5G PPP

European Commission | Horizon 2020 European Union funding for Research & Innovation

5G-XHaul

Dynamically Reconfigurable Optical-Wireless Back-haul/Fronthaul with Cognitive Control Plane for Small Cells and Cloud-RANs

D2.3 Architecture of Optical/Wireless Backhaul and Fronthaul and Evaluation

This project has received funding from the European Union's Framework Programme Horizon 2020 for research, technological development and demonstration

Advanced 5G Network Infrastructure for the Future Internet

Project Start Date: July 1st, 2015 | Duration: 36 months

H2020-ICT-2014-2 671551 | December 31st, 2016 – Version 1.0

Project co-funded by the European Commission

Under the H2020 programme

Dissemination Level: Public

Grant Agreement Number:	671551
Project Name:	Dynamically Reconfigurable Optical-Wireless Backhaul/Fronthaul with Cognitive Control Plane for Small Cells and Cloud-RANs
Project Acronym:	5G-XHaul
Document Number:	D2.3
Document Title:	Architecture of Optical/Wireless Backhaul and Fronthaul and Evaluation
Version:	Final
Delivery Date:	December 31st, 2016 (February 1st, 2017)
Responsible:	University of Bristol, High Performance Networks Group (**UNIVBRIS-HPN**)
Editor(s):	Anna Tzanakaki (**UNIVBRIS-HPN**)
Authors:	Markos Anastasopoulos (**UNIVBRIS-HPN**), Anna Tzanakaki (**UNIVBRIS-HPN**), Dimitra Simeonidou (**UNIVBRIS-HPN**), Jens Bartelt (**TUD**), Jim Zou (**ADVA**), Michael Eiselt (**ADVA**).
Keywords:	converged optical and wireless network, transport infrastructure, frontal, backhaul, flexible functional splits, multi-objective optimisation, performance evaluation.
Status:	Final
Dissemination Level	Public
Project URL:	http://www.5g-xhaul-project.eu/

Table of Contents

List of Figures

List of Tables

Executive Summary

To address the predicted global mobile data traffic increase by a factor of eight between 2015 and 2020 5G-XHaul focuses on developing a converged optical and wireless network solution supported by a flexible and scalable control plane with the aim to form a flexible transport infrastructure. This infrastructure will be able to jointly support the backhaul (BH) and fronthaul (FH) functionalities required to cope with the future challenges imposed by fifth generation (5G) Radio Access Networks (RANs).

Recognizing the benefits of the C-RAN architecture and the associated challenges, currently mobile FH solutions are expanded to adopt more effective wireless technologies operating in the Sub-6 GHz and 60 GHz frequency bands enhanced with advanced beam tracking and Multiple-Input Multiple-Output (MIMO) techniques, new versatile Wavelength Division Multiplexing (WDM) optical network platforms as well as novel control and management frameworks that allow service driven customization offering increased granularity, end-to-end optimization and guaranteed Quality of Service (QoS). However, to relax the stringent FH requirements of C-RAN architectures, while taking advantage of its pooling and coordination gains, solutions relying on architectures adopting flexible functional splits have been proposed. 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.

In this deliverable, the concept of supporting flexible functional splits is addressed through a combination of small scale servers (cloudlets) embedded in the wireless access and relatively large-scale Data Centres (DCs) placed in the metro network domain. As shown through relevant studies flexible functional splits impose the requirement of fine bandwidth granularity and elastic resource allocation at both the wireless and the optical transport network domains. On the other hand 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, 5G-XHaul proposes a converged optical-wireless 5G network infrastructure interconnecting compute resources with fixed and mobile users, to support both operational network (C-RAN and flexible functional split options) and end-user services and defines an innovative heterogeneous network architecture adopting a variety of wireless and optical technologies able to support a wide range of 5G services. In view of this, this deliverable reports on the optical/wireless backhaul and fronthaul 5G-XHaul architecture. In addition, a detailed evaluation of the performance of this architecture is provided through modelling and simulations studies.

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 that supports jointly BH and FH services and the concept of flexible functional splits, 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) modelling 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 architecture is evaluated in terms of operational expenditure related with FH services associated with power consumption under strict delay constraints, end-to-end service delay of BH services and monetary capital and operational costs, while relevant trade-offs are identified and discussed.

1. 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 CApital EXpenditures (CAPEX) and Operational EXpenditures (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 standardised through the Common Public Radio Interface (CPRI), the Open Base Architecture Initiative (OBSAI) and the Open Radio Interface (ORI), being currently CPRI the most widely used standard. More details can be found in deliverable D2.2 [D2.2]. 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 60 GHz frequency bands enhanced with advanced beam tracking and MIMO techniques, new versatile Wavelength Division Multiplexing (WDM) optical network platforms [3] 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 have been proposed [4], [5]. 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, it has been recently proposed,. A typical example includes the OpenAirInterface (OAI), an open source 4G/5G radio stack able to be executed on general purpose servers hosted in data centres (DCs) [6]. 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) embedded in the wireless access and relatively large-scale DCs placed in the metro network domain. It should be noted that flexible splits impose the requirement of fine bandwidth granularity and elastic resource allocation at both the wireless and the optical transport network domains. On the other hand 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, 5G-XHaul proposes a converged optical-wireless 5G network infrastructure interconnecting compute resources with fixed and mobile users, 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, is 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. 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 [8] that supports jointly BH and FH services and the concept of flexible functional splits, 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) modelling 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 minimise end-to-end service delay of BH services.

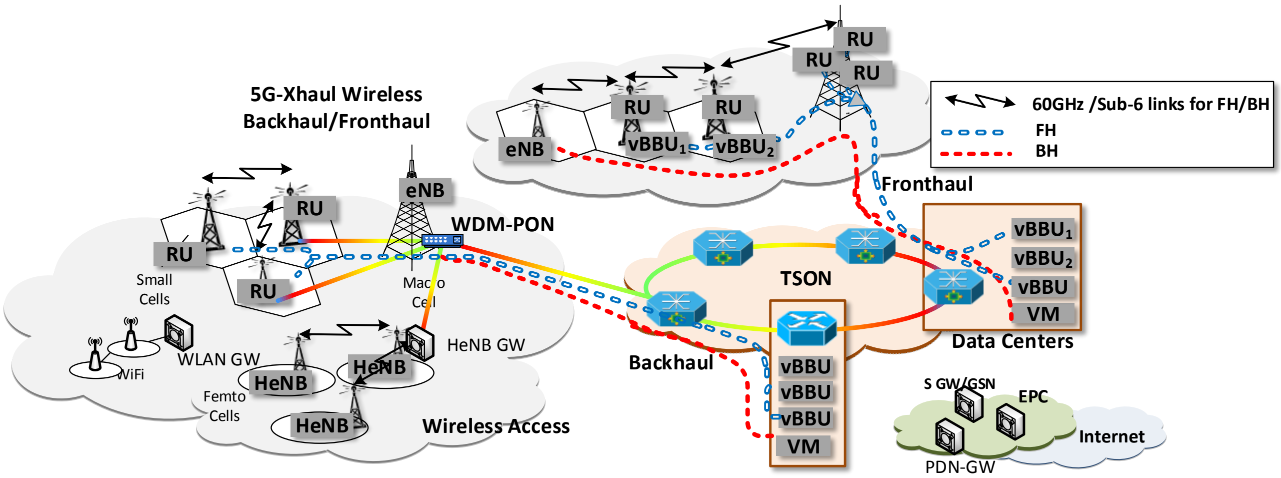
The rest of the document is structured as follows. Section 2 describes in detail the 5G-XHaul data plane architecture including heterogeneous optical and wireless network domains and technologies. Section 3 provides a description of the control and management plane. Section 4 describes in detail the modelling framework developed for the evaluation of the 5G-XHaul infrastructure, section 5 focuses on the support of BH services, section 6 focuses on the analysis of FH services and their evaluation and section 7 focuses on the joint support of fronthaul and backhaul services through the 5G-XHaul infrastructure providing some trade-off discussion and presenting a set of relevant evaluation results. Finally, section 8 summarises the conclusions of the work.

2. 5G-XHaul 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 500 metres apart. Small cells can be wirelessly backhauled to the macro-cell site using a combination of mmWave 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 platform adopts a dynamic and flexible/elastic frame-based optical network solution combined with enhanced capacity WDM PONs [9], and can support demanding capacity and flexibility requirements for traffic aggregation and transport. Through this architecture, already incorporated in the overall data plane architecture of the 5G-PPP architectural vision [7], 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). The 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 synchronisation requirements. 5G-XHaul proposes to support BH and FH jointly in a common infrastructure, maximising 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 in terms of transport capacity, delay and synchronisation. 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-centralised 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. More detail on the optimal split options and FH services can be found in section 6 of this document.

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) [9] enhanced with novel features for improved granularity, elasticity and synchronisation. These can provide the required connectivity, capacity and flexibility to offer jointly FH and BH functions. A high level view of the 5G-XHaul infrastructure is provided in Figure 1, also discussed in detail in deliverable D2.2.



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 (PI): 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.

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 and 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 included in the TSON edge node. 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) is assumed. 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 [9], [15].

3. 5G-XHaul: Control and Management

As stated previously, the proposed infrastructure shown in Figure 1 exhibits a great degree of heterogeneity in terms of technologies. To address the challenge of managing and operating this type of complex heterogeneous infrastructure, we propose 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 centralised controller that has a holistic view of the network [20]. In early SDN deployments the data plane implementations only supported packet forwarding related functionalities. However, the advent of new high performing technologies such as WiGig [21] and dynamic optical metro solutions necessitate the execution of much more complex networking functions such as scheduling, network monitoring and management, resource virtualisation, isolation, etc. In response to this, SDN controlled programmable hardware infrastructures can now effectively support implementation of these functionalities using high level programming languages.

At the same time, NFV enables the execution of network functions on compute resources by leveraging software virtualization techniques [28]. 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). These are controlled by the SDN controller, to allow on demand resource allocation, able to support dynamically changing workloads [28]. SDN network elements can be treated as VNFs, since they can be implemented as software running on general-purpose platforms in virtualized environments. Both SDN and non-SDN models can be supported by SDN network elements. On the other hand, network applications can include SDN controller functions, or interact with SDN controllers and can themselves provide VNFs. Network elements controlled by SDN controllers can also provide Physical Network Functions (PNFs). Service Chaining (SC), combining and orchestrating physical and virtual network functions to support end-to-end service provisioning over heterogeneous environments, is considered to be one possible network application.

Taking advantage of the SDN concept and the benefits of cross-technology virtualisation, 5G-XHaul proposes an overarching layered architecture able to efficiently and effectively support 5G services presented in Figure 2.

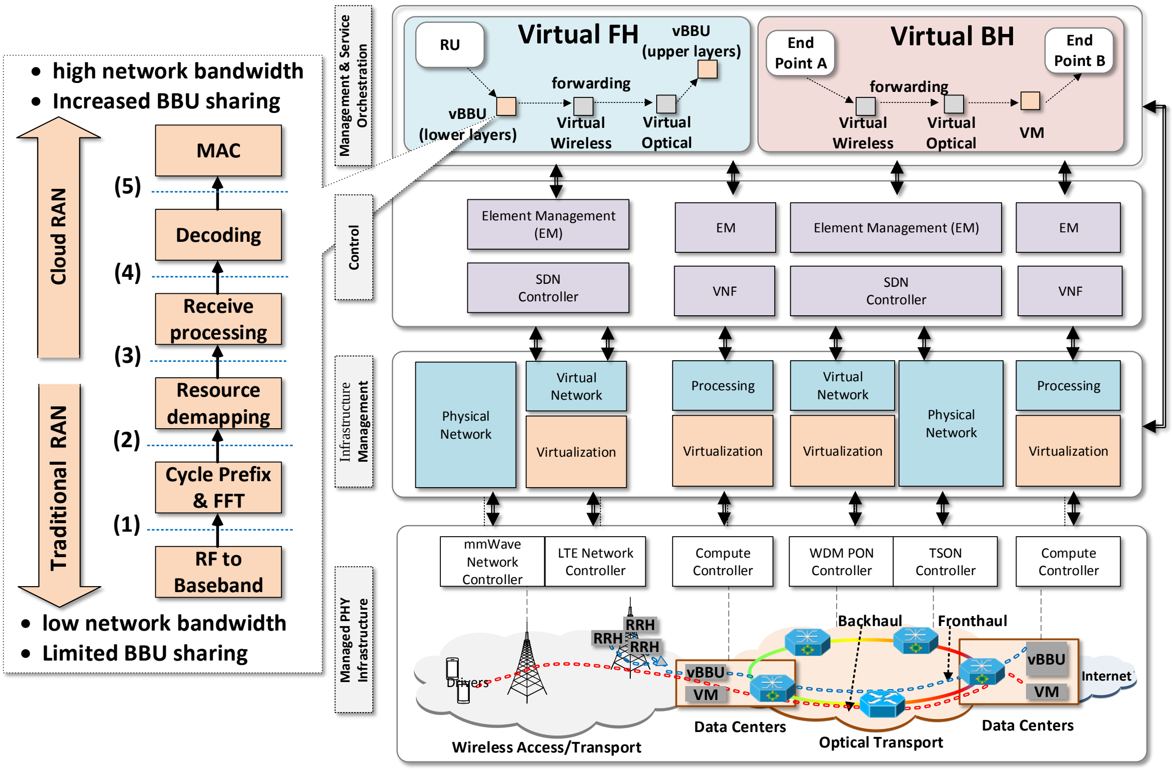


Figure 2: The 5G-XHaul overarching architecture supporting functional split processing [4]-[5].

