer is described in section II.A.

The Infrastructure Management Layer (IML) is responsible for the management of the different technology domains and the creation of virtual infrastructure (VI) slices, comprising heterogeneous resources. These VI slices enable multi-tenancy operation models providing both FH and BH services. This layer communicates with the various network and computes controllers that are responsible for retrieving information and communicating with the individual domains. Once the information has been collected, the resources are abstracted and virtualized. From the architectural and functional perspective, IML addresses all virtualization and virtual resource management functions. Management of traditional non-virtualized physical infrastructures is also supported.

Cross-domain orchestration of the virtual and physical infrastructures, created and exposed by the IML to the higher layers, is carried out by the Control Layer. This layer has a holistic view of all network segments and technology domains and implements converged control and management procedures for dynamic and

automated provisioning of end-to-end connectivity services (i.e. service chaining), according to specific QoS considerations. Configuration of virtualized (or non-virtualized) wireless and optical network resources, as well as legacy devices, is carried out by a set of distributed SDN controllers. Apart from network configuration capabilities, offered by the SDN controllers, further enhanced VNFs, running on top of the VIs, can be developed to operate the entire heterogeneous infrastructure in a seamless manner.

Finally, the Management and Service Orchestration Layer is responsible for the converged orchestration of the computation and network services. It is also used for the composition and delivery of multi-tenant chains of virtualized network functions and should support interoperability with legacy software and hardware technologies and architectures.

III. USE CASE: JOINT OPTIMIZATION OF FH/BH

As already discussed, one of the main innovations of the proposed solution (Figure 2) is that of the data plane architecture, offering the ability to jointly support FH and BH services . This has been achieved through a hybrid optical transport, deploying passive and active optical network solutions, as described in the "Data Plane Architecture" subsection above. In addition to this, the higher layers of the architecture that facilitate access and management of both network and compute resources also play a key role. More specifically, the ability of the IML to create VI slices across heterogeneous domains and to expose these to the upper layers (as described in section II.B) is an instrumental architectural tool, facilitating the delivery of FH and BH services.

The identification of optimal VIs that can support the required FH or BH services in terms of both topology and resources includes:

- Ordering, referred to as SC, of the relevant functions (VNF or PNF) that need to be applied to the traffic flows traversing the VIs.

- Estimating the virtual resources required to support SC and executing the corresponding applications over the PI.

- Mapping of the virtual resources to the physical resources.

A graphical illustration of this process is shown in the upper part of Figure 2, where two different VIs, supporting independently FH and BH functions, are deployed over a common PI. These VIs can be e.g. assigned to different virtual operators.

As discussed in Section II.A 5G-XHaul proposes the interconnection of RUs and end-users with a set of geographically distributed compute resources. These are general-purpose servers hosted by DCs (as showcased

e.g. by Alcatel-Lucent, Intel, China Mobile and Telefónica, Mobile World Congress 2015), through a multi-technology transport network. To support virtual FH (VFH) services, over the 5G-XHaul PI, RUs generate demands that need to be supported by compute resources. Compute resources are responsible for executing the various BB functions in a predefined order (left part of Figure 2). As already discussed, based on the split option adopted, these functions can be partly executed locally at the RU or centrally through general-purpose servers at a DC. The split choice will dictate the processing allocation, to local and central compute resources, and enforces the corresponding SC graph. A graphical representation of a typical virtual BH (VBH) service that supports content delivery (CDN) to end users is also illustrated in Figure 2. The SC graph indicates that the VBH service allows mobile traffic, generated at the wireless access domain, to traverse a hybrid multi-hop wireless/optical transport network, before it reaches the compute resources.

