ΟΙΚΟΝΟΜΙΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΘΗΝΩΝ

ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS

Department of Informatics

(MSc) in INFORMATION SYSTEMS

MSc. Thesis

“Network Functions Virtualization (NFV) for Mobile Networks”

Andreas G. Papidas

Athens, June 2016

Supervisor: George Xylomenos

Contents

Abstract

The target of Network Function Virtualization (NFV) is to decouple network hardware and software, enable the development of network solutions by using multi-vendor hardware that can effectively cooperate with multi-vendor software platforms and replace current standard networking operations for fixed and mobile networks. It is hoped that CAPEX and OPEX savings can be achieved by using NFV, since virtual network appliances can run in a cloud-based environment, with hardware residing in totally geographically dispersed points of interest.

NFV derives from Server Virtualization, a technique that enables users to share the resources and functionalities of a server concurrently through Virtual Machines (VMs) emulating a standard server. In a similar way, NFV migrates the standard network functions to a virtual or cloud based architecture, leading to high flexibility.

NFV, IoT (Internet of Things), SDN (Software Defined Networking), SON (Self-Organized/Optimized Networks) and 5G are predicted to shape the future of telecommunication networks as many analysts, research centers and vendors claim. The role of NFV in this path is possibly the most important, since the vast traffic volume expected to be created through IoT connected devices and 5G networks cannot be efficiently and cost effectively supported with existing infrastructure deployments.

Moreover, NFV is tightly coupled with SDN, with the latter having less stringent requirements as far as real time processing is concerned. Their combination can lead to flexibility with the development of non-proprietary communication protocols and the automation of network functions. The main difference between the two is that NFV aims at the decoupling of software and hardware, while SDN aims at the separation of packets and interfaces from the network control plane.

In the first chapter of this thesis we present the background for network function virtualization, the key benefits that it can bring to the operators and telecommunication service providers as well as the correlation with software defined networking (SDN). Several standardization and architecture design approaches are described as well.

In the second chapter we present the evolution from 2G (GSM) to 3G (UMTS) and 4G (LTE) and describe the main functions of the core and LTE radio network, since its specific elements are very good candidates for virtualization. We focus on LTE networks, since virtual EPCs (LTE core network) as well as virtual eNodeBs (Radio Access Part) are already deployed and operators and vendors understand that there shall be a clear benefit from virtualizing these elements due to their flat IP based architecture, simplicity of interfaces and CAPEX and OPEX savings that virtualization can provide.

The third chapter describes the key drivers behind the virtualization of EPC as well as the basic architectures proposed by research and industry while the fourth chapter refers to the idea of

virtualization in the radio access part of mobile networks deployed through C-RAN. We present the main reasons that led to the need for C-RAN deployments including a CAPEX and OPEX analysis and the basic deployment architectures up to this point in time.

Περίληψη Διπλωματικής Εργασίας

Ο κύριος στόχος των τεχνικών εικονικοποίησης δικτυακών λειτουργιών (Network Functions Virtualization - NFV) είναι η ανεξαρτητοποίηση του λογισμικού από το υλικό, η δημιουργία λογισμικού που παρέχει τις ίδιες λειτουργίες που μέχρι στιγμής παρέχονται από συγκεκριμένες μονάδες δικτυακού εξοπλισμού και η ανάπτυξη λύσεων λογισμικού οι οποίες μπορούν να λειτουργούν πάνω σε υλικό οποιουδήποτε κατασκευαστή εγκατεστημένο σε διαφορετικές περιοχές. Η υπάρχουσα δομή του διαδικτύου αναμένεται να αλλάξει υιοθετώντας την συγκεκριμένη τεχνική ενώ αναμένονται ιδιαίτερα οικονομικά οφέλη σε επίπεδο μείωσης του κόστους δικτυακών εγκαταστάσεων.

Η τεχνική εικονικοποίησης δικτυακών λειτουργιών προέρχεται από τις ήδη υπάρχουσες τεχνικές εικονικοποίησης των διακομιστών (servers) που επιτρέπουν στους χρήστες να μοιράζονται τους πόρους του συστήματος μέσω εικονικών μηχανών (Virtual Machines) που αποτελούν εξομοιώσεις των διακομιστών. Με παρόμοιο τρόπο η τεχνική εικονικοποίησης δικτυακών λειτουργιών μετατρέπει τις υπάρχουσες λειτουργίες του δικτυακού υλικού σε εξομοιώσεις παρέχοντας με αυτόν τον τρόπο ευελιξία υλοποιήσεων.

Η **τεχνική εικονικοποίησης δικτυακών λειτουργιών (NFV) ,το 'Διαδίκτυο των Πραγμάτων (IoT)', η δικτύωση που βασίζεται μόνο σε λογισμικό (Software Defined Networking), τα αυτό-οργανωμένα / αυτό-βελτιστοποιημένα δίκτυα (Self-Organized/Optimized Networks) και τα δίκτυα πέμπτης γενιάς (5G)** είναι οι τεχνολογίες αιχμής που αναμένεται να επικρατήσουν τα επόμενα έτη όπως συνάγεται από μελέτες ερευνητικών κέντρων και κυρίως της βιομηχανίας τηλεπικοινωνιών. **Σε αυτό το σημείο αξίζει να σημειωθεί ότι τα δίκτυα δίκτυα πέμπτης γενιάς (5G) θα είναι βασισμένα σε τεχνικές εικονικοποίησης δικτυακών λειτουργιών (NFV).**