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 compute 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 virtualised (or non-virtualised) 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.

4. 5G-XHaul Infrastructure modelling

The proposed network infrastructure deploys of a set of optical and wireless network technologies to interconnect a variety of end-devices and IT resources. The wireless access domain comprises cellular LTE operating in the 2 GHz frequency band, a Sub-6 GHz access technology (e.g. WiFi) and a high-frequency mmWave technology (e.g. WiGig). These exhibit a high degree of heterogeneity [21] as they differ both in terms of operational and performance parameters, including spectrum use; antenna characteristics, physical layer encoding, sharing of the available spectrum by multiple users as well as maximum bit rate and reach. LTE is among the prime wireless access cellular technologies in 4G networks as it is anticipated to offer a theoretical net bit-rate capacity of up to 100 Mbps per macro-cell in the downlink and 50 Mbps per macro-cell in the uplink if a 20 MHz channel is used. These data rates can be further increased through Multiple-Input Multiple-Output (MIMO) technology. At the same time, LTE can provide improved QoS characteristics such as low packet transmission delays, fast and seamless handovers supporting high speed vehicular communications scenarios and operation with different bandwidth allocations. LTE can also support a wide range of services and performance metrics in a wide range of environments such as indoor, urban and rural. As already discussed, BB signal processing functionalities in LTE systems can be performed by the BBUs that can be either co-located with the antenna remote radio heads (Traditional-RAN deployment) or located remotely exploiting the concept of C-RAN [22]. In the present study both network deployments are evaluated emphasizing on metrics such as energy, delay, mobility and density. A high level view of the basic components considered in 5G-XHaul together with the associated interfaces is provided in Figure 3.

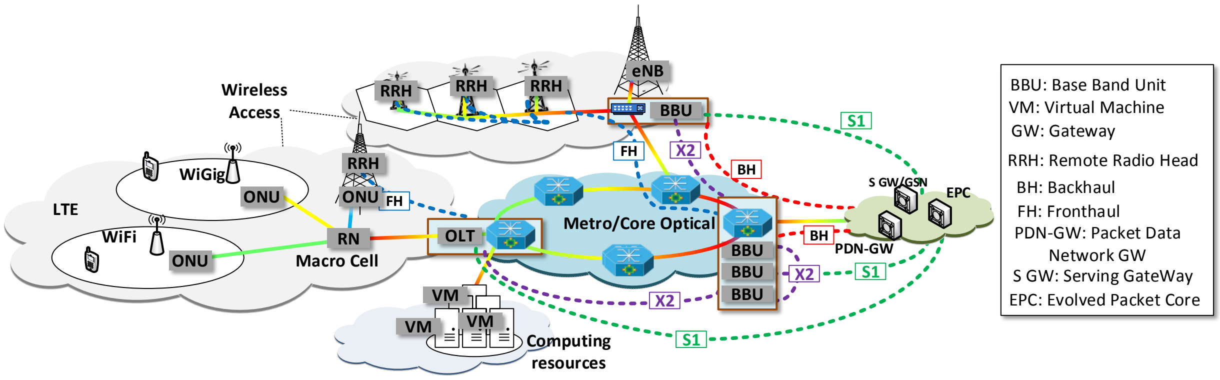


Figure 3: Converged Heterogeneous Network and Compute Infrastructures.

As stated previously, to support the transport network requirements associated with C-RAN [24] we propose the adoption of an optical transport solution comprising passive and active elements, offering high capacity and advanced features including dynamic bandwidth allocation both in the time and frequency domain [26]. In the wireless access domain, spectral efficiency is enhanced by complementing macro-cells with small cells as they allow higher rates of frequency reuse over carefully designed geographical areas with easy access to the network backbone. In addition to small cells, given that WiFi networks are readily available in almost every public or private area and are easy to install and manage, significant benefits are expected by the joint consideration of WiFi and LTE systems. The small cell concept can easily be extended to WiGig to overcome the high inter-cell interference in dense network deployments.

At the same time, given the technology heterogeneity of the proposed infrastructures, a critical function is interfacing between technology domains, including isolation of flows, flexible scheduling schemes, QoS differentiation mechanisms and mapping of different QoS classes across different domains. This can be achieved adopting flexible hardware functions that allow hardware repurposing through concepts such as hardware programmability. Hardware programmability can potentially enable dynamic and on demand sharing of resources guaranteeing also the required levels of isolation and security. In this context, programmable Network Interface Controllers (NICs), commonly used to bridge different technology domains at the data plane, can play a key role. These controllers have a unique ability to provide hardware level performance exploiting software flexibility and can offer not only network processing functions (i.e. packet transactions [26]), but also hardware support for a wide variety of communication protocols and mechanisms [27] such as Virtual Ouput Queuing (VOQ). Through VOQ, a single physical buffer traversed by different flows can be divided into several separate queues with guaranteed performance facilitating hardware sharing through, e.g. virtualization. In 5G-XHaul, the concept of VOQ is introduced both at the Optical Network Units (ONUs) as well as at the interfaces between the Optical Line Terminals (OLTs) and the TSON edge nodes enabling optimized BH and FH service provisioning over the underlying physical resources.

In the following subsection, a modeling framework for the various system components comprising the 5G-XHaul transport network, including the WDM-PON, the TSON as well as the multi-technology wireless access. Is presented. Once this framework has been presented, a detailed evaluation of the performance of the overall solution 5G-XHaul solution under different operational scenarios (i.e. scenario I with BH only, scenario II with FH only and scenario III with joint FH/BH services) is provided.

4.1 WDM-PON

5G-XHaul considers a high capacity WDM PON solution for the interconnection of the RUs with the TSON edge nodes and the BBUs. In contrast to 10G-EPON systems, that rely on sharing a single wavelength over time, the WDM-PON solution considered offers i) the option to assign one or more dedicated wavelength to the RUs supporting connections with very high capacity requirements (i.e. CPRI FH connections) ii) enhanced security as wavelengths may be allocated to the RUs over separated paths and iii) simplified MAC as the ONUs and the OLT are interconnected on a point-to-point basis.

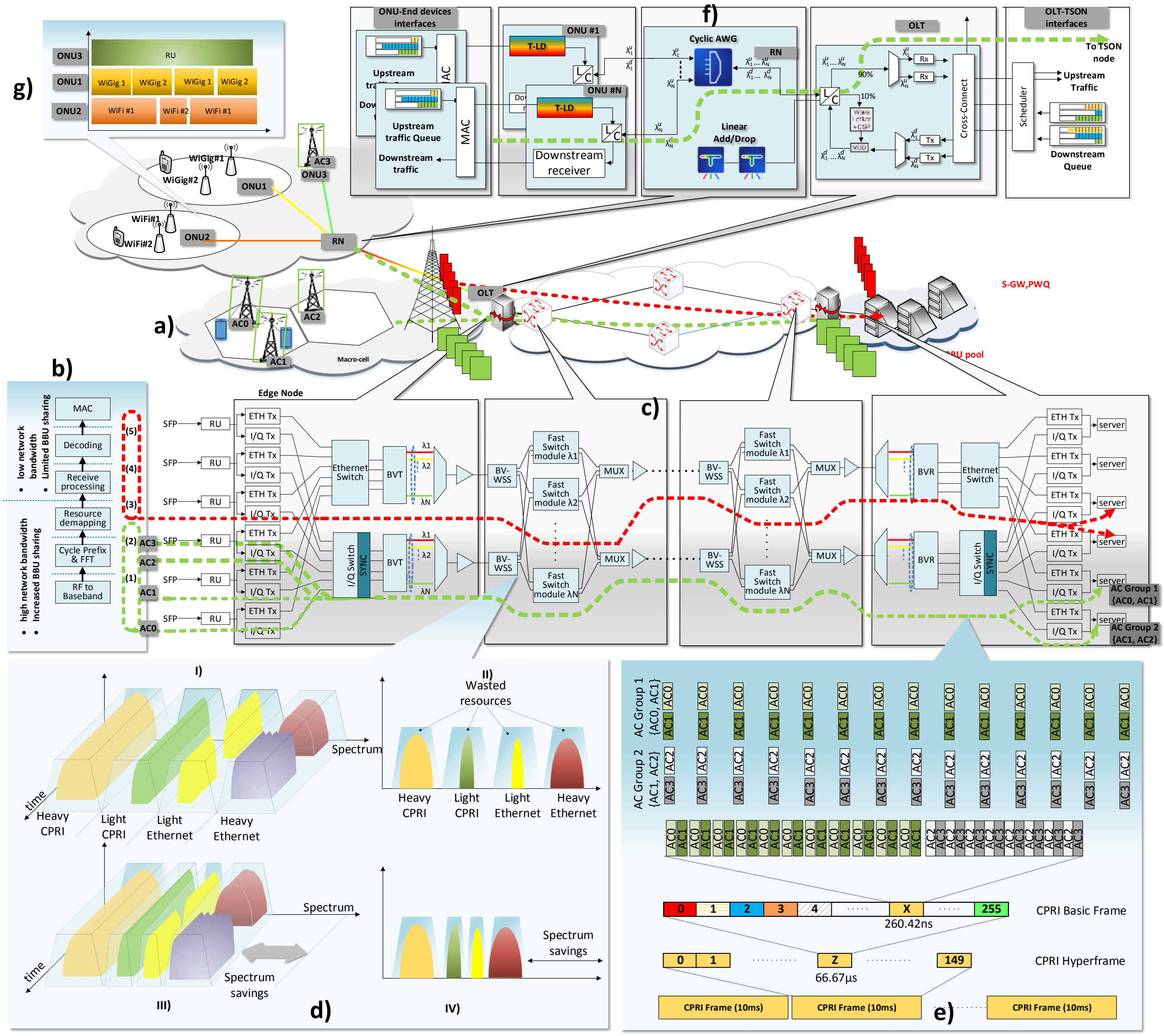


Figure 4: Modelling of 5G-XHaul components and resources: a) Unified mobile FH/BH over converged optical data centre networks, b) Functional split of RU processing [3]-[4], c) Data path interfaces, d) Joint BH/FH resource allocation without and with elastic bandwidth allocation: I-II) Fixed grid bandwidth allocation for FH (Light, Heavy CPRI flows) and BH services (light and heavy Ethernet

traffic, III-IV) Support of the same services through flexible spectrum allocation, e) AxC data stream generation (upper part), multiplexing (middle) and mapping into the CPRI framing structure, f) WDM-PON components, g) allocation of WDM-PON resources for heavy CPRI traffic (ONU 3) and shared lightweight Ethernet flows (ONU1, ONU2).

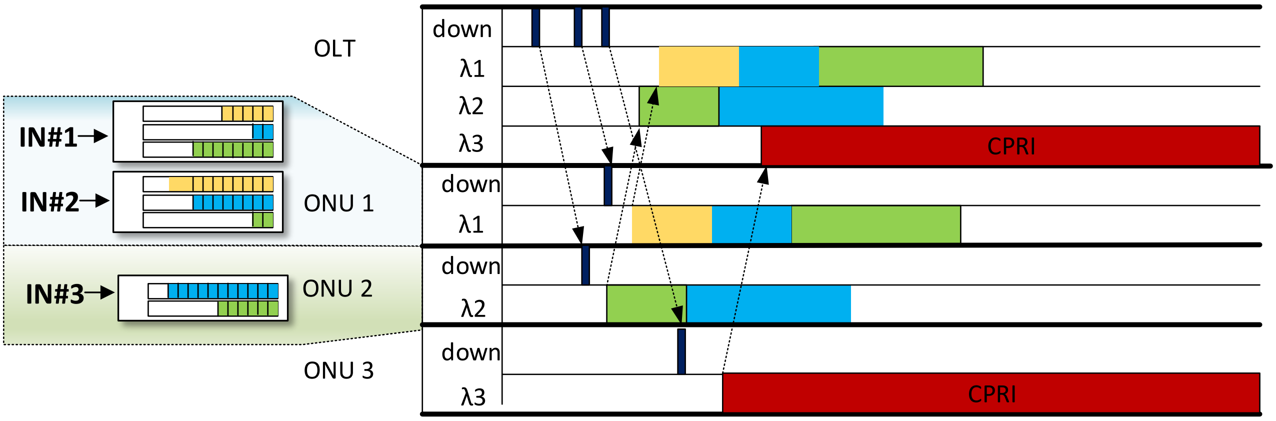


Figure 5: Wavelength and scheduling assignment in uplink WDM-PON.

The WDM-PON key component block diagrams, considered and the associated resource allocation framework is presented in Figure 4 f) and g). In the uplink, traffic generated from the access points is multiplexed through suitable interfaces. These interfaces are required to address the granularity mismatch between the wireless access and the PON network domains and are equipped with buffers that allow the creation of VOQs. The output of these interfaces are used as input at the ONUs. Each ONU is able to adaptively select its transmission wavelength to fit the connected filter port by employing an out-of-band communication channel [44]. The uplink traffic passes through the L/C filters and the Remote Nodes (RNs) before it reaches the OLTs. RNs are equipped with cyclic AWGs that allow multiplexing and demultiplexing of WDM channels. To simultaneously support bidirectional traffic, AWGs are configured to operate in the L and C-bands providing 40 channels per band. Through this configuration in the uplink, channels belonging to a specific waveband are multiplexed and forwarded to the OLTs L/C bandpass filters, whereas in the downstream the reverse operation is performed. Each OLT is also equipped with a cross-connect that allows a large number of external ports to be flexible interconnected with the OLT transmitter and receiver arrays.