To evaluate the performance of this type of infrastructures and the proposed architecture, we have developed a mathematical framework, based on MOP, for the integrated wireless and optical network domains, considering the data plane described in Section II.A and the details of the compute resources required. Our study focuses on optimal planning of VFH and VBH infrastructures, in terms of both topology and resources considering overall power consumption and end-to-end delay associated performance metrics. The proposed model takes a holistic view, considering jointly mobile FH and BH functions, to ensure appropriate allocation of the required resources across all domains. To identify the best performing FH and BH VIs, detailed power consumption and end-to-end delay models, describing the optical and wireless network as well as the compute domains and the associated interfaces, are considered [9]. The results obtained using this modeling framework, focus on the specific use case of joint FH and BH optimisation, assuming delivery of CDN services to the end-users.

The joint VFH/VBH planning problem is solved taking into account a set of constraints that guarantee efficient and stable operation of the resulting VIs. The relevant constraints included in our analysis are summarized below:

- VFH and VBH infrastructures have a specific set of service requirements. In response to this, different types of VNFs are deployed and chained together, each having specific processing and bandwidth requirements. To realize SC, sufficient network bandwidth and processing capacity must be allocated to the planned VIs, corresponding to specific physical resources, for the interconnection and deployment of VNFs. VNFs are processed in the order defined by the corresponding SC.

- The physical resources that should be reserved to ensure SC depend on the users' mobility model assumed, the size of the wireless cells and the traffic model adopted. In the ideal scenario, seamless

handovers for a mobile user can be 100% guaranteed only if the required amount of resources is reserved for all its neighboring cells. However, to limit overprovisioning of resources, a more practical approach is to relate the reserved resources in the neighboring cells with the handoff probability. In this paper, the amount of resources that are leased in the wireless domain is assumed to be an increasing function of the handoff probability [12]. Given that both the RUs and the end-users need to be supported by remotely located compute resources, the additional resource requirements also propagate in the transport network and the compute domain.

- During the VI planning process, flow conservation as well as mapping and aggregation/de-aggregation of traffic between different domains is taken into consideration.

- Given that both FH and BH services, assumed in this use case, require compute processing there is a clear need to include the impact of the compute domain in the overall infrastructure evaluation. Therefore, the traffic associated with these services has to be mapped, not only to network resource requirements, but also to compute resource requirements. This introduces an additional constraint, associated with the conversion of network to compute resource requests. To achieve this, a mapping parameter, defined as network-to-compute, is introduced that provides the ratio of network requirements (measured in Mbps) and computational requirements (measured in Operations Per Second (OPS)), of a specific service demand. This parameter takes low values, for cloud services requiring high network bandwidth, but low processing (e.g. video streaming) and high values for tasks requiring intensive processing but are not bandwidth intensive (e.g. data mining, Internet of Things). Regarding BH services, the Standard Performance Evaluation Corporation recently established the Cloud subcommittee to develop benchmarks that are able to measure these parameters. Taking a similar approach the authors in [14] measured the average ratio between computational and network bandwidth requirements for various "Big Data" analytics workloads. Similarly, FH services require specific computing resources to support BB processing. The processing power depends on the sub-components of the BBU [3]-[4] including processing tasks related to Fast Fourier Transforms (FFT), error correction, processing-resource mapping/de-mapping etc. calculated in Giga OPS (GOPS). The resulting processing power depends on the configuration of the Long Term Evolution (LTE) system (i.e. number of antennas, bandwidth, modulation, coding, number of resource blocks), more details can be found in [4].

- To address the importance of end-to-end delay for specific services e.g. FH or real time BH services, end-to-end delay is considered in our analysis. In highly loaded heterogeneous networks, such as

the 5G-XHaul solution, end-to-end delay can be greatly influenced by queuing delays associated with the interfaces across the infrastructure domains. Therefore, applying specific queuing policies and scheduling strategies at the interfaces, across technology domains, can offer significant delay benefits. Traditionally, the adoption of queuing models and scheduling strategies can be mathematically modelled applying queuing theory and modeling the different network domains as an open/closed mixed queuing network. However, such a model is not able to abide to the strict FH latency constraints. To address this issue, the queuing delays for the VFH are modelled under worst case operational conditions, using network calculus theory [13].