Η τεχνική εικονικοποίησης δικτυακών λειτουργιών (NFV) είναι συναφής με τις τεχνικές δικτύωσης που βασίζονται μόνο σε λογισμικό (Software Defined Networking - SDN) οι οποίες ωστόσο έχουν χαμηλότερες απαιτήσεις στην επεξεργασία λειτουργιών σε πραγματικό χρόνο. Ο συνδυασμός των δύο μπορεί να οδηγήσει στην ανάπτυξη ανοιχτών πρωτοκόλλων επικοινωνίας και στην αυτοματοποίηση λειτουργιών του δικτύου. Η διαφορά των δύο έγκειται στο ότι το NFV στοχεύει στην ανεξαρτητοποίηση λογισμικού και υλικού ενώ το SDN στον διαχωρισμό πακέτων και διεπαφών από το επίπεδο διαχείρισης δικτύου.

Στο πρώτο κεφάλαιο παρουσιάζονται οι βασικές αρχές της εικονικοποίησης δικτύων (NFV) και οι βασικές αρχιτεκτονικές. Στο δεύτερο κεφάλαιο παρουσιάζεται η δομή του δικτύου τέταρτης γενιάς καθώς τα παραδείγματα εικονικοποίησης στα οποία θα αναφερθούμε αφορούν την συγκεκριμένη τεχνολογία. Στο τρίτο κεφάλαιο περιγράφουμε λύσεις εικονικοποίησης του δικτύου κορμού του δικτύου τέταρτης γενιάς (4G) ενώ στο τελευταίο κεφάλαιο περιγράφονται οι τεχνολογίες εικονικοποίησης του δικτύου πρόσβασης.

Introduction

Until now, network equipment in fixed and mobile communications evolved on the rationale of standard vendor defined software running on standard hardware. This fact sets constraints on service providers and operators, since the dependency between network hardware and software cannot be broken. Moreover, as networks expand and new services need to be deployed, operators and telecommunication service providers need to install new proprietary hardware appliances, reserve space and power resources, leading to excessive CAPEX and OPEX losses, while the management of the hardware platforms becomes more and more complicated.

The target of **Network Function Virtualization (NFV)** is to decouple the latter dependency between hardware and software, enable the development of network solutions by using multi-vendor hardware that can effectively cooperate with multi-vendor software platforms, and replace current standard networking operations for fixed and mobile networks. It is expected that CAPEX and OPEX savings can be increased by using NFV as a solution, since NFVs can run in cloud based hardware in totally geographically dispersed points of interest. The current global Internet architecture is expected to change gradually, since NFV shall have an important impact on the traditional network structures as well.

NFV derives from Server Virtualization, a technique that enables users to share the resources and functionalities of a server concurrently through Virtual Machines (VMs) which are emulations of a standard server. In a similar way, NFV migrates the standard network functions to a virtual or cloud based architecture leading to high flexibility. At this point we should stress the relation of NFV with cloud computing and Software Defined Networking (SDN) which shall be explained later in more detail.

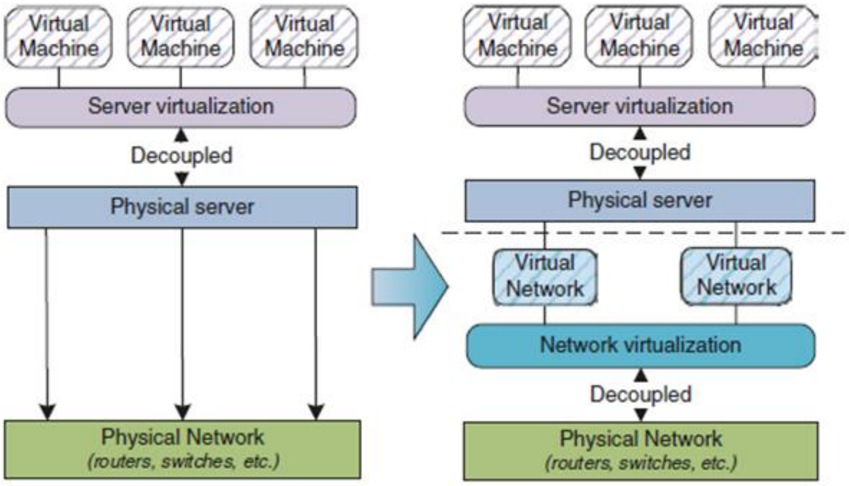


Figure 1. Server virtualization versus network virtualization. From [1].

NFV, IoT (Internet of Things) SDN (Software Defined Networking), SON (Self-Organized/Optimized Networks) and 5G are predicted to shape the future of telecommunication networks as many analysts, research centers and vendors claim, while NFV role in this path is possibly the most important since the vast traffic volume expected to be created through IoT connected devices and 5G networks cannot be efficiently and cost effectively supported with the existing standard infrastructure deployments policy. [42].

Through network virtualization, multiple virtual networks can share the same network element hardware resources, thus leading to a split of network services from the hardware. Based on this fact, multiple services can be reused. An example might be the efficient management of server network interconnection or even the virtualization of wireless networks functionalities including both core and radio access nodes.