To further improve multiplexing gains and address the granularity mismatch between the TSON and the PON network domains the system is also equipped with a second set of interfaces that also allow the creation of VOQs. VOQs allow downlink traffic to be stored and multiplexed before it is forwarded to the suitable transmitter. A typical example of this process is shown in Figure 5, where each ONU is assigned with a dedicated wavelength. For simplicity, it is assumed that the number of receivers at the OLT is equal to the number of wavelengths assigned at the ONUs. However, the analysis can be easily extended to cover the case where wavelengths are shared between the different ONUs. In the uplink each ONU-wireless access interface (left-side of Figure 4 f)) reports the status of its VOQ to the OLT-TSON interface. Following this, the OLT-TSON interface communicates the bandwidth allocation plan produced based on the scheduling policy adopted (weighted fair queuing, FIFO, LIFO, etc.) to each ONU.

4.2 TSON

At the optical network ingress TSON edge node the interfaces receive traffic frames generated by fixed and mobile users (Figure 4 1a). The incoming traffic is aggregated into optical frames, which are then assigned to suitable time-slots and wavelengths for further transmission. At the egress point the reverse function takes place.

The optical edge node is also equipped with elastic bandwidth allocation capabilities supported through the deployment of Bandwidth Variable Transponders (BVTs) [34]. In addition to providing BH functionalities, this infrastructure also interconnects a number of RUs and end-users with a set of general purpose servers [34].The use of general purpose servers enables the concept of virtual BBUs (vBBUs), facilitating efficient sharing of compute resources to support both operational and end-user services. This joint functionality is facilitated by the edge nodes that comprise a hybrid subsystem of an I/Q switch and an Ethernet switch. The I/Q switch handles transport classes with strict synchronisation and bandwidth constraints – i.e. split options

(1) and (2) – while the Ethernet data switch handles BH traffic and relaxed FH transport classes, i.e. split options (3)-(5). An analysis of these splits is provided in [33]. The multi-protocol interfacing requirements described above are facilitated through a Serializer/Deserializer (SerDes) module that performs serial-to-parallel data conversion from the RU data stream to the parallel hybrid I/Q-Ethernet switch [45] The I/Q data streams are supported by a synchronisation block that manages the synchronisation signals between the end points. The I/Q switch is also capable of routing data streams from any ingress to any egress port and timeslot with very low latency [45]. This allows mapping of the Antenna Carrier (AxC) streams to the suitable optical network or BBU resources. A typical example is illustrated in the lower right-part of Figure 4c, where through I/Q switching wireless samples AxC1-AxC4 are grouped, forwarded and assigned to suitable vBBUs (Figure 4e). At the ingress part of the Ethernet switch module a set of gigabit links is provided to interconnect RUs and BBUs, whereas at the egress part of the module, traffic is aggregated and placed to a suitable transmission queue. As already mentioned, data from I/Q and Ethernet switches are transported via the elastic optical network to the vBBUs. The elastic optical network solution adopted comprises BVTs, BV optical cross-connects and fast optical switching modules. This approach enables the allocation of variable size-spectral/time slots [11] thus supporting services with continuous channel allocation at various bit rates (i.e. heavy and light CPRI) and services with sub-wavelength time-slot allocation (Ethernet flows). A relevant discussion was also provided in deliverable D2.2.

4.3 End-to-end network modelling and Optimization

To address the great diversity of requirements introduced by the upcoming services in a cost-effective and energy efficient manner, optimal resource assignment considering the unique application and device characteristics is needed. In achieving this goal, the development of intelligent optimisation algorithms considering different Key Performance Indicators (KPIs) (i.e. capacity, latency, energy consumption) for all physical and virtual network providers can play a key role. In the SDN/NFV architecture, this process is located at the infrastructure management and control layers, offering to network service providers suitable tools that can assist in performing a broad range of tasks, including [29]:

- activities related to service chain (SC) management, able to Create/Delete/Update network SCs and a set of other relevant network functions,
- management of SCs considering virtual and/or physical resources and definition of traffic rules to dictate the selection of the optimal chain out of a set of possible chains,
- scale-in/out functionalities such as the, ability to bring up/down multiple network functions on an on-demand basis,
- traffic offloading from one forwarding entity to another,
- unified orchestration of compute and network elements,
- service orchestration with legacy or third party Operation Support System (OSS).

The combination of these tools facilitates the support of any mix of services, use cases and applications and can assist in addressing both technical and business challenges anticipated to arise in future network infrastructures. A specific use case that can be used to highlight the role of these tools is the provisioning of Content Delivery (CD) Services in mobile cloud computing (MCC) environments, deploying a heterogeneous network infrastructure. MCC enables mobile devices to overcome their inherent processing, storage and power limitations by offloading intensive tasks (or accessing hosted content) to remotely located DCs.

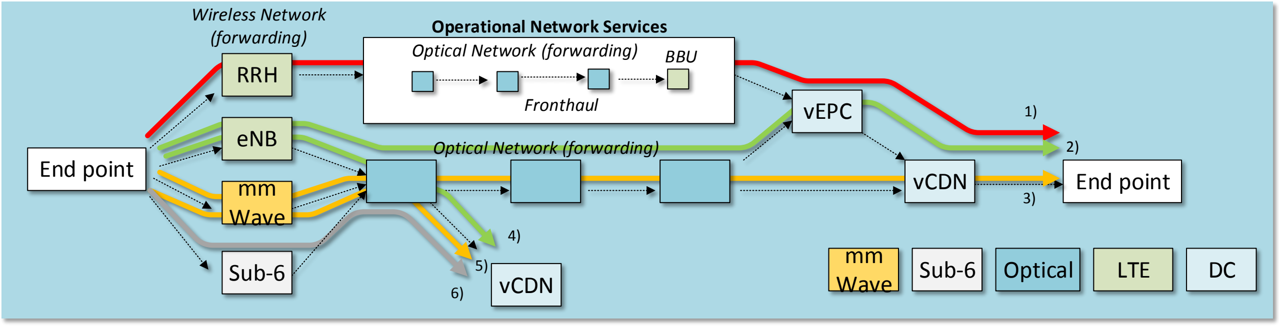


Figure 6: Service chaining over heterogeneous network infrastructures in support of content delivery services: 1) vCDN over C-RAN, 2) vCDN over LTE, 3) vCDN hosted at remote DCs through WiGig, 4)-6) vCDN hosted at local DCs over LTE, mmWave and Sub-6 wireless access respectively.

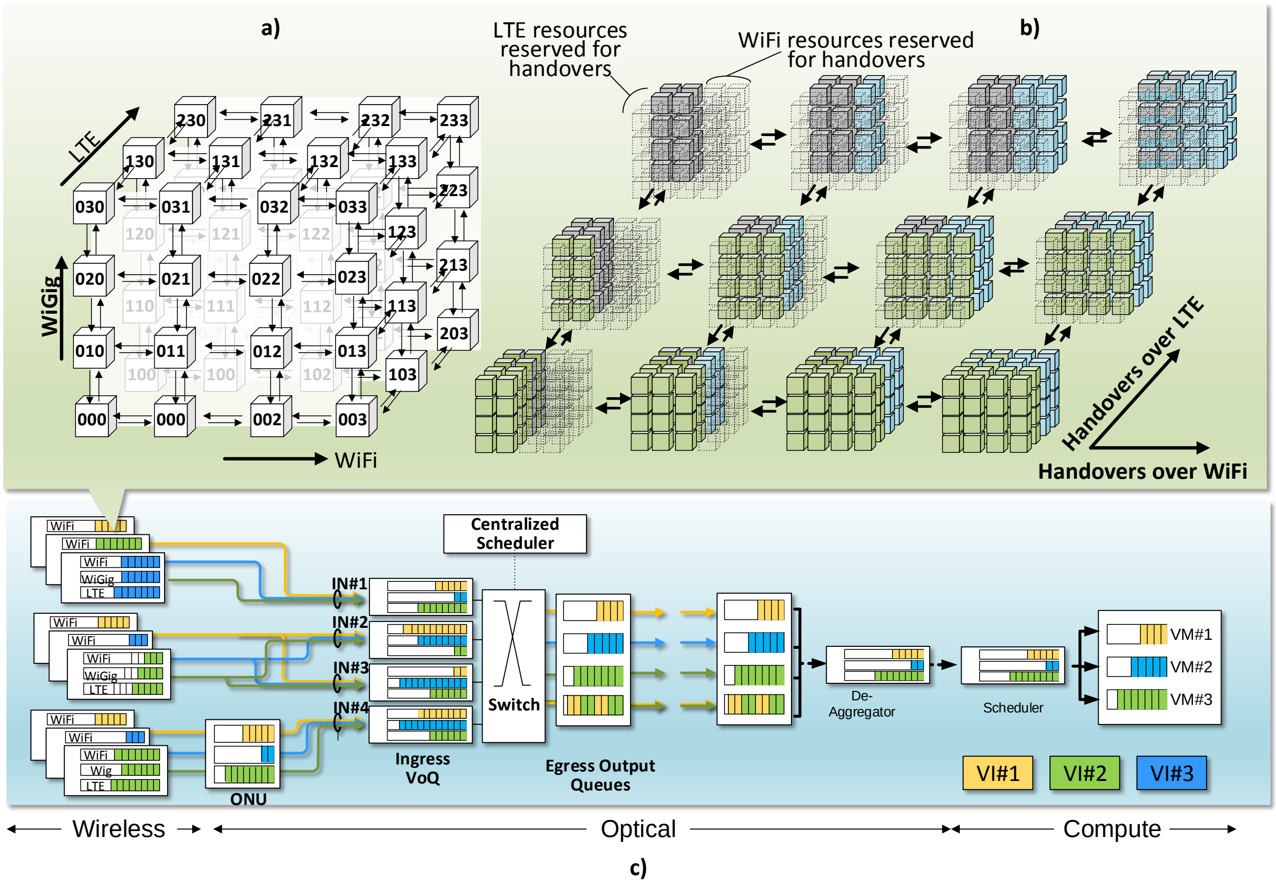


Figure 7: Modelling queuing delays in converged network environments a) three-dimensional Markov chain for estimating delays in the wireless access network with traffic offloading capabilities from one forwarding entity to another b) five-dimensional Markov chain modelling mobility, c) End-to-end model as a network of queues (uplink).

To provide CD services, the orchestrator instantiates different type of VNFs that are deployed and chained together, each having specific processing and bandwidth requirements. Based on the type of wireless access technology (i.e., RRH, eNB, WiFi, WiGig) used to forward data from mobile devices (End-Point A in Figure 6 to the optical transport and the Data Centres (DCs) where CD services are hosted, multiple candidate SCs can be created. To realize each SC, sufficient network bandwidth and processing capacity must be allocated, corresponding to specific physical resources for the interconnection and deployment of VNFs. VNFs are then processed in the order defined by the corresponding SC. For example, SC1 in Figure 6 illustrates the case of mobile data offloading to remote DCs over the RRHs. To realise this, wireless signals received by the RRHs are forwarded over an optical transport network to the BBU pool and then to the DC location. Flow conservation as well as mapping and aggregation/de-aggregation of traffic between different domains should be also satisfied.

Apart from network and capacity constraints, end-to-end delay is an important KPI that needs to be also considered in the analysis. In highly loaded heterogeneous networks, end-to-end delay can be greatly influenced by queuing delays associated with the interfaces. Therefore, applying specific queuing policies and scheduling strategies at these locations is very important. Significant delay benefits can be achieved by instantiating the necessary network functions and reserving the required virtual/physical resources. End-to-end delay can be mathematically modelled through queuing models and the adoption of closed-form approximations derived by modeling the different network domains as open, closed and/or mixed queuing networks. An example is illustrated in Figure 7 (c), where a three-dimensional Markov chain is adopted to model the three wireless access technology domains, i.e. LTE, WiGig and WiFi. Each dimension of the Markov chains corresponds to a different virtualised wireless access domain with its state space defined as $\mathcal{S} = \{(i,j,k)|i \leq \mathcal{I}, j \leq \mathcal{J}, k \leq \mathcal{K}\}$, where i, j and k correspond to the virtualised resources used across the LTE, WiGig and WiFi dimension respectively, and (i,j,k) is a feasible state in $\mathcal{S}$. Note that $\mathcal{I}$, $\mathcal{J}$ and $\mathcal{K}$ correspond to the maximum set of resources that can be allocated to a specific provider. A key characteristic of the proposed scheme is that it allows modeling of traffic offloading decisions from one entity to another (i.e. WiGig to WiFi or LTE) as well as modeling of the arrival of a new service request by modifying the corresponding state, i.e. $(i,j,k) \rightarrow (i+1,j,k)$, when a new forwarding decision is applied through the LTE network. The steady state probabilities of the Markov process can be determined in a unique way using the well-known matrix-geometric solution techniques and the corresponding service delay can be determined.

Markov chain models can be effectively used to evaluate the performance of domains where statistical independence between arrivals and services exists. Therefore, they can be applied to describe scenarios where virtual resources are realized through isolated physical resources such as different channels, spectrum, wavelengths etc. However, these models cannot be extended to technology domains where common buffers are shared among multiple virtual flows. A typical example for this exception applies to the edge nodes of the optical domain where common FIFO queues can be traversed by several virtual flows. A solution to this problem is to adopt the concept of VOQs that can achieve traffic isolation among flows, providing at the same time flow-level bandwidth provisioning with strict delay guarantees (see Figure 7 c) [30]. It should be noted that VOQs do not refer to physical entities, but correspond to pointers pointing to specific packets of the physical queues. In practice, they can be implemented in programmable hardware through the development of appropriate flow scheduling algorithms whereas centralised control can be implemented in OpenFlow.