- As described in detail in Section II.A, 5G-XHaul proposes to exploit the benefits of flexible FH processing splits. In practice, flexible processing splits refer to the choice of a single split option for every time instance. Once the split option has been selected, the corresponding service chaining is applied across the network, deploying specific network and compute resources as dictated by the relevant split. For example, assuming an LTE system with transmission bandwidth B_w=20 MHz, sampling frequency of 30.72 MHz, bit resolution per I/Q 15, oversampling factor 2 and 2 antennas under split option (1) (left part of Figure 2) the transport network bandwidth for the RU-CU interconnection has to be at least 2.46 Gbps. However, when employing split option (2), this is reduced to 720 Mbps, assuming 1200 subcarriers and FFT period of 66.67 μsec [3], [4].

Based on the functional split adopted, part of the processing can be performed by compute resources either at a local cloudlet [12] c, $c \epsilon \mathcal{C}$ ($\mathcal{C}$ denotes the set of cloudlets) with cost w_c per GOPS or at a remote regional DC $s \epsilon \mathcal{S}$ ($\mathcal{S}$ denotes the set of DCs) with cost w_s per GOPS. Assuming that the cost for operating FH capacity $u_{FH,e}$ of physical link $e \epsilon \mathcal{E}$ ($\mathcal{E}$: set of physical links) is w_e and $\pi_{FH,s}$ $\pi_{FH,c}$ are the processing capacities used for BBU processing at the remote server and the cloudlet, respectively, the optimal VFH infrastructure is identified through the minimization of the cost:

$$min \, \mathcal{VFH}(\boldsymbol{u}, \boldsymbol{\pi}) = \sum_{e \epsilon \mathcal{E}} w_e \, u_{FH,e} + \sum_{s \epsilon \mathcal{S}} w_s \pi_{FH,s} + \sum_{c \epsilon \mathcal{C}} w_c \pi_{FH,c} \quad (1)$$

subject to virtual and physical capacity constraints, functional split and demand constraints described above.

Due to the inherent energy efficiency of the optical transport, improved performance is achieved for a higher degree of centralization (C-RAN) compared to the traditional RAN approach. However, this comes at the expense of overloading the transport network to support FH requirements, leaving to reduced available resources for the BH functions. To address this issue, the secondary optimization objective is to minimize the end-to-end delay in the BH, subject to demand processing and capacity constraints:

$$\min \mathcal{VBH}(u, \pi) = \sum_{e \epsilon \mathcal{E}} \frac{1}{\mathcal{U}_e - u_{FH,e} - u_{BH,e}} + \sum_{s \epsilon \mathcal{S}} \frac{1}{\Pi_s - \pi_{FH,s} - \pi_{BH,s}} + \sum_{c \epsilon \mathcal{C}} \frac{1}{\Pi_c - \pi_{FH,c} - \pi_{BH,c}} \quad (2)$$

where $u_{BH,e}$, $\pi_{BH,s}$ represent the network and server capacity allocated to the BH, respectively, $\mathcal{U}_e$, is the total capacity of e and Π_s, Π_c are the total processing capacity of the DC s and the cloudlet c, respectively.

The MOP problem described through equations (1)-(2) can be written as $\min \mathcal{F}(\boldsymbol{u}, \boldsymbol{\pi}) = [\mathcal{VFH}(\boldsymbol{u}, \boldsymbol{\pi}), \quad \mathcal{VBH}(\boldsymbol{u}, \boldsymbol{\pi})]$ subject to the constraints previously discussed. This problem, can be transformed from an MOP problem into a single objective optimization, using traditional scalarization techniques (e.g. the Pascoletti-Serafini scalarization method [10]), and be solved using relaxation schemes (e.g. Lagrangian Relaxation). In the following section, the performance of the overall architecture is evaluated through the modelling framework described above. The joint FH and BH service provisioning problem is modeled through the MOP (1)-(2) and is evaluated, in terms of power consumption and service delay, adopting a realistic network topology using a measured traffic dataset [15].