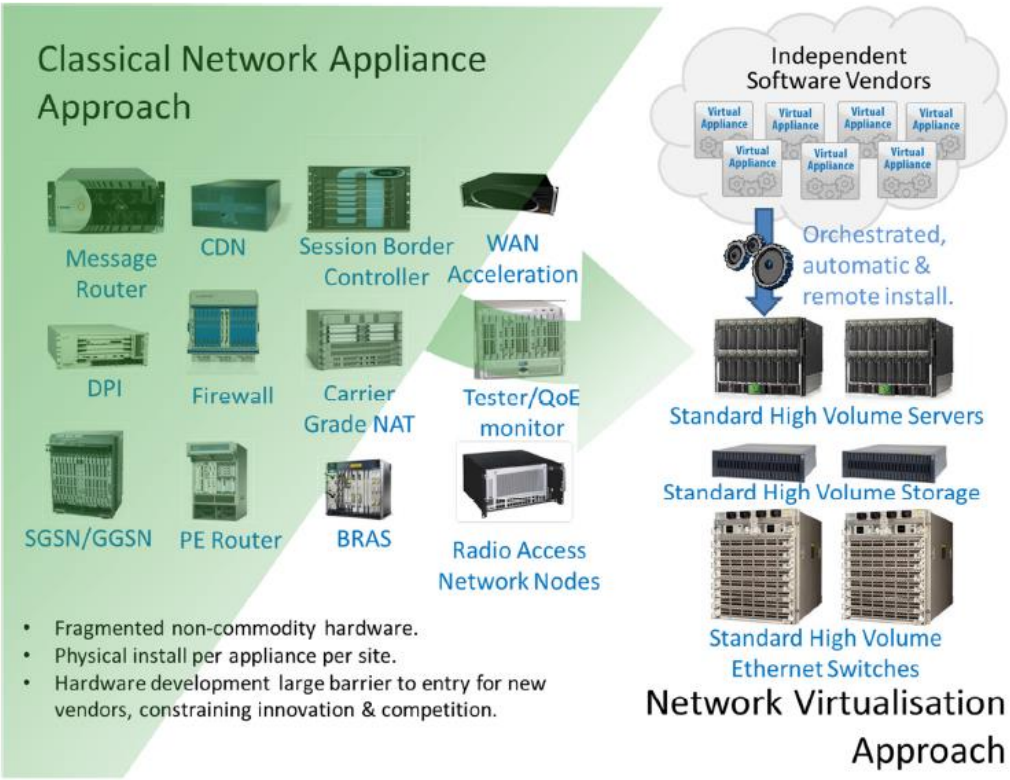


Figure 2. Classical network appliance approach versus NFV approach.

NFV aims to replace standard switching, routing and load balancing functionalities in the fixed domain, while for the mobile domain end to end solutions including the IMS (IP Multimedia System), LTE 4G EPC (Evolved Packet Core) and the RAN (Radio Access Network) are proposed [5]. EPC's all - IP flat architecture and simplicity of interfaces is a key enabler for the deployment of NFV solutions, while world leading vendors have a special interest on virtual EPC development through virtualizing LTE core network elements. Through such virtualization, the EPC can be offered as a service (EPCaaS) instead of standard unified hardware solution, higher interoperability can be achieved and more services can be deployed, while the CAPEX and OPEX savings are clear.

As already mentioned, virtualization can be applied in the radio access part as well through the evolutionary C-RAN approach. The basic functional radio access network elements of a base station (GSM/UMTS/LTE) are the baseband unit (BBU) and the radio unit (RRU or RRH). C-RAN aims to virtualize the baseband units of the base stations and aggregate them in a baseband unit pool serving multiple RRUs. Low latency, high bandwidth and effective resource allocation can be achieved by using advanced fronthaul optical interconnections between the BBU pool and the RRUs.

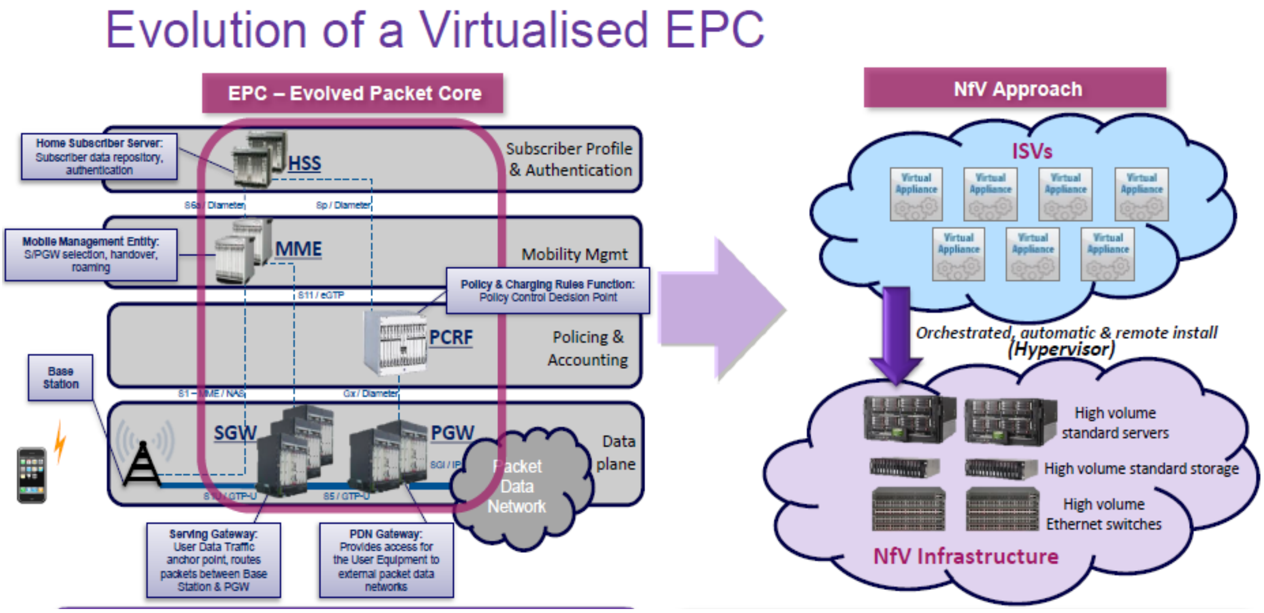


Figure 3. Transition from standard EPC to Virtualized EPC (vEPC).