4.4 Mobility considerations

An additional consideration to be taken into account during the operation of this type of infrastructures is end-user mobility. To handle mobility, redundant physical resources should be reserved to support uninterrupted Service Chaining (SC). The amount of redundant resources increases with the speed of end-user mobility, the size of the wireless cells (mobile users associated with small cells will exhibit very frequent handovers) and the traffic model adopted. Based on their technical characteristics, the wireless access technologies adopted in this work, can address end-user mobility with different levels of effectiveness. E.g. WiGig providing high capacity levels is most suitable for indoor environments with limited end-user mobility, whereas WiFi and LTE can support lower capacities but higher mobility levels, with LTE being the most suitable technology for high speed vehicular communications. To maximise the benefits provided by the available technologies, users with low mobility are offloaded to the WiGig domain releasing WiFi and LTE resources for mobile users. It is clear that 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 handover probabilities across WiGig and LTE cells, and to reserve a specific set of resources for handover purposes.

Similarly to the static cases described above, a five-dimensional Markov chain can be adopted to evaluate the performance of the virtualised wireless access network under mobility where two additional dimensions have been introduced to model handovers across WiFi and LTE. Given that users are much more sensitive to call dropping than to call blocking, a percentage of the virtualised resources is reserved for handovers. Thus, new service requests can use resources up to a specific threshold above which new requests are dropped. On the other hand, mobile users are dropped when all resources are already in use. In this case, also, a closed-form approximation of the systems' state probabilities can be extracted using dimensional reduction techniques. The redundant resource requirement imposed for mobility purposes also propagates from the wireless access domain to the optical network and compute domains as depicted in Figure 7 c.

Taking into account the aforementioned considerations, a MOP problem can be formulated in such a way that optimizes the performance of the converged network and computation infrastructure considering also the battery lifetime of the mobile devices under delay and mobility constraints. The output of this optimisation problem can drive the selection of the optimal SC out of a set of multiple chains. In addition, it can identify possible locations where VNFs or PNFs can be placed as well as the optimal wireless access technology that should be used.

5. Backhaul services: Performance Evaluation

As already discussed, 5G-XHaul proposes a widely heterogeneous but converged optical-wireless 5G network infrastructure interconnecting compute resources with fixed and mobile users. To evaluate the performance of BH services over the heterogeneous 5G-XHaul infrastructure the modelling framework described above has been adopted and a set of BH service scenarios have been studied.

5.1 Simulation Environment and Parameters

The proposed framework is evaluated using the infrastructure topology illustrated in Figure 1. This infrastructure covers a 10x10 km^2 area over which 50 RU units are uniformly distributed and comprises a set of optical edge nodes in the optical segment, and optical point-to-point links for fronthauling the RRHs. The optical network technology adopted deploys a single fiber per link, 4 wavelengths per fibre, wavelength channels of 10 Gbps each, minimum bandwidth granularity of 100 Mbps and maximum link capacity of 40 Gbps. The power consumption model for the optical nodes is provided in [25]. Furthermore, a 2×2 MIMO transmission with adaptive rank 10 MHz bandwidth adjustment has been considered, while users are distributed and generate traffic over the serviced area according to real datasets reported in [32]. This traffic, corresponding to wireless access traffic generated by a set of mobile devices, needs to be processed by a specific set of compute resources. Each mobile device is equipped with a processing unit with capacity of 400 Million Instructions per second (MIPS). Its power consumption is 0.3 Watts (W) and 1.3 W under idle and transmission/reception mode, respectively, whereas under processing mode it consumes 0.9 W. The whole area is also covered by a set of WiFi access points offering 135 Mbps capacity with power consumption of 1.28 W during data transmission, 0.94 W during data reception, 0.82 W in idle mode and 64 mW in sleep mode. In addition to this, there exist a set of WiGig access points providing indoor coverage offering 1 Gbps data rate with 2 W power consumption under peak load Finally, each DC has a processing capacity of 80 Giga IPS and its power consumption follows the step-size power consumption model [25].

5.2 Simulation Results

The objective of the optimisation framework is to identify the optimal SCs in order to jointly optimise the performance of converged network and computation infrastructure as well as the battery lifetime of the mobile devices. The former can be achieved by identifying the optimal routing paths and the location of the DCs where demands need to be processed, whereas the latter can be achieved by ensuring that computation offloading is always beneficial for the mobile devices in terms of energy consumption. It should be noted that computation offloading is beneficial if the total energy consumed by a mobile device, for transmitting and receiving data, is at least equal to the total energy consumed for data processing by the mobile device itself without violating QoS specifications.

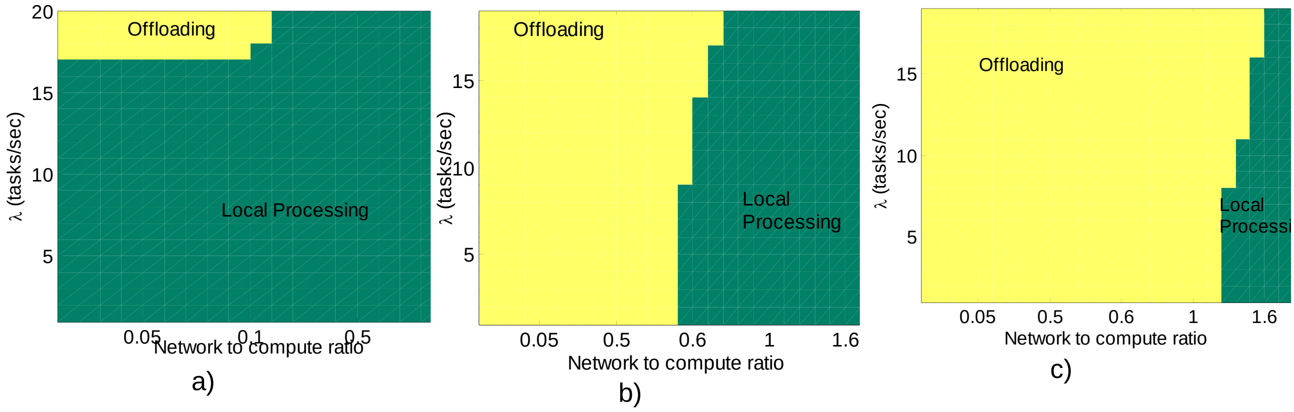


Figure 8: Traffic offloading regions as function of the service type under various levels of convergence a) Converged LTE with optical DC networks, b) Converged LTE/WiFi with optical DC networks and c) Converged LTE/WiFi and WiGig with optical DCs.

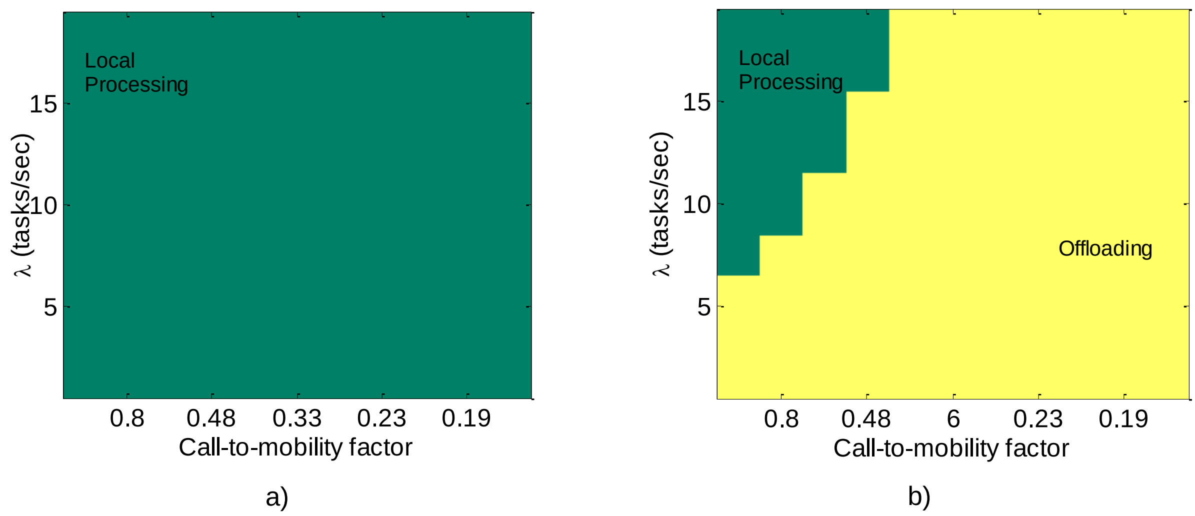


Figure 9: Traffic offloading regions as function of mobility over a) Converged LTE/WiFi with optical DC networks and c) Converged LTE/WiFi and WiGig with optical DCs (network to compute ratio=1).

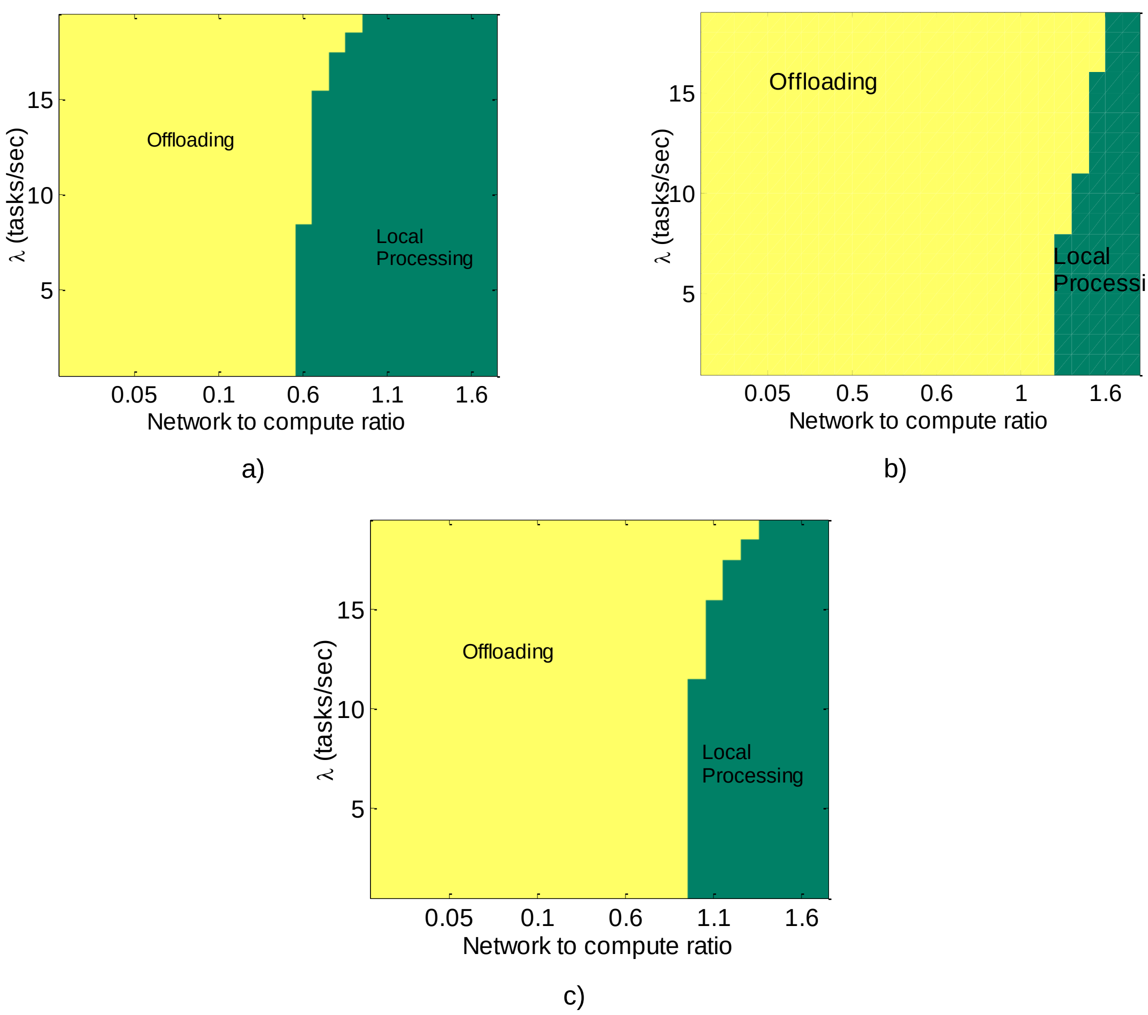


Figure 10: WiGig Penetration a) 10%, b) 50%, c) 100%.

Figure 8-Figure 10 illustrate the impact of service requirements in terms of network and computational resources, as well as that of mobility on the decision of a mobile device to offload its tasks to the cloud. To achieve this, the Network-to-compute resources and the Call-to-mobility [31] ratios are introduced to capture the communication cost and the speed of the mobile devices, respectively. The former is used to capture the relation between computational and network bandwidth requirements, while the latter is defined as the fraction of the service holding time over the cell residence time. From Figure 8 it is observed that traffic offloading to the cloud is beneficial for the mobile devices when communication cost is low (low networking requirements) and the processing load is high. It is also seen that for higher degree of network convergence we are able to support mobile cloud services with extremely high networking requirements, e.g. multimedia services. At the same time, with the increase of the penetration of WiGig technology, we are able to unlock mobile cloud computing to a broader range of services (Figure 9).