IV. PERFORMANCE EVALUATION

The network topology assumed is the Bristol 5G city infrastructure, illustrated in Figure 3 (a) that will host the final proof-of-concept 5G-XHaul demonstrator. This infrastructure covers a 10x10 km^2 area over which 50 APs are uniformly distributed and comprises TSON nodes in the optical segment, and microwave point-to-point links for backhauling the APs. TSON deploys a single fiber per link, 4 wavelengths per fibre, wavelength channels of 10Gbps each, minimum bandwidth granularity of 100Mbps and maximum link capacity of 40Gbps. In the present study, w_e is associated with the power consumption of link $e \epsilon \mathcal{E}$. The TSON power consumption model is provided in [7], [13]. The microwave transceivers considered, support maximum bandwidth of 2Gbps and their typical power consumption is 45W (Huawei OptiX RTN310). Furthermore, a 2×2 MIMO scheme with adaptive number of transmission elements, carrier frequency 2.6GHz, 20 MHz bandwidth adjustment and capacity per cell up to 201.6Mbps has been considered, while users are distributed and generate traffic over the serviced area according to real datasets reported in [15] (Figure 3 (b)). This traffic, corresponding to wireless access traffic, needs to be processed by specific compute resources. The proposed optimization scheme is focusing on the following two scenarios:

a) *"Traditional RAN,"* where power consumption per AP ranges between 600 and 1200 Watts under idle and full load conditions, respectively, and commodity servers are used to support CDN services.