The NFV approach can be applied to all layers of the standard OSI (Open systems Interconnection) model while hybrid approaches are feasible and possibly more preferable at this time, till virtualization deployments become mature.

Various use cases related with NFV exist, however in this thesis we shall focus on the mobile network field and more specifically we shall refer to the virtualization of LTE including both core network, through the virtualization of Evolved Packet Core (EPC), but Radio Access Network (RAN) as well, via the C-RAN (Centralized RAN) approach which is regarded as the next big thing after LTE in mobile communication networks.

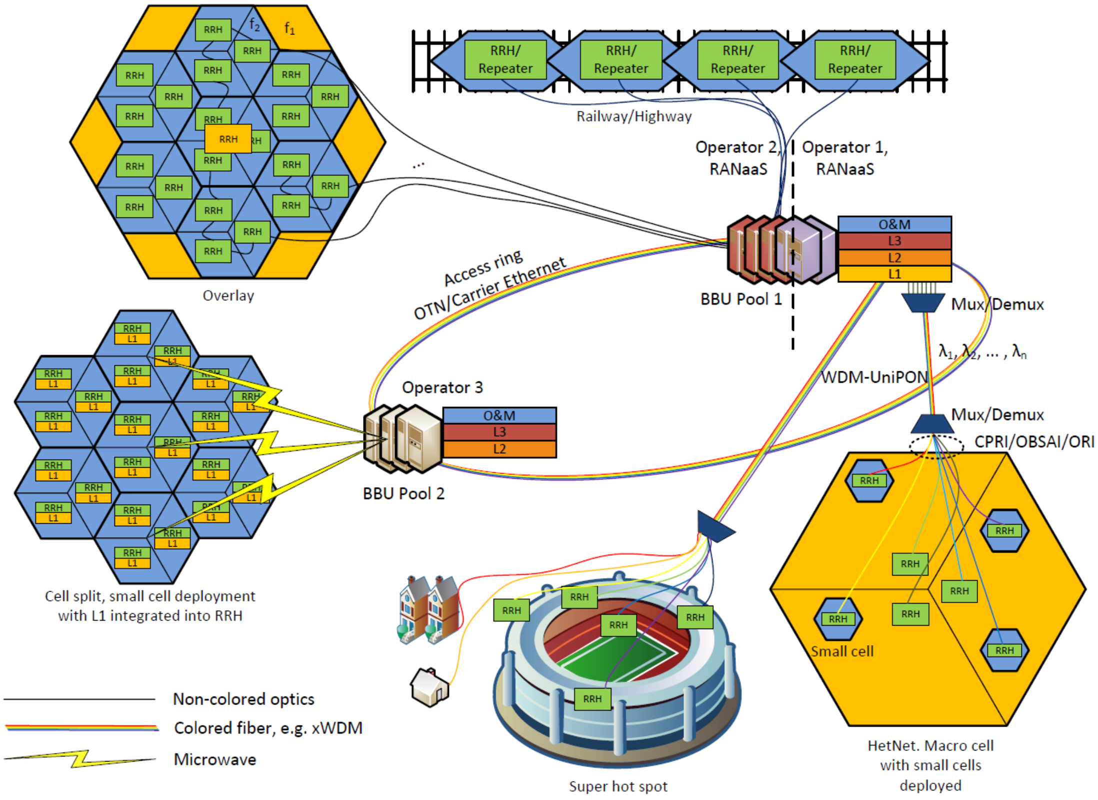


Figure 4. C-RAN deployment scenarios.

In the first chapter we present the background for NFV, the key benefits that the technology can bring to the operators and telecommunication service providers as well as the correlation with software defined networking (SDN). In addition several standardization and architecture design approaches are described.

In the second chapter we present the evolution from 2G (GSM) to 3G (UMTS) and 4G (LTE) and describe the main functions of core and radio network of LTE since the specific elements consist very good candidates for virtualization. We focus on LTE networks since virtual EPCs (LTE core network) and virtual eNodeBs (Radio Access Part) are already deployed from industry, while operators and vendors understand that there shall be a clear benefit from virtualizing the specific elements due to their flat IP based architecture, simplicity of interfaces and CAPEX and OPEX savings that virtualization can provide.

Chapter 3 describes the key drivers behind the virtualization of EPC as well as the basic architectures proposed by research and industry, while Chapter 4 refers to the idea of virtualization in the radio access part of mobile networks deployed through C-RAN. We present the main reasons that led to the need for C-RAN deployments including a CAPEX and OPEX analysis and the basic deployment architectures up to this point in time.