As also discussed in [25], when mobility is high (lower call-to-mobility factor), additional resources are required to support the seamless handovers in the wireless access domain. This additional resource requirement also propagates in the optical metro network and the DC domain in order to ensure availability of resources in all domains involved (wireless access and backhauling, optical metro network, and DCs) to support the requested services and enable effectively seamless and transparent end-to-end connectivity between mobile users and the computing resources. This leads to underutilisation of network resources and therefore increased delays, which deteriorate the benefits associated with traffic offloading. The present approach, through the higher degree of consolidation and the better utilisation of the network resources it offers, is able to handle high degrees of mobility and also support services with significant communication requirements in a very efficient manner. It is also observed that with the increase of service processing requirements (Figure 10) offloading is beneficial for larger mobility and network-to-compute ratio regions. It is clear that, for the case where computational intensive data is processed locally, in-device processing latencies will be higher than the transmission delays, leading to increased power consumption for the mobile device.

6. Frontraul Services

As discussed in section 1, FH describes the segment of the transport network that interconnects RUs without full BB functionality with a CU where the remaining BB functionality is located. The CUs are usually located in central offices (COs) (c.f. deliverable D2.4 [35]). 5G-XHaul focuses on the so-called functional splits, which define a variable distribution of BB functionality between RUs and CUs. In deliverable D2.1 [1], several functional splits have been investigated in detail, being three of those selected in deliverable D2.2 [34] for further investigation. This section presents the final results on the functional split investigation and evaluation. To this end, the selected functional splits are briefly recaptured and compared to the state-of-the-art. Based on recent progress in 5G RAN standardisation, updated requirements in terms of data rates are presented. From 5G-XHaul measurements performed in a live LTE network, we evaluate the requirements in comparison to the state-of-the-art and present final results on statistical multiplexing gains.

6.1 Functional Splits and Final Peak Requirements

The three functional splits selected in 5G-XHaul for evaluation are referred to as Split A, B and C, and are illustrated in Figure 11. The state-of-the-art split is standardised in the CPRI standard [38] and is also illustrated for comparison.

The main difference between Split A and CPRI is that beamforming is performed at the RU and, hence, only one stream per analogue-to-digital converter (ADC) chain/antenna port instead of per antenna element is transported, which is especially beneficial for massive MIMO.

Split B prevents unused resource blocks from being transported over the FH and, hence, the data rate on the transport link depends on the current cell utilisation. This is important to enable statistical multiplexing.

In Split C, Forward Error Correction (FEC) encoding/decoding is performed at the RU, leading to transport data rates depending on the currently selected modulation and coding schemes (MCS), i.e. on the users' channel quality.

In deliverable D2.1, the theoretic peak requirements for the three splits were derived for three different potential future radio access technology (RAT) configurations. Based on recent developments in 5G pre-standardisation, namely 3GPP TR 38.319 [37], the parametrisation and resulting peak requirements have been updated in Table 1.

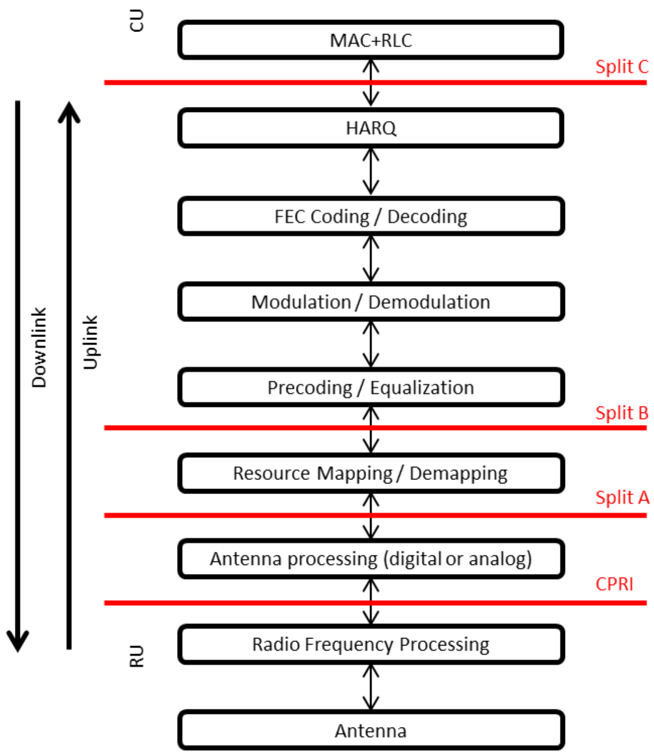


Figure 11: Functional splits.

Table 1: Current and potential 5G radio access parameters.

Parameter	Symbol	LTE	Sub-6	Low mmWave	High mmWave
Carrier Frequency [GHz]	f_C	2	2	30	70
Channel Size [MHz]	BW	20	100	250	500
Sampling Rate [MHz]	f_S	30.72	150	375	750
# Antennas	N_A	4	96	128	256
# ADC/DAC chains	N_P	4	16	12	10
# Layers	N_L	4	16	12	10
Overhead	γ	1.33	1.33	1.33	1.33
Quantizer resolution time domain	$N_{Q,T}$	15	15	12	10
Quantizer resolution frequency domain	$N_{Q,F}$	9	9	8	7
Modulation order	M	64	1024	256	64
Max. code rate	R_C	0.85	0.85	0.85	0.85
Frame duration [ms]	T_F	1	1	1	1
FFT size	N_{FFT}	2048	2048	2048	2048
# Active subcarriers	$N_{SC,act}$	1200	1300	1300	1300
# Data symbols per frame	N_{Sy}	14	70	150	300
Peak utilization	μ	1	1	1	1
Formula data rate split CPRI	$D_{CPRI} = 2 \cdot N_A \cdot f_S \cdot N_{Q,T} \cdot \gamma$				
Formula data rate split A	$D_A = 2 \cdot N_P \cdot f_S \cdot N_{Q,T} \cdot \gamma$				
Formula data rate split B	$D_B = 2 \cdot N_P \cdot N_{SC,act} \cdot N_{Sy} \cdot N_{Q,F} \cdot T_F^{-1} \cdot \mu \cdot \gamma$				
Formula data rate split C	$D_C = N_L \cdot N_{SC,act} \cdot N_{Sy} \cdot R_c \cdot \log_2 M \cdot T_F^{-1} \cdot \mu \cdot \gamma$				
Peak data rate split CPRI [Gbps]	D_{CPRI}	4.9	574.6	1532.2	5107.2
Peak data rate split A [Gbps]	D_A	4.9	95.8	143.6	199.5
Peak data rate split B [Gbps]	D_B	1.6	34.9	49.8	72.6
Peak data rate split C [Gbps]	D_C	0.46	16.5	21.2	26.5

The main changes concern the carrier frequency of the high mmWave carrier, which was changed from 80 GHz to 70 GHz, the bandwidth and number of antennas, which were reduced in light of the assumptions in [37], the maximum modulation order, and the number of subcarriers, which is assumed to be higher in 5G due to the utilization of potential new waveforms [41] or more efficient filtering. In addition, the parameters are compared to a standard LTE configuration in Table 1.

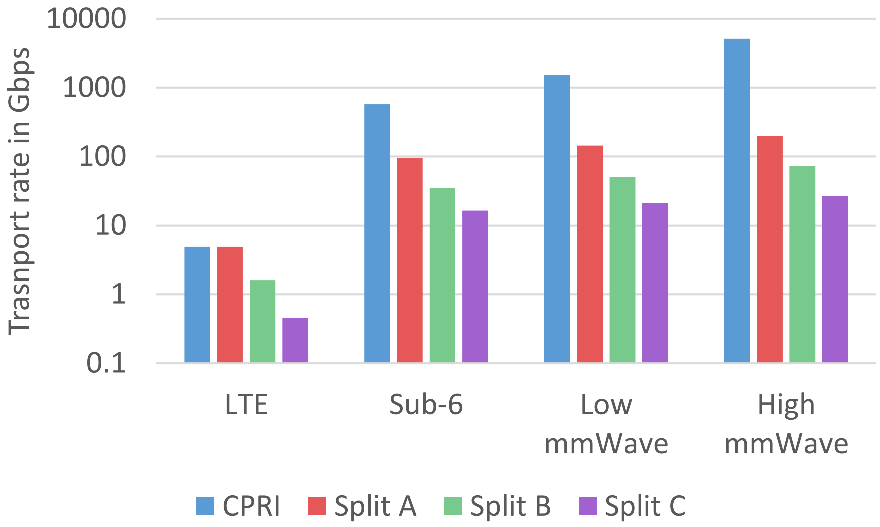


Figure 12: Peak transport data rate requirements for different functional splits and different radio access technologies.

Figure 12 graphically illustrates the peak data rate requirements for all four parametrisations and the four functional splits. Two important observations can be made. First, the introduction of new RATs for 5G will lead to an increase in transport data rates. Due to the envisioned increase of user data rates of factor 10-100 compared to current systems, a similar increase in the transport segment can be considered inevitable. Second, maintaining a CPRI-like split could increase the transport data rates into the range of several Tbps, which no current or near-future technology would be able to transport (compare e.g. deliverable D2.2, Sec. 4). However, as can be seen, the functional splits investigated by 5G-XHaul dramatically reduce the peak data rates, providing the necessary prerequisite for a feasible and cost-efficient 5G transport network.

The need for new functional splits has also been realized by relevant standardization bodies, namely IEEE 1914.1 [39] and eCPRI [38], both of which consider new splits. 5G-XHaul provided its expertise to the former, thereby actively impacting standardisation.

It is necessary to note that, even with the new functional splits, the higher data rates in Table 1 are well above currently available technology. Accordingly, 5G-XHaul is not focusing to implement technology to transport data rates in the 100 Gbps range. Instead, 5G-XHaul focuses on the flexibility of the FH links, which is the second major requirement for 5G transport networks, as will be described in the next section.

6.2 Evaluation of Statistical Multiplexing Gains

6.2.1 Introduction and Summary of Previous Findings

Flexibility is a key requirement for future 5G networks given the different requirements arising from the parallel use of different functional splits and RATs, relying on the simplification in network management that can be achieved through SDN. Packet-based FH is hence considered an important enabler for future transport networks (c.f. deliverable D3.1 [36]) and, accordingly, more flexible transport technologies such as TSON (c.f. deliverable D2.2, Section 4) are being developed within 5G-XHaul. Especially in the FH segment, packet-based FH, in connection with the new functional split, will lead to improvements via so-called statistical multiplexing gains. These gains will be evaluated here.

In deliverable D2.1, measurements were performed within 5G-XHaul that represent usage statistics from a live LTE network. These were extrapolated to a 5G scenario, and a methodology was described to evaluate these statistics to evaluate statistical multiplexing in the transport network. Statistical multiplexing gains occur in the aggregation segment of the transport network, i.e. after nodes, which aggregate the traffic from multiple cells. If the traffic in the individual cells varies, so will the aggregated traffic. However, as it is unlikely that all cells will exhibit peak load at the same time, the aggregation link does not have to be dimensioned for sum peak traffic. This requires a certain outage probability to be accepted as in very rare cases it is possible that all cells do exhibit peak traffic at the same time. However, a certain outage due to other effects (power failure etc.)

needs to be considered in any case this is acceptable in practice. The effect of statistical multiplexing is based on the central limit theorem, and is also a motivation for BB pooling and virtualisation in C-RAN, as the same effect is observed for required processing power in the CU. In summary, statistical multiplexing requires several factors: a varying traffic, aggregation of several cell, the acceptance of an outage probability, and, first and foremost, a transport network flexible enough to facilitate the statistical multiplexing. Current point-to-point fibre FH links can neither aggregate traffic from several cells, nor can they reflect the variation of traffic, as they operate with a fixed rate. The TSON and WDM-PON technologies developed in 5G-XHaul are hence important enablers to support the gains described in the following.

6.2.2 Final Results on Statistical Multiplexing

Figure 13 shows the required transport data rate versus the number of aggregated cells derived from the measurements reported in deliverable D2.1, which are updated with respect to the parameters in Table 2. Recall that two scenarios were investigated, one with the low load measure in the LTE network, another for a higher load reflecting a 5G scenario. The results are given for all four RATs and the three functional splits investigated within 5G-XHaul. Recall that the required transport data rate of C cells was given as (c.f. deliverable D2.1, Eq. (5.9))

$$R_{\text{req}} = \max(Q_{100}(1), Q_{95}(C)), \tag{6.1}$$

with $Q_x(C)$ being the x-th percentile of the aggregated data rate of C cells, resulting in an outage probability of $P_O = (100 - x)$ %. The term max accounts for the fact that at least the peak capacity of a single cell should be provided (see [40]), as statistical multiplexing cannot be relied upon for low numbers of cells.

The most important observation to be made is that 5G networks will increase transport traffic up two orders of magnitude. This increase can be mitigated by the aforementioned introduction of the new functional splits, leading to less demanding transport data rates.