b) *"C-RAN with virtual BBUs (vBBUs)"* where commodity servers are used to support both FH (through the creation of vBBUs [1]) and BH CDN services.
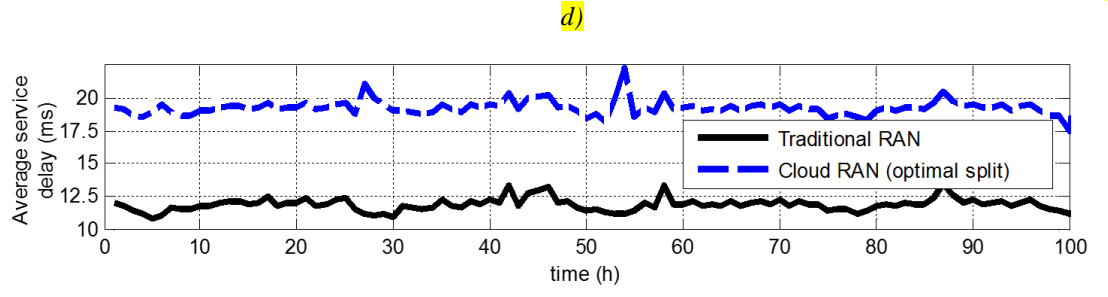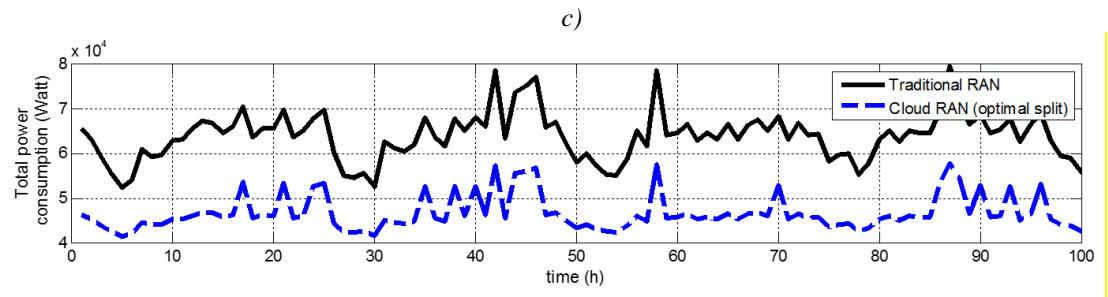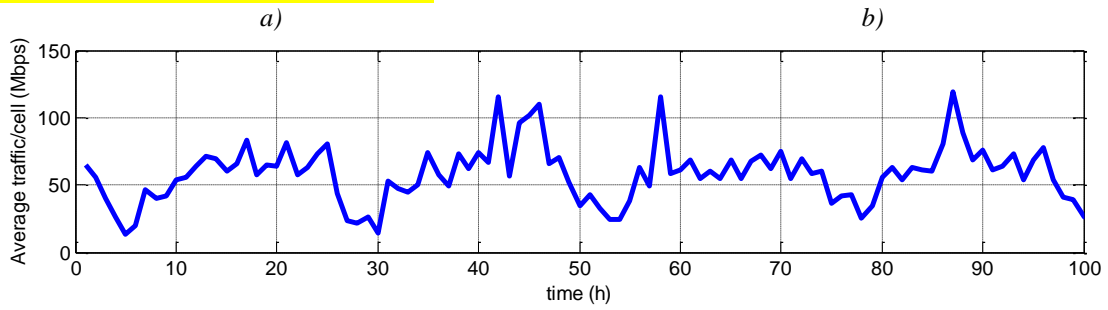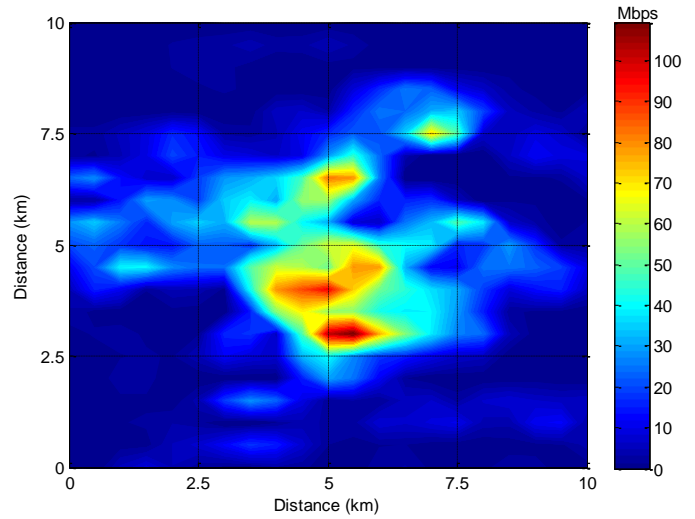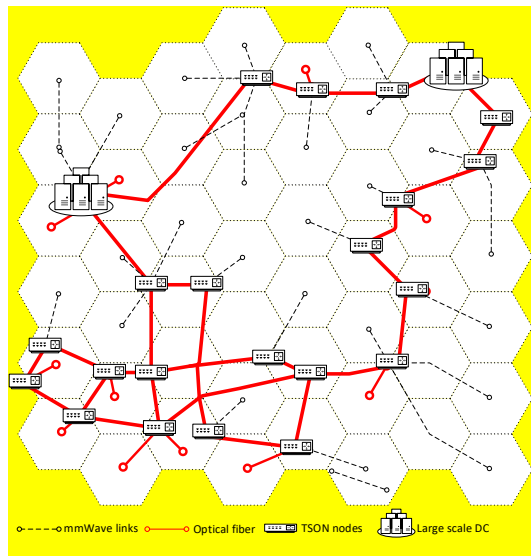
Figure 3 a) Bristol 5G city network topology with mmWave backhauling, b) Snapshot of spatial traffic load and c) average traffic/BS based on the dataset [10] during 8/2012, d)-e) Total power consumption and total service delay over time for the traditional RAN.

In the case of *C-RAN with vBBUs*, where optimal split options are deployed, two types of servers have been considered: a) small scale commodity servers (cloudlets) close to the APs and, b) commodity servers hosted by large scale DCs (Figure 3 (a)) with an average cost equal to 2Watts/GOPS and 1.6Watts/GOPS, respectively. Details regarding the numerical values used in the simulations are provided in [7]. Although both types of servers can provide the necessary processing power for C-RAN and CDN services, large scale DCs provide superior performance per Watt, compared to cloudlets. Figure 3 (d) shows that significant energy savings (ranging between 60-75%) can be achieved when adopting the C-RAN approach using the integrated wireless-optical infrastructure, compared to traditional RAN. However, due to sharing of network resources between BH services and high priority FH services, C-RAN leads in increased BH service delays that remain below 25 ms (Figure 3 (e)). On the other hand, traditional RAN provides minimum end-to-end BH service delays, as no sharing with FH services is required, but at the expense of increased energy consumption due to the lack of BBU sharing.