Chapter 1: Network Function Virtualization principles and design approaches

In the first chapter of this thesis we present the background for network function virtualization, the key benefits that can bring to the operators and telecommunication service providers, as well as its correlation with software defined networking (SDN). We proceed by presenting the first standardization and architecture design approaches delivered by organizations such as the European Telecommunications Standards Institute (ETSI), the design considerations that must be examined, the basic modules / entities that must be developed and finally we describe the research and vendors approach and activities on the topic.

1.1 The roots of NFV, key benefits and relation with Software Defined Networks (SDN)

History and roots of NFV

The first approaches and design proposals in network virtualization derive from industry as well as academic research centers, and in this thesis we refer to both perspectives. The next figure describes these two approaches and the possible application fields.

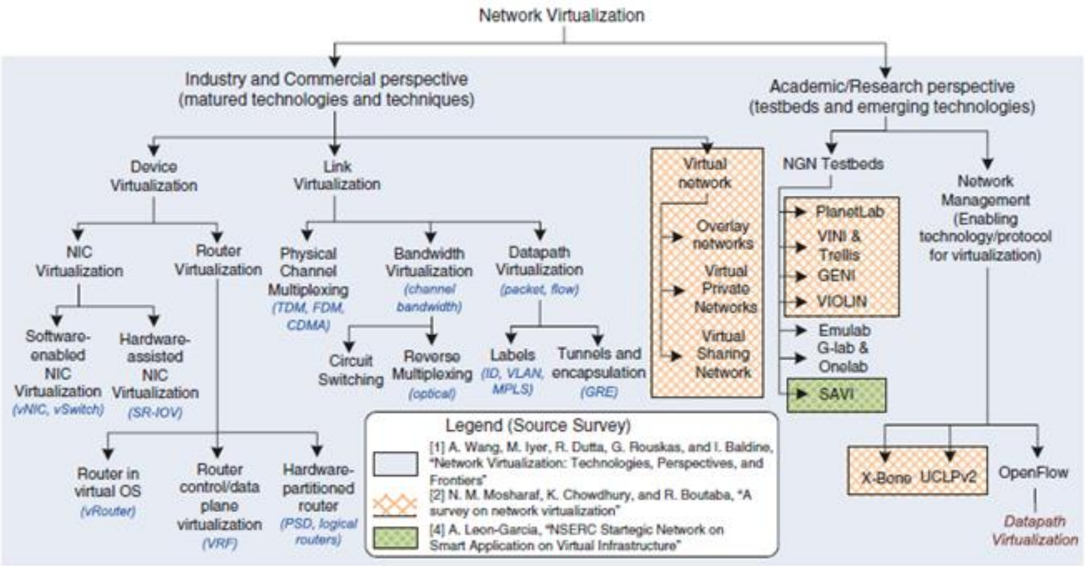


Figure 5. Industry and Academic perspective approach to NFV. From [2].

The idea of NFV was introduced during 2012 after the worldwide leading operators called research centers and vendors to set the roots and basic architecture of NFV solutions. ETSI (European Telecommunication Standards Institute) was selected to be the head of all specifications related to NFV and due to this the ETSI ISG NFV (ETSI Industry Specification Group) was created, leading to 11 group specifications releases. It is expected that 3GPP shall adopt the specific standards and proceed with the release of 3GPP standards as well.

Key benefits to telecommunication service operators through NFV

NFV solutions can bring to telecommunication service operators many benefits including the flexibility to evolve and deploy new network services in a faster and more cost effective manner under the following key points [2]:

- **Hardware and software decoupling**: As already mentioned, with NFV deployment a network element does not consist anymore of a combination of hardware and software network modules, thus separate planning and maintenance activities can be planned.
- **Flexibility in network functions operations and design**: The software and hardware decoupling leads to the redesign of hardware resources and elements and their usage for multiple concurrent network operations. Due to this, operators can deploy network services faster for their clients running on the same hardware entities.
- **Effective network scaling**: The dynamic scaling capabilities that are provided with NFV enable the usage of NFV instances with different granularity and per traffic application scenario, which right now is one of the key issues to be faced from operators since traffic continues increasing but hardware and software appliances remain the same.

At this point it should be noted that the decoupling between software and hardware does not mean obligatory resource virtualization of all network elements. Operators can still develop or buy software and run it on the current commodity hardware structures, however the gain of running the software in virtualized modules is the key point that leads to better performance results and CAPEX/OPEX profits. Finally, hybrid scenarios where functions running on virtualized resources can mutually operate and coexist with functions running on standard physical resources are suggested, till a full transition to virtualization takes place.

NFV and Software Defined Networking (SDN) relation

NFV is tightly coupled with SDN (Software Defined Networks), a recent approach to networking which leads to independence between network functions control and data forwarding in the traditional manner that we know. For example, right now a packet that arrives in a standard network element such as a router or switch must follow a set of forwarding or routing rules related to error correction, NAT (network address translation), QoS (Quality of Service) as well as standard routing protocols rules [2]. SDN and NFV have started being developed independently, however it can be said that the former acts as complementary to the latter.

With the use of SDN the above mentioned operations of data forwarding, routing and network function control are decoupled. To put it in a simpler way, data plane and control plane are decoupled, leading to centralized management of control plane and de-centralized data plane management.