Figure 14 shows the multiplexing gain, which can be expressed as the ratio of the required aggregated data rate assuming 5 % outage over the n-fold peak data rate of a single cell:

$$g_{\text{mux}} = \frac{\max(Q_{100}(1), Q_{95}(C))}{C \cdot Q_{100}(1)} \tag{6.2}$$

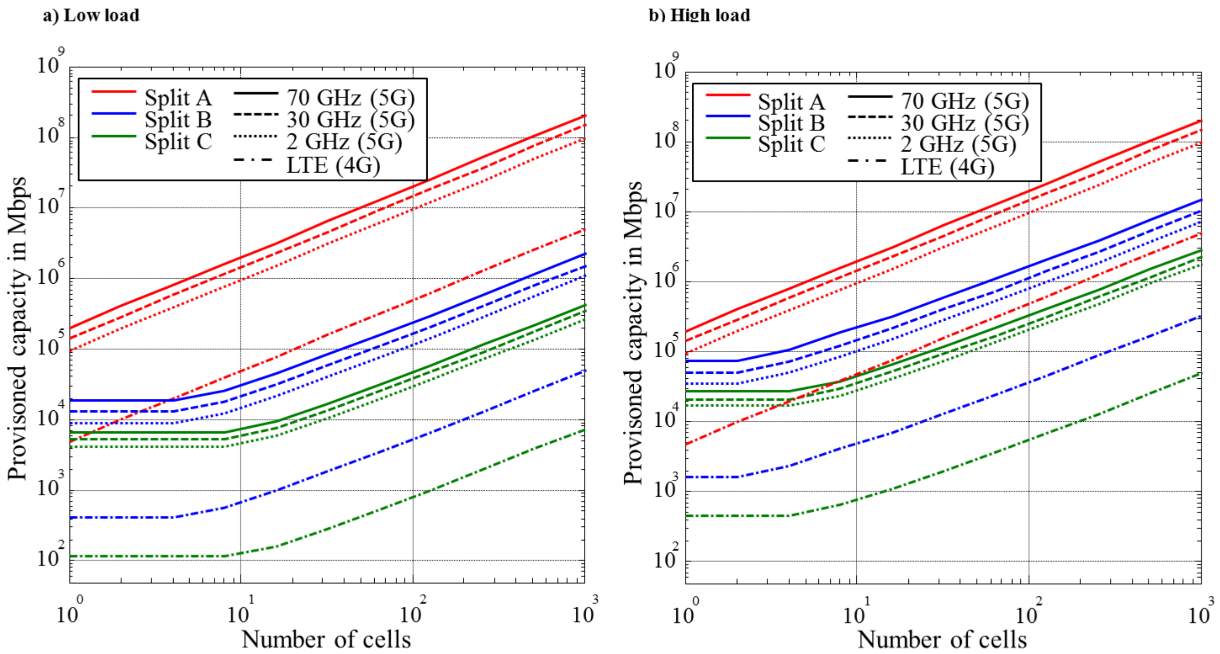


Figure 13: Aggregated data rate requirements for a) low load b) high load scenario.

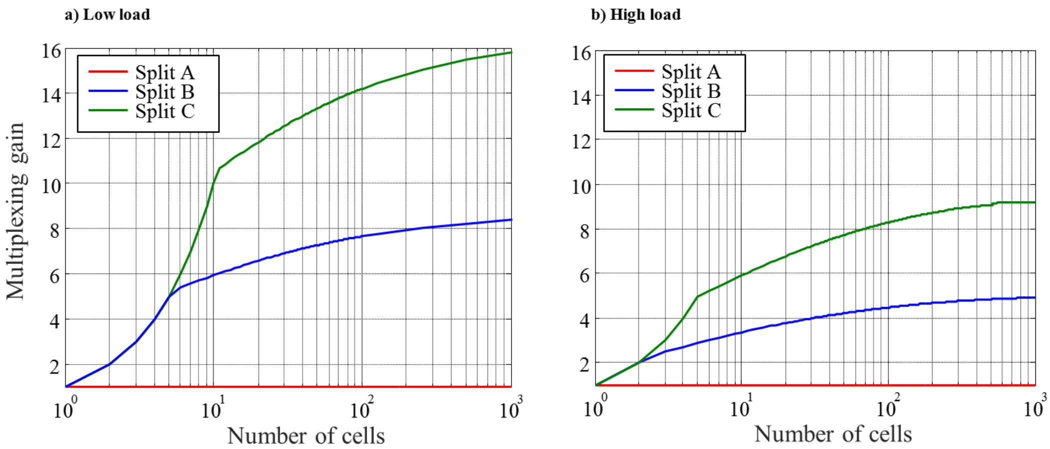


Figure 14: Statistical multiplexing gain for $P_O = 0.05$ for a) low load and b) high load scenario.

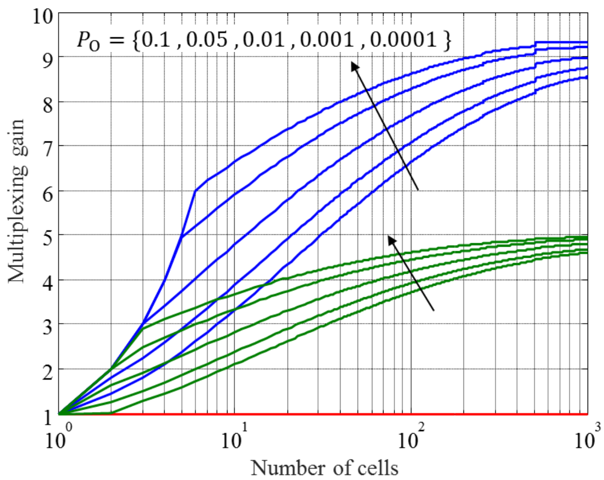


Figure 15: Statistical multiplexing gains for high load scenario and different outage probabilities.

First, it is apparent that split A does not yield multiplexing gains as the transport rate is fixed for this split. Second, the gain is higher for split C, as here the traffic depends not only on the resource utilization the air interface, but also on the coding overhead based on the chosen MCS. As discussed in D2.1, the higher variability introduced by this in general leads to higher multiplexing gains. Third, it can be seen that the multiplexing gain increases with the number of cells, up to a factor of 16 for low load and 1000 cells. It is also higher in the low load scenario, due to the overall lower load there. However, even for the high load scenarios and fewer cells, considerable gains can be achieved. Consider as a first example that three cells can already multiplexed at a three-sector site, this would yield a multiplexing gain of factor 3 and 2.5 (for split C and B, respectively) between the site and the central office/ active node even in the high load scenario. Consider as a second example D2.4, Sec. 5, where at least 69 cells are aggregated at an active node, this would yield multiplexing gains of approximately factor 8 and factor 4.3 (for split C and B, respectively).

In these examples, an outage probability of 5 % was assumed, lower outage probabilities might be desired, e.g. for the ultra-reliability use-case. To illustrate the effect of this, Figure 15 shows the statistical multiplexing gains for the high load scenario for outage probabilities of 10 % - 0.01 %. Table 2 lists the corresponding outage times in the busy hour per year. Figure 16 additionally illustrates the multiplexing gains for the two previous examples of 3 and 69 cells for different outage probabilities. Note that, in general, the term outage represents the offered traffic of a given time instance, e.g. hour, minute, packet duration, Transmission Time Interval (TTI), etc., that cannot be transported by the FH network during that time instance. This could be reflected in the QoS by different effects, as a potentially tolerable increase in delay or as a packet loss due to buffer overflow.

Table 2: Outage probabilities and outage time.

Outage probability P_O	Outage time per year
0.1	36.5 h
0.05	18.25 h
0.01	3.65 h
0.001	21.9 min
0.0001	2.19 min

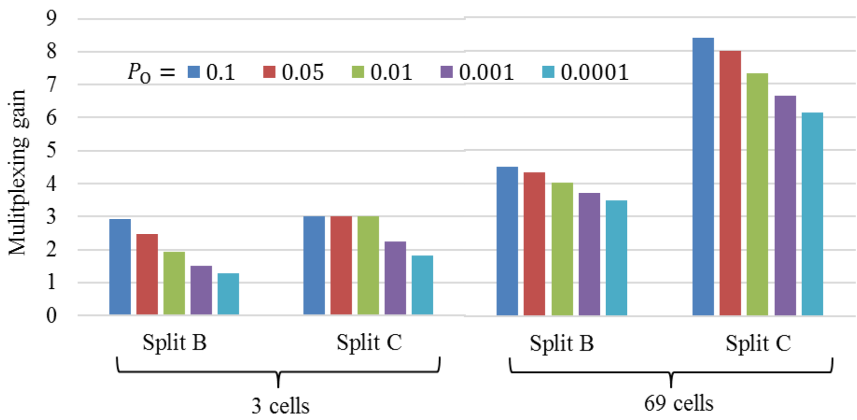


Figure 16: Statistical multiplexing gains for 3-cell site and a 69-cell central office.

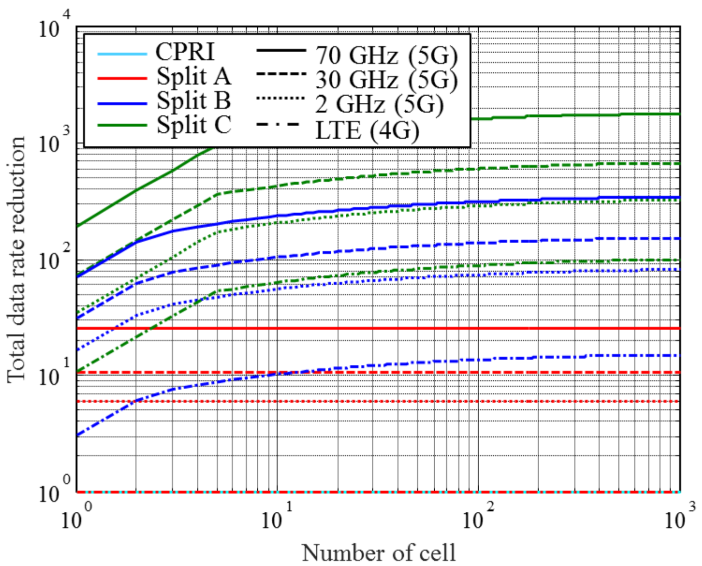


Figure 17: Total transport data rate reduction by functional splits and statistical multiplexing for high load scenario and $P_O = 0.05$.

In addition to the statistical multiplexing effect, the proposed splits also reduce transport traffic in general compared to CPRI, as was illustrated in Figure 17. To show the overall transport capacity saving that the 5G-XHaul approach enables, Figure 17 illustrates the overall data rate reduction compared to CPRI, given as:

$$g_{\text{total}} = \frac{C \cdot Q_{100,\text{CPRI}}(1)}{\max(Q_{100,\text{Split X}}(1), Q_{95,\text{Split X}}(C))} \tag{6.3}$$

As can be seen, the overall reduction is considerable and can reach up to three orders of magnitude. These reductions lead to the following benefits:

- The reductions **enable 5G C-RAN deployment,** as the currently utilised CPRI approach would have led to data rates that cannot be transported by current technologies. Without these reductions, only decentralized deployments would be possible for 5G networks.
- The reduction in required transport capacity **lowers costs and energy consumption** in future transport networks. While the specific savings depend on many factors, e.g. deployment scenario, utilized transport technology, and availability of pre-existing transport infrastructure, it can be expected that the savings will be considerable.

7. Joint Fronthaul and Backhaul Services

As already discussed, one of the main innovations of the proposed 5G-XHaul solution is at data plane level, 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 (Section 2). In addition to this, the higher layers of the architecture, which facilitate access and management of both network and compute resources, also play a key role. More specifically, the ability of the IML to create Virtual Infrastructure (VI) slices across heterogeneous domains and to expose these to the upper layers is an instrumental architectural tool, facilitating the delivery of FH and BH services.

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

- Ordering, referred to as Service Chaining (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 Physical Infrastructure (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 1 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 by, e.g. 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 Physical Infrastructrue (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). 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 6. 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 1 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 [11]. The results obtained using this modelling 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 summarised 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 [14]. Given that both RUs and 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 MIPS) 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 [16] 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 [4]-[5], 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), and more details can be found in [5].
- 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 [15].
- 5 G-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 of 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 [4], [5].

Based on the functional split adopted, part of the processing can be performed by compute resources either at a local cloudlet [14] 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 the 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 minimisation of the cost:

$$min\, \mathcal{VFH}(\boldsymbol{u},\boldsymbol{\pi}) = \sum_{e\epsilon\mathcal{E}} w_e\, u_{FH,e} + \sum_{s\epsilon\mathcal{S}} w_s \pi_{FH,s} + \sum_{c\epsilon\mathcal{C}} w_c\, \pi_{FH,c}, \tag{7.1}$$

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 centralisation (C-RAN) compared to the traditional RAN approach. However, this comes at the expense of overloading the transport network to support FH requirements, entailing reduced available resources for the BH functions. To address this issue, the secondary optimisation objective is to minimise the end-to-end delay in the BH, subject to demand processing and capacity constraints:

$$\min \mathcal{VBH}(\mathrm{u}, \pi) = \sum\nolimits_{\mathrm{e}\epsilon\mathcal{E}} \frac{1}{\mathcal{U}_{\mathrm{e}} - \mathrm{u}_{\mathrm{FH,e}} - \mathrm{u}_{\mathrm{BH,e}}} + \sum\nolimits_{\mathrm{s}\epsilon\mathcal{S}} \frac{1}{\Pi_s - \pi_{FH,s} - \pi_{BH,s}} + \sum\nolimits_{\mathrm{c}\epsilon\mathcal{C}} \frac{1}{\Pi_c - \pi_{FH,c} - \pi_{BH,c}} \qquad (7.2)$$

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 (7.1)-(7.2) can be expressed as $\min \mathcal{F}(\mathrm{u}, \pi) = [\mathcal{VFH}(\mathrm{u}, \pi),\ \ \mathcal{VBH}(\mathrm{u}, \pi)]$, subject to the aforementioned constraints. This problem can be transformed from an MOP problem into a single objective optimisation, using traditional scalarisation techniques (e.g. the Pascoletti-Serafini scalarisation method [12]), and it can be solved using relaxation schemes (e.g. Lagrangian Relaxation). In the section 7.1, the performance of the overall architecture is evaluated through the modelling framework described above. The joint FH and BH service provisioning problem is modelled through the MOP (7.1)-(7.2) and is evaluated in terms of power consumption and service delay, adopting a realistic network topology using a measured traffic dataset [17].