The impact of mobility on the total power consumption and the optimal split option adopted is shown in Figure 4 (a) and (b), respectively. The call-to-mobility factor is defined as the fraction of the service holding time over the cell residence time [13], with low call-to-mobility factor values indicating high degree of mobility. It is known that high degree of mobility introduces additional resource requirements in the wireless domain. To ensure seamless end-to-end connectivity between end-users, RUs and compute resources, these additional resource requirements also propagate across the transport network and the compute domains. In Figure 4 (b) it is observed that lower split options are beneficial for higher mobility, enabling a larger number of BB processing tasks to be offloaded to remote DCs. Given that BB processing requirements increase with mobility, a higher degree of centralization benefits the system, due to the increased consolidation and the improved performance per watt that large scale remote DCs offer, compared to local cloudlets.
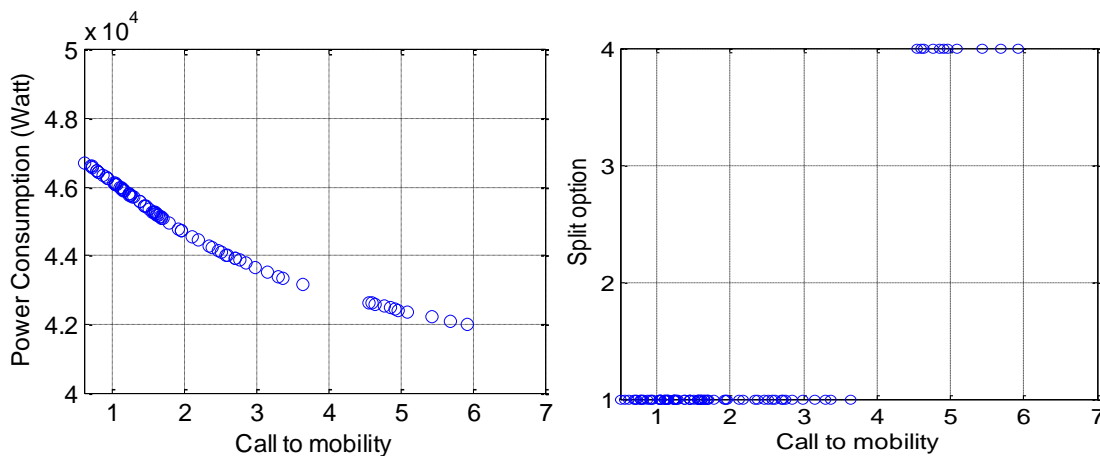

Figure 4: Impact of mobility power consumption and optimal split option (load 18Mbps/cell).