SDN focuses on Layer 2 and 3 network elements and operations while a SDN controller provides the northbound interface on which many additional services can be built as shown below:

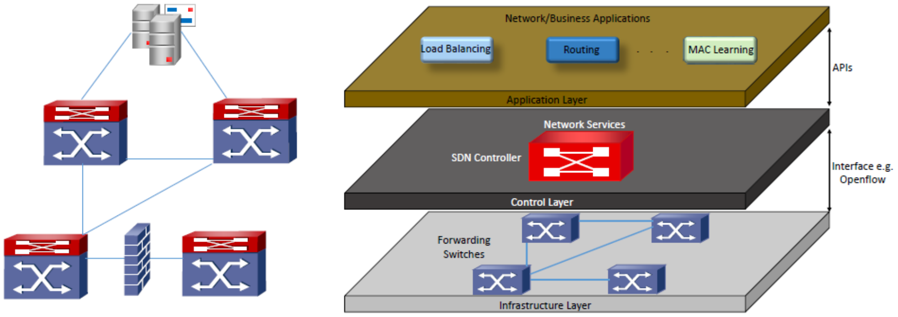


Figure 6. Traditional network layers versus logical layers in a SDN Network. From [2].

The key point in the relation of SDN and NFV is to understand that many IT services already run on cloud services, however for the NFV case and based on the fact that we focus on the telecommunications market where real time performance requirements are more stringent, a detailed mapping of current cloud models to telecom models is needed.

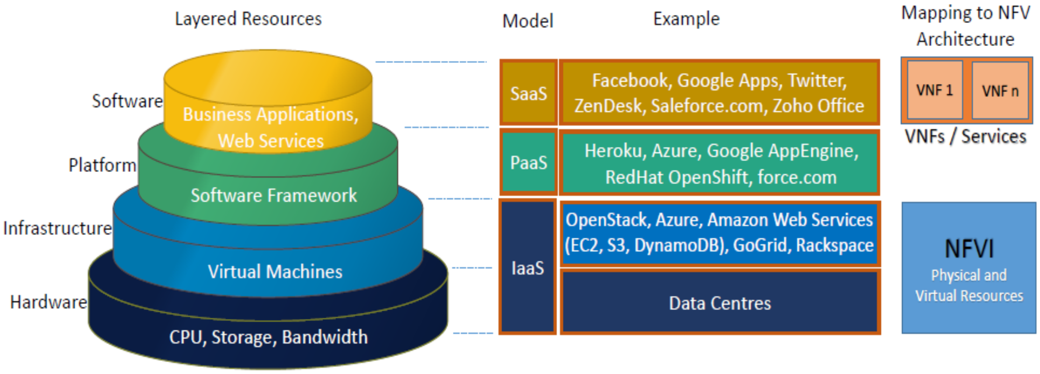


Figure 7. Indicative SDN architecture and mapping to NFV. From [2].

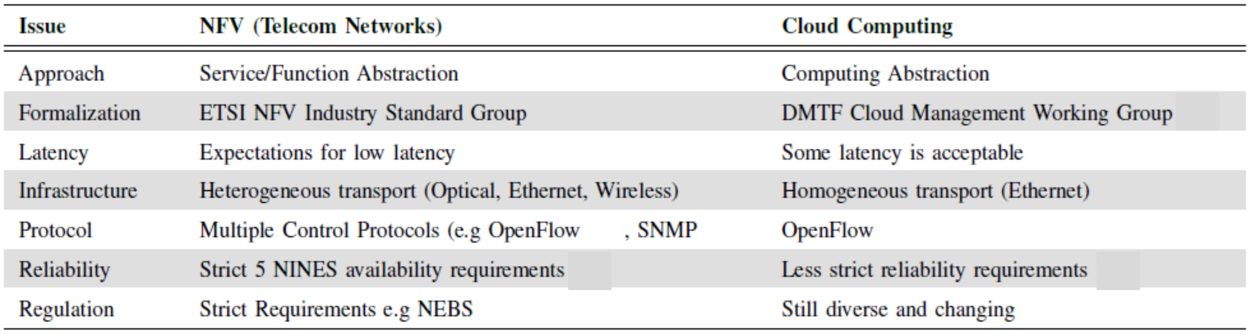
Until now, research activities of vendors as well as some research papers conclude that *Openstack*, a cloud operating system that can control large computing deployments, presents some deficiencies in the resource reservation policy and in the Quality of Service (QoS) demands, thus proposals that include hardware acceleration are already adopted for the needs of telecommunications market application scenarios [3].

It can be said that there are a lot of similarities between SDN and NFV with the clearest being the effort to use open source software and standard network infrastructure with non-proprietary protocols. Their combination can lead to even better results, since SDN can chain several network functions in a NFV deployment and provide further automation of functions.

The main difference between the two is that NFV aims at the decoupling of software with hardware while SDN aims at the separation of packets and interfaces from the network control plane. In more detail:

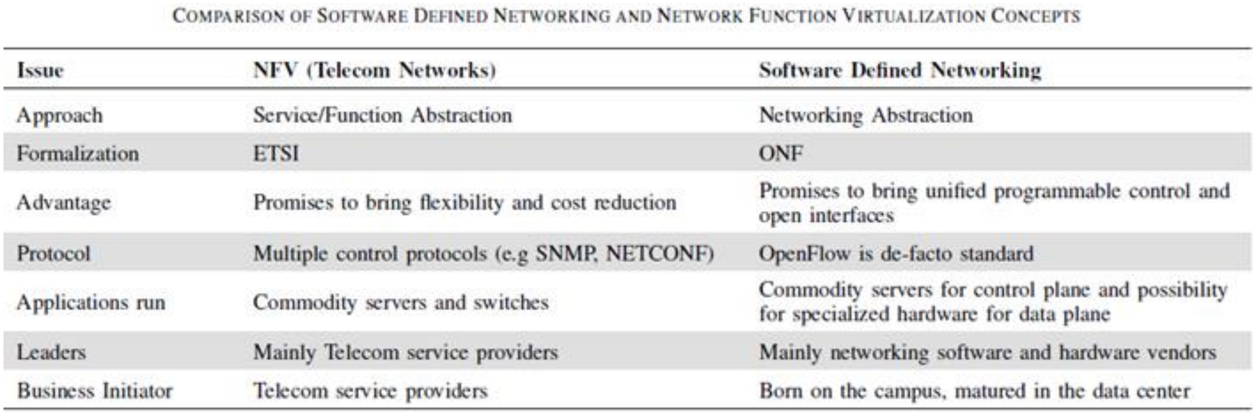
“The NFV concept differs from the virtualization concept as used in the SDN architecture. In the SDN architecture, virtualization is defined as the allocation of abstract resources to specific clients or applications. On the other hand, in NFV the goal is to abstract network functions from dedicated hardware, for example to allow them to be hosted on server platforms in cloud data centers” [4].

Some additional key differences between SDN, Cloud Computing and NFV are described by the following tables as well.



Issue	NFV (Telecom Networks)	Cloud Computing
Approach	Service/Function Abstraction	Computing Abstraction
Formalization	ETSI NFV Industry Standard Group	DMTF Cloud Management Working Group
Latency	Expectations for low latency	Some latency is acceptable
Infrastructure	Heterogeneous transport (Optical, Ethernet, Wireless)	Homogeneous transport (Ethernet)
Protocol	Multiple Control Protocols (e.g OpenFlow , SNMP	OpenFlow
Reliability	Strict 5 NINES availability requirements	Less strict reliability requirements
Regulation	Strict Requirements e.g NEBS	Still diverse and changing

Table 1. Key points comparison of NFV and Cloud Computing. From [2].



COMPARISON OF SOFTWARE DEFINED NETWORKING AND NETWORK FUNCTION VIRTUALIZATION CONCEPTS

Issue	NFV (Telecom Networks)	Software Defined Networking
Approach	Service/Function Abstraction	Networking Abstraction
Formalization	ETSI	ONF
Advantage	Promises to bring flexibility and cost reduction	Promises to bring unified programmable control and open interfaces
Protocol	Multiple control protocols (e.g SNMP, NETCONF)	OpenFlow is de-facto standard
Applications run	Commodity servers and switches	Commodity servers for control plane and possibility for specialized hardware for data plane
Leaders	Mainly Telecom service providers	Mainly networking software and hardware vendors
Business Initiator	Telecom service providers	Born on the campus, matured in the data center

Table 2. Key points comparison of NFV and SDN. From [2].

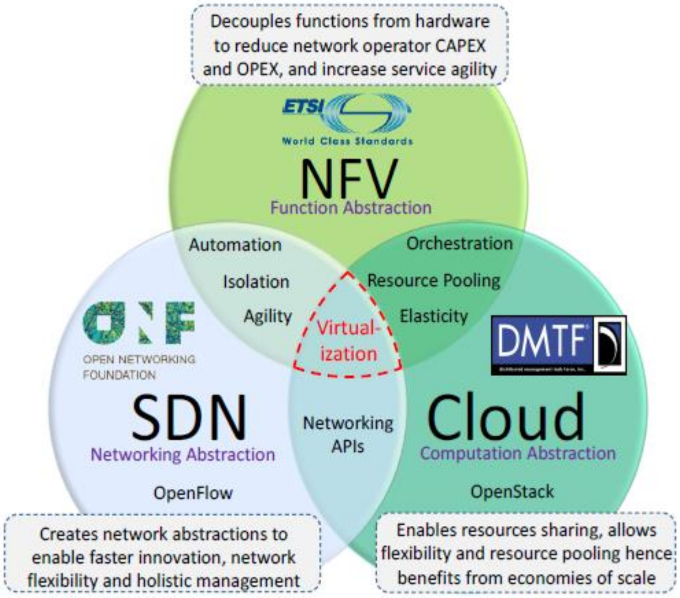


Figure 8. Relationship of NFV, SDN and Cloud Computing. From [2].

1.2 NFV architectures and key modules

In a high level design approach, NFV is based on the following domains [7]:

1. A software implementation of a network function (Virtualized Network Function) that is able to run over a NFV infrastructure.
2. The NFV Infrastructure (NFVI), consisting of the hardware resources where virtual networks can reside on and run.
3. A NFV Management and Orchestration module (MANO) which is assigned to handle and orchestrate the lifecycle of virtual network functions.

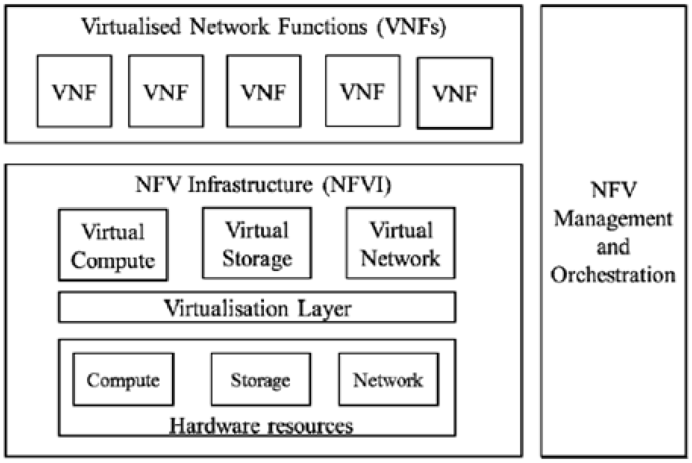


Figure 9. High Level NFV Domains. From [7].