7.1 Performance Evaluation

The network topology assumed is the Bristol 5G city infrastructure, illustrated in Figure 18 (a) that will host the final proof-of-concept 5G-XHaul demonstrator. This infrastructure covers a 10x10 km2 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 fibre per link, 4 wavelengths per fibre, wavelength channels of 10 Gbps each, minimum bandwidth granularity of 100 Mbps and maximum link capacity of 40 Gbps. In the present study, w_{e} is associated with the power consumption of link $\mathrm{e}\epsilon\mathcal{E}$. The TSON power consumption model is provided in [9], [15]. The microwave transceivers considered, support maximum bandwidth of 2 Gbps and their typical power consumption is 45W (Huawei OptiX RTN310). Furthermore, a 2×2 MIMO scheme with adaptive number of transmission elements, carrier frequency 2.6 GHz, 20 MHz bandwidth adjustment and capacity per cell up to 201.6 Mbps has been considered, while users are distributed and generate traffic over the serviced area according to real datasets reported in [17] (Figure 18 (b)). This traffic, corresponding to wireless access traffic, needs to be processed by specific compute resources. The proposed optimisation scheme focuses 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 [2]) and BH CDN services.

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 18 (a)) with an average cost equal to 2 Watts/GOPS and 1.6 Watts/GOPS, respectively. Details regarding the numerical values used in the simulations are provided in [9]. 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 18 (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 on to increased BH service delays that remain below 25 ms (Figure 18 (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.

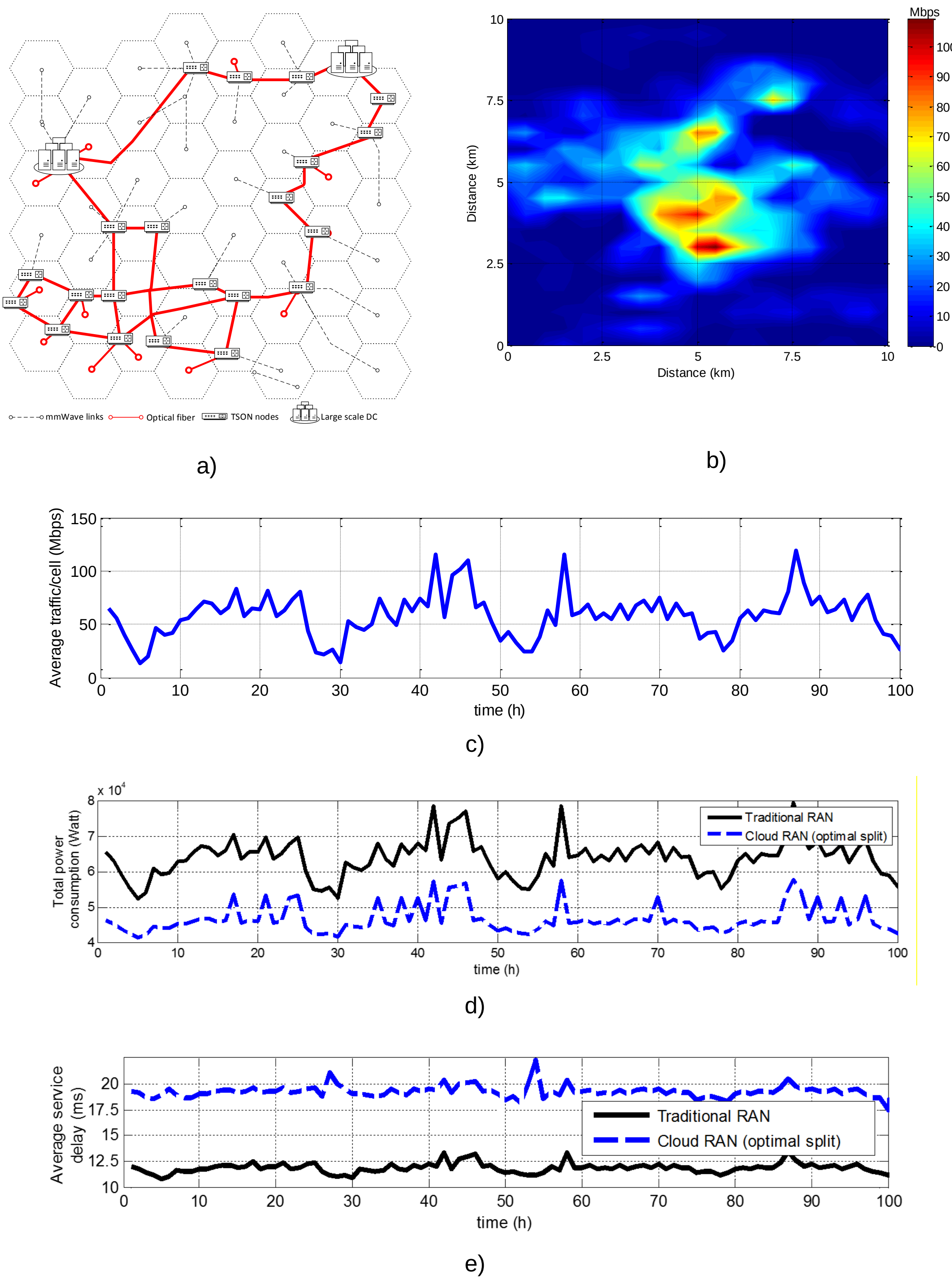


Figure 18: a) Bristol 5G city network topology with mmWave backhauling, b) Snapshot of spatial traffic load and c) average traffic/BS based on the dataset [10] during 8/2012, d)-e) Total power consumption and total service delay over time for the traditional RAN.

The impact of mobility on the total power consumption and the optimal split option adopted is shown in Figure 19 (a) and (b), respectively. The call-to-mobility factor is defined as the fraction of the service holding time over the cell residence time [15], 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 19 (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 centralisation benefits the system, due to the increased consolidation and the improved performance per Watt that large scale remote DCs offer, compared to local cloudlets.

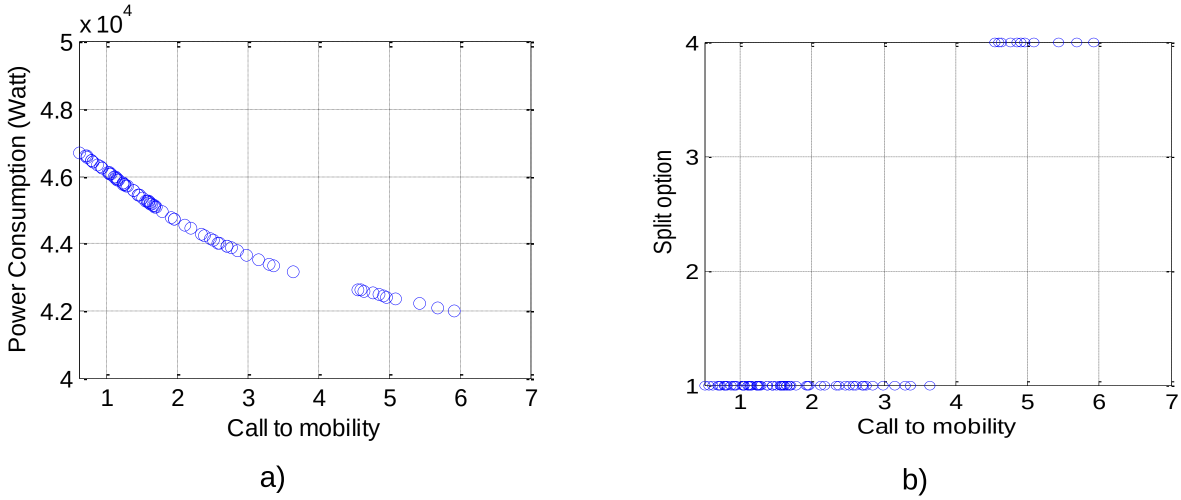


Figure 19: Impact of mobility power consumption and optimal split option (load 18 Mbps/cell).

Figure 20 (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 20 (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.

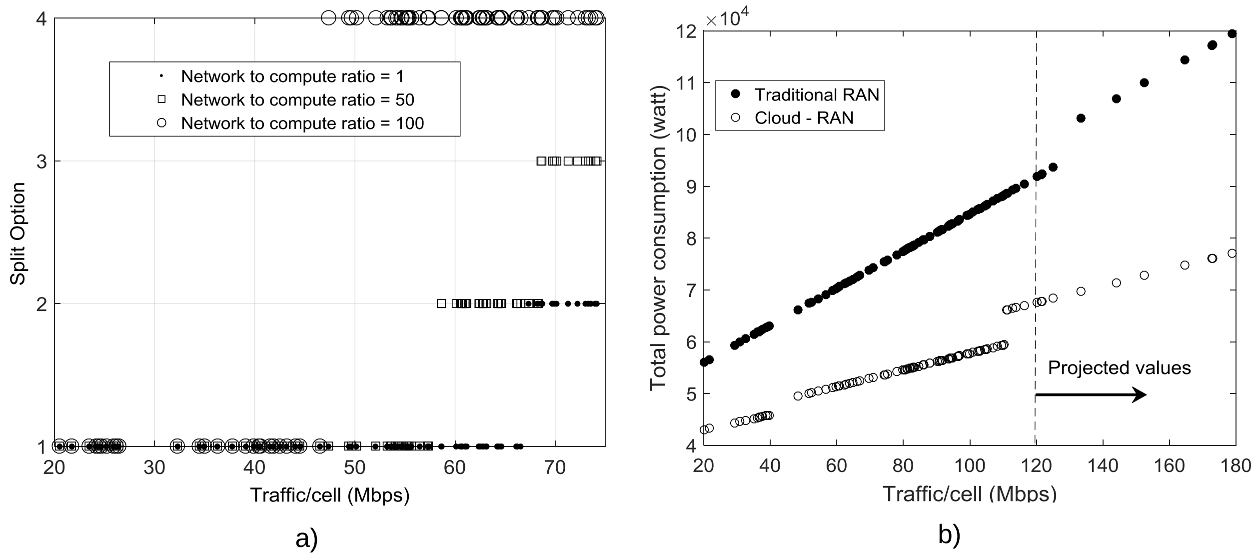


Figure 20: Split option as a function of load for different Compute to network ratios.

Finally, the impact of the relative processing and transport network cost, on the optimal split option, is illustrated in Figure 21. 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 21. 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.

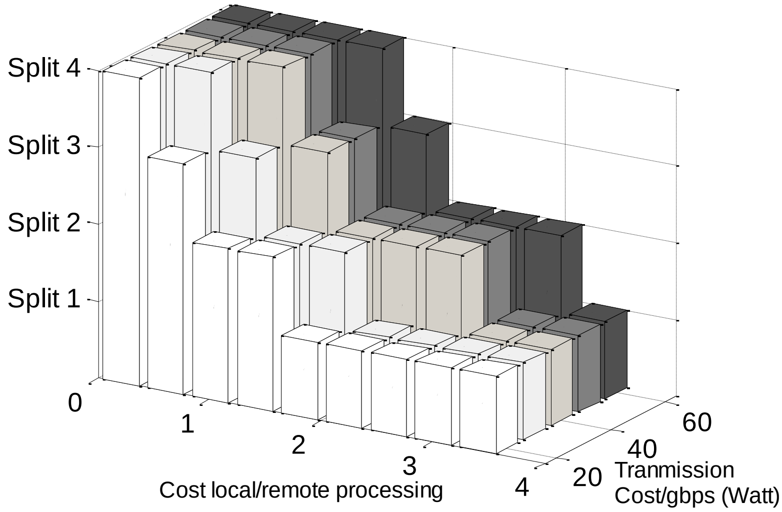


Figure 21: Split options for various processing and transmission costs.

7.2 Extensions: Optimal transport network deployment

5G-XHaul considers a multi-technology transport network solution comprising point-to-point microwave links and optical network technologies for the interconnection of the RUs with the BBUs. Optical technologies outperform wireless in terms of energy efficiency and capacity, offering at the same time deterministic performance as they are not affected by whether conditions, etc. However, they suffer significant deployment and installation costs which increase proportionally with the distance covered. On the other hand, microwave links can be easily installed almost everywhere offering significant benefits in terms of flexibility and upgradability.

Given the significant role that these technologies are expected to pay in future 5G systems, this subsection aims at identifying the optimal deployment strategies for both mmWave and optical fiber transport solutions. To keep the analysis tractable, the present study focuses on the part of the transport network interconnecting the RUs with the aggregation network (see Figure 22 a) and two basic options are evaluated. The first option considers the RUs to be interconnected to the metro/optical network using mmWave links while the second assumes the deployment of PONs. For the equipment and installation costs of both technologies (i.e. mmWave tower, optical equipment, fiber trenching costs, etc.), the values reported in [42] have been adopted. The associated curve trends, including installation costs for fiber and mmWave links as a function of distance, are shown in Figure 22 (b). It is seen that the fiber installation costs increase linearly with distance, whereas for microwave links these costs remain almost constant as they primarily depend on the initial tower set-up costs. Despite the initial high installation cost of optical technologies their daily operational costs are lower compared to mmWave due to their much lower power consumption. For example, for an end-user requesting 1.25 Gbps data rate the power consumption for PON ranges between 10 and 22 W/user whereas this may reach 180 W/user when the same services are provided over mmWave links [43].