Figure 5 (a) illustrates the impact of service requirements in terms of network and compute resources on the optimal split option adopted. CDN services with high network-to-compute ratios (e.g. video analytics) require significant network resources to operate, leading to overutilization of transport capacity. This effect is counter-balanced by the selection of higher split options (i.e. options 3, 4) that require lower bandwidth for the interconnection of RUs with CUs compared to the bandwidth requirements of lower split options (i.e. options 1, 2). The impact of the traffic load on the total power consumption is illustrated in Figure 5 (b). As expected, for higher traffic load, the total power consumption increases, and a step-like increase is observed, above 45 Mbps per cell traffic load. Beyond this threshold, the preferable system split option becomes split 4 (rather than split 1) and a large number of cloudlets per geographic region are activated to support BB processing requirements.
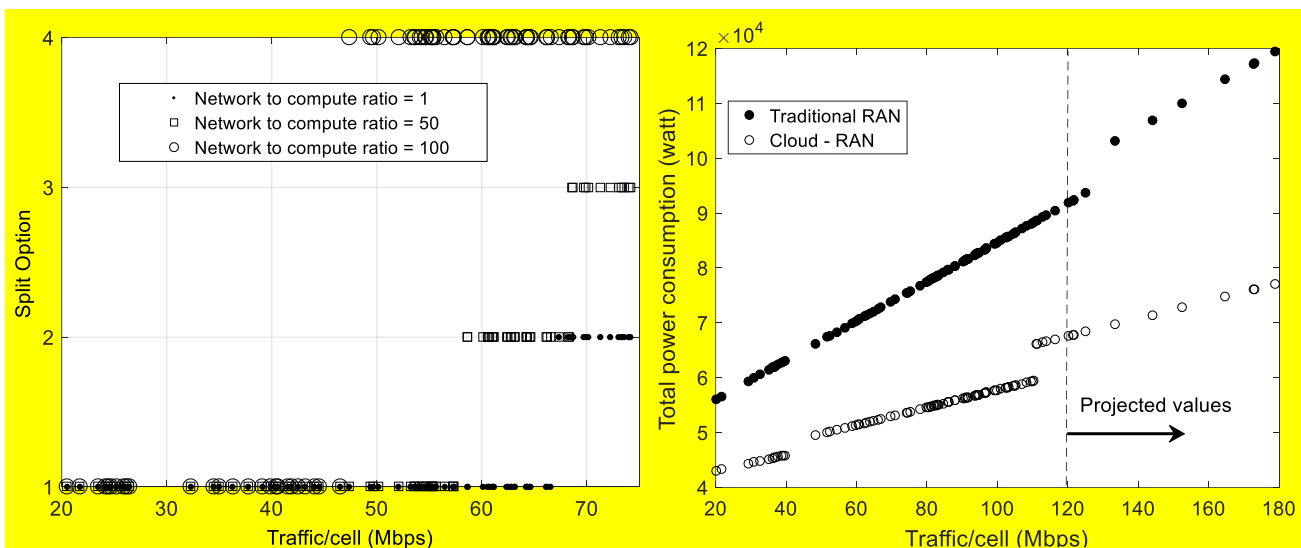

Figure 5 Split option as a function of load for different Compute to network ratios.

Finally, the impact of the relative processing and transport network cost, on the optimal split option, is illustrated in Figure 6. The relative local to remote processing cost, is defined as the ratio of the power consumed for processing of data (associated with a specific service) at the local cloudlet over the power consumed for processing of the same data remotely at large scale DCs. It is seen that increasing this ratio makes it beneficial to perform more processing functions at large scale remote DCs. Thus a lower split option is preferable. To include the impact of the network cost in this analysis, the end-to-end transmission cost is also plotted in Figure 6. As the transmission cost increases (higher number of wireless hops), it is beneficial to perform more processing functions at the local cloudlets and hence adopt a higher split option.
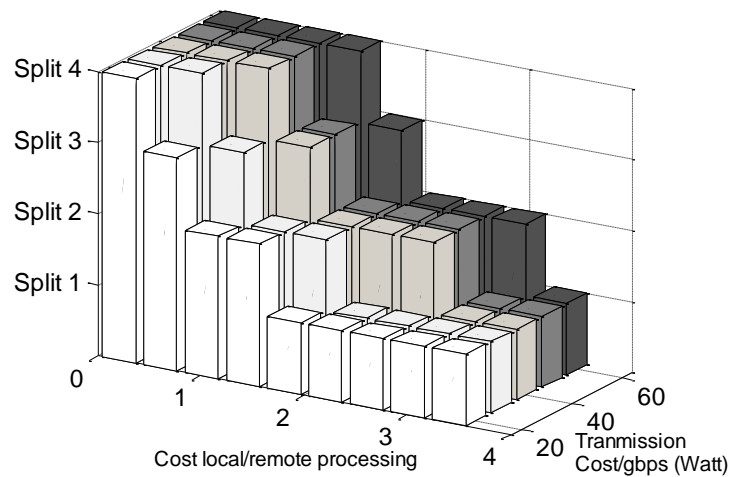
Figure 6 Split options for various processing and transmission costs.