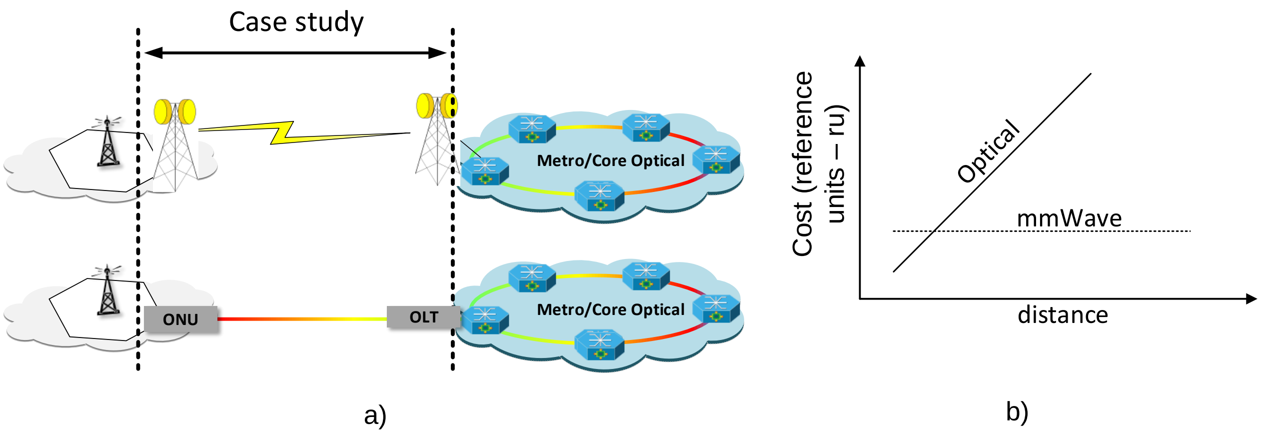


Figure 22: a) Transport network technologies domain. b) Microwave and fibre installation costs as a function of distance.

The set $\mathcal{E}$ of physical links introduced in (7.1)-(7.2) is partitioned into the set of links $\mathcal{E}_w$ in the mmWave domain, the set $\mathcal{E}_o$ of the passive optical links, and the set $\mathcal{E}_a$ of the active optical links, so that $\mathcal{E} = \mathcal{E}_w \cup \mathcal{E}_o \cup \mathcal{E}_a$. In this section, a third objective is introduced that aims at identifying the optimal mix of mmWave and PON technologies so that the total CAPEX and OPEX are minimised. To achieve this, a binary variable, $\wp_r$, is first introduced to indicate whether an RU, r, is interconnected to the metro optical network using PON or mmWave technologies:

$$\wp_r = \begin{cases} 0 & \text{if } r \text{ is interconnected to the metro network using PON} \\ 1 & \text{if } r \text{ is interconnected to the metro network using mmWave} \end{cases}$$

Assuming that CAPEX and OPEX for PON and mmWave network supporting RU r is $\mathcal{C}_r^{PON}$ and $\mathcal{C}_r^{mmWave}$, respectively, the third objective is to select $\wp_r$ so that the following cost can be minimised:

$$\min \sum_r [\mathcal{C}_r^{PON}(1 - \wp_r) + \mathcal{C}_r^{mmWave}(\wp_r)], \qquad (7.3)$$

subject to the constraints of the problem formulated in Section 7.

The impact of the share of the transport network technologies chosen on the total CAPEX and OPEX is illustrated in Figure 23. In the numerical results, it has been assumed that OPEX is associated with the power consumption and has been converted to monetary values by multiplying with 0.02r.u/ kWh. Note that the initial investment cost for installing PON and mmWave technologies is depreciated over a 10 year period. In Figure 23 it is shown that when RUs are fully backhauled through mmWave technologies the total CAPEX and OPEX is high due to the relatively high power consumption of mmWave links. On the other hand the increase of PON penetration reduces the total cost due to the energy efficient operation of PONs. However, exceeding a specific number of RUs backhauled by PON solutions, would make the total CAPEX and OPEX increase due to the significant fibre optic trenching costs.

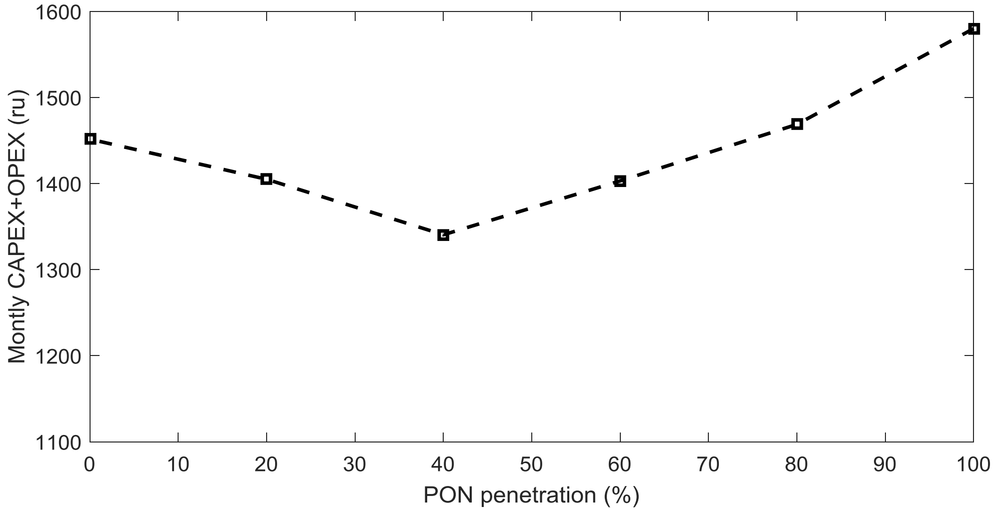


Figure 23: Impact of transport technologies and total CAPEX+OPEX.

8. Conclusions

A converged optical-wireless 5G network infrastructure has been proposed by the EU 5G-PPP project 5G-XHaul, to support jointly operational network (C-RAN) and end-user services. 5G-XHaul considers a layered architecture, inspired by the ETSI NFV standard and the SDN reference architecture.

The 5G-XHaul infrastructure interconnecting compute resources with fixed and mobile users, and defines an innovative heterogeneous network data plane adopting a wide variety of wireless and optical technologies. In view of this, this deliverable reports on the optical/wireless backhaul and fronthaul 5G-XHaul architecture. The concept of C-RAN is expanded to also address the option of flexible functional splits. In addition, a detailed evaluation of the performance of this architecture is provided through modelling and simulations studies.

A novel MOP modelling framework has been developed to evaluate the performance of the 5G-XHaul architecture. Our study has considered a variety of fronthaul and backhaul options, spanning from the distributed RAN to solutions adopting fully or partially the C-RAN approach. Modelling 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. In addition, optimal technology deployment choices have been investigated in these environments with regards to monetary capital and operational costs, comparing mmWave and PON solutions.

9. References

[1] J. Chen et al "Demonstration of Analog Millimeter-wave Fronthaul Link for 64-QAM LTE Signal Transmission" *in Proc. IEEE VTC*, Fall 2015

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

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

[4] 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.

[5] 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.

[6] 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.

[7] 5G PPP architecture WG, View on 5G Architecture, https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-5G-Architecture-WP-July-2016.pdf

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 5G Vision. The 5G Infrastructure Public Private Partnership: the next generation of communication networks and services. 2015. [Online] https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf

[19] GSMA Intelligence, Understanding 5G:Perspectives on future technological advancements in mobile, Dec.2015.
https://www.gsmaintelligence.com/research/?file=c88a32b3c59a11944a9c4e544fee7770&download

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

[21] Wireless Gbit Alliance (WiGig): http://wirelessgigabitalliance.org

[22] CHINA mobile White paper, C-RAN: The Road Towards Green RAN, [Online]: http://labs.chinamobile.com/cran/

[23] Huawei, Building 4G on High-Speed Railways, [Online]: http://www1.huawei.com/en/static/HW-371912.pdf.

[24] A. Tzanakaki et al., "5G infrastructures supporting end-user and operational services: The 5G-XHaul architectural perspective," 2016 IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, 2016, pp. 57-62.

[25] A. Tzanakaki, et al., "Virtualization of heterogeneous wireless-optical network and IT support of cloud and mobile services," IEEE Commun. Mag., 51(8), pp. 155-161, 2013.

[26] A. Sivaraman et al., Packet Transactions: High-Level Programming for Line-Rate Switches. ACM SIGCOMM 2016, pp. 15-28.

[27] G. Brebner, "Programmable hardware for high performance SDN," Optical Fiber Communications Conference and Exhibition (OFC), 2015, Los Angeles, CA, 2015, pp. 1-3.

[28] ETSI GS NFV 002 v1.1.1 Network Function Virtualisation (NFV); Architectural Framework, 2013-10.

[29] ETSI GS NFV-EVE 005 V1.1.1, Network Functions Virtualisation (NFV); Ecosystem; Report on SDN Usage in NFV Architectural Framework. 2015-12.

[30] H. Jin, D. Pan, J. Liu and N. Pissinou, "OpenFlow-Based Flow-Level Bandwidth Provisioning for CICQ Switches," in IEEE Transactions on Computers, vol. 62, no. 9, pp. 1799-1812, Sept. 2013.

[31] Y. Fang and I. Chlamtac, "Analytical Generalized Results for Handoff Probability in Wireless Networks," IEEE Trans. Commun., vol. 50, no. 3, Mar. 2002, pp. 369–99.

[32] 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.

[33] 5G-XHaul, Deliverable D2.1 – Requirements Specifications and KPIs Document, November 2015, [Online]. Available: http://www.5g-xhaul-project.eu/download/5G-XHaul_D_21.pdf. Accessed Dec. 5, 2016.

[34] 5G-XHaul, Deliverable D2.2 – System Architecture Definition, July 2016, [Online]. Available: http://www.5g-xhaul-project.eu/download/5G-XHaul_D_22.pdf. Accessed Dec. 5, 2016.

[35] 5G-XHaul, Deliverable D2.4 – Network Topology Definition, December 2016.

[36] 5G-XHaul, 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, June 2016, [Online]. Available: http://www.5g-xhaul-project.eu/download/5G-XHaul_D_31.pdf. Accessed Dec. 5, 2016.

[37] 3GPP, TR 38.913: Study on Scenarios and Requirements for Next generation Access Technologies, 3GPP Technical Report, 2016.

[38] Common Public Radio Interface, [Online]. Available http://www.cpri.info/. Accessed Aug. 1, 2016.

[39] IEEE Next Generation Fronthaul Interface (1914) Working Group, [Online]. Available: http://sites.ieee.org/sagroups-1914/. Accessed Oct. 1, 2016.

[40] NGMN Alliance, "Guidelines for LTE Backhaul Traffic Estimation," White paper, July 2011. [Online]. Available: http://www.ngmn.de/uploads/media/NGMN_Whitepaper_Guideline_for_LTE_Backhaul_Traffic_Estimation.pdf., Accessed Dec. 12, 2016.

[41] G. Wunder et al., "5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications," IEEE Communications Magazine, vol. 52, no. 2, pp. 97-105, February 2014.

[42] https://www.digitalairwireless.com/files/Fiber-vs-Microwave-White-Paper_1333235596.pdf

[43] F. Heliot, E. Katranaras, O.Onireti, M.Imran, "On the Energy Efficiency-Spectral Efficiency Trade-off in Cellular Systems", "Green Communications: Theoretical Fundamentals, Algorithms and Applications", Chapter 14, pp.353-404. 2012.

[44] J. Zou, "Optical Fronthauling for 5G Mobile: A Perspective of Passive Metro WDM Technology", OFC 2017.

[45] Xilinx, CPRI IP Core, https://www.xilinx.com/products/intellectual-property/do-di-cpri.html.

10. Acronyms

Acronym	Description
ADC	Analogue-to-Digital Converter
ADVA	ADVA Optical Networking
AP	Access Point
AxC	Antenna Carrier
3GPP	Third Generation Partnership Project
BB	Base Band
BBU	Base Band Unit
BVT	Bandwidth Variable Transponder
CAPEX	CAPital EXpenditures
CD	Content Delivery
CPRI	Common Public Radio Interface
C-RAN	Cloud Radio Access Network
CO	Central Office
CPRI	Common Public Radio Interface
CU	Central Unit
DC	Data Centre
eCPRI	Enhanced Common Public Radio Interface
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FH	Fronthaul
Gbps	Gigabit per second
GOPS	Giga OPS
IEEE	Institute of Electrical and Electronics Engineers
IML	Infrastructure Management Layer
IPS	Instructions Per Second
KPI	Key Performance Indicator
LTE	Long Term Evolution
MCC	mobile cloud computing
MCS	Modulation and Coding Scheme
MIMO	Multiple-Input Multiple-Output
mmWave	Millimetre Wave
MIPS	Million Instructions per second
MOP	Multi Objective oPtimization
NFV	Network Function Virtualization
NIC	Network Interface Controller

OLT	Optical Line Terminal
ONU	Optical Network Unit
OPEX	OPerational EXpenditures
OPS	Operations Per Second
OSS	Operations Support System
PI	Physical Infrastructure
PNF	Physical Network Function
PON	Passive Optical Network
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RN	Remote Node
RRH	Remote Radio Head
RU	Remote Unit
SC	Service Chaining
SDN	Software Defined Networking
Tbps	Terabit per second
TTI	Transmission Time Interval
TSON	Time Shared Optical Network
TUD	Technische Universität Dresden
UNIVBRIS-HPN	University of Bristol, High Performance Networks Group
vBH	virtual BH
vFH	virtual FH
VI	Virtual Infrastructure
VNF	Virtual Network Function
VOQ	Virtual Output Queuing
WDM	Wavelength Division Multiplexing