V. CONCLUSIONS

This paper presents a converged optical-wireless 5G network infrastructure to support jointly operational network (C-RAN) and end-user services proposed by the EU 5GPPP project 5G-XHaul. A layered architecture, inspired by the ETSI NFV standard and the SDN reference architecture is also presented. A novel MOP modeling framework has been developed to evaluate the performance of the 5G-XHaul architecture. Our study has considered a variety of FH and BH options, spanning from the distributed RAN to solutions adopting fully or partially the C-RAN approach. Modeling results show that the proposed architecture can offer significant energy savings, but there is a clear trade-off between overall energy consumption and end-user service delays.

ACKNOWLEDGMENT

REFERENCES

[1]   C-RAN. The Road Towards Green RAN. White Paper 3.0 (Dec. 2013), http://goo.gl/Sw6bfw

[2]   Nokia Press Release, Nokia accelerates centralized RAN deployment with expanded mobile fronthaul solution #MWC16, Feb. 2016, [Online]: http://tinyurl.com/h5kmbvb

[3]   U. Dötsch et al., Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE, *Bell Labs Tech. J,* vol. 18, no.1, pp. 105-128, May 2013

[4]   D. Wubben et al., "Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible centralization through cloud-RAN," *IEEE Signal Process. Mag*., vol.31, no.6, pp.35-44, Nov. 2014

[5]   N. Nikaein, M. K. Marina, S. Manickam, A. Dawson, R. Knopp, and C. Bonnet, "OpenAirInterface: A Flexible Platform for 5G Research", *SIGCOMM Comput. Commun. Rev.* Vol. 44, No. 5, Oct. 2014.

[6]   ETSI GS NFV-SWA 001 V1.1.1 (2014-12), "Network Functions Virtualisation (NFV); Virtual Network Functions Architecture", 2014.

[7]   H2020 Project 5G-Xhaul, Deliverable 2.2 "System Architecture Definition", July 2016 [Online]: http://www.5g-xhaul-project.eu/download/5G-XHaul_D_22.pdf

[8]   Bo Han, V. Gopalakrishnan, J. Lusheng, L. Seungjoon, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Commun. Mag*., vol.53, no.2, pp.90-97, Feb. 2015

[9]   B. R. Rofoee, G. Zervas, Y. Yan, M. Anastasopoulos, A. Tzanakaki, S. Peng, R. Nejabati, and D. Simeonidou, "Hardware Virtualized Flexible Network for Wireless-DataCenter (invited)" , *IEEE/OSA J. Opt. Commun. Netw.*, March 2015, vol. 3, pp A526 - A536

[10]   A. Pascoletti, P. Serafini, "Scalarizing vector optimization problems", J Optimization Theory Appl, vol. 42, pp. 499-524, 1984.

[11]   M. Savi, M. Tornatore, G. Verticale, "Impact of processing costs on service chain placement in network functions virtualization," *in proc. of IEEE NFV-SDN*, 2015 pp.191-197, 18-21 Nov. 2015

[12]   J.-Y. Le Boudec, P.Thiran, "Network calculus: a theory of deterministic queuing systems for the internet", Springer, 2001

[13]   A. Tzanakaki et al., "Virtualization of heterogeneous wireless-optical network and IT infrastructures in support of cloud and mobile cloud services," *IEEE Commun. Mag*. , vol.51, no.8, pp.155-161, Aug. 2013.

[14]   J. Chang, K. T. Lim, J. Byrne, L. Ramirez, and P. Ranganathan. "Workload diversity and dynamics in big data analytics: implications to system designers," *in Proc. of ASBD* 2012.

[15]   X. Chen et al., "Analyzing and modeling spatio-temporal dependence of cellular traffic at city scale," *in proc. of IEEE ICC*, pp.3585-3591, 2015.