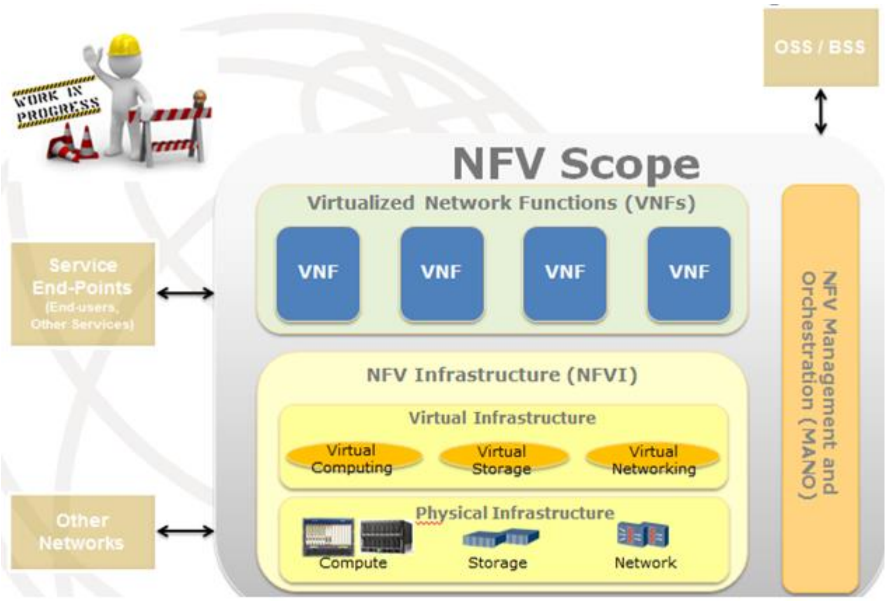


Figure 10. High Level architecture of NFV.

Additionally, a VNF Forwarding Graph (VNF-FG) defines and controls the connectivity between the VNFs while a description of an end-to-end network service (e.g. mobile voice / data, Internet access) consists of VNFs and end points as shown below, with the dotted lines representing the logical interfaces created between the NFVI and the NFVs which can be either wired or wireless.

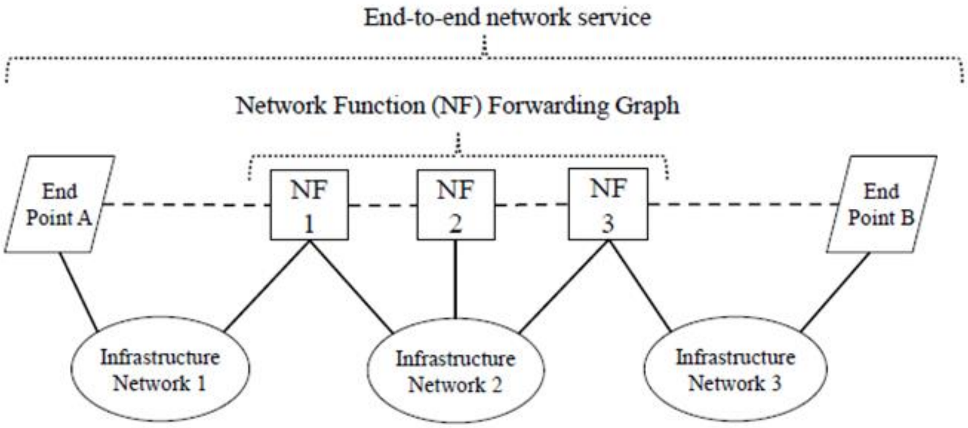


Figure 11. End to end service Graph representation. From [7].

NFV is based on the idea that the physical infrastructure deployment of a VNF is not visible from the end to end service perspective while redundant infrastructures can reside in different locations. Based on this a VNF can run on totally different geographically dispersed physical appliances.

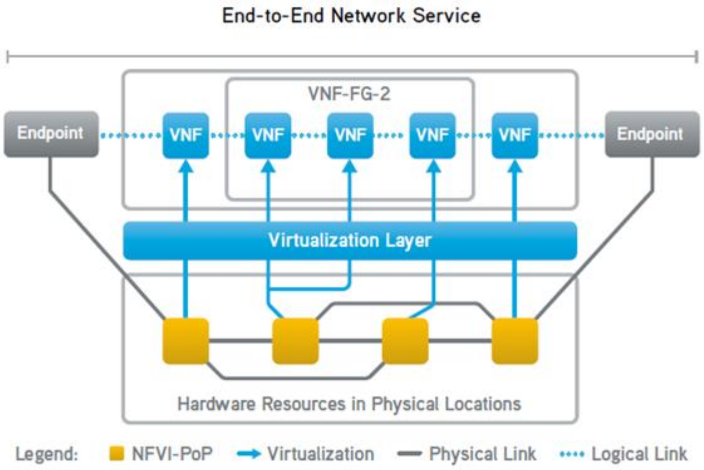


Figure 12. End to end network service with VNFs and nested graphs. From [7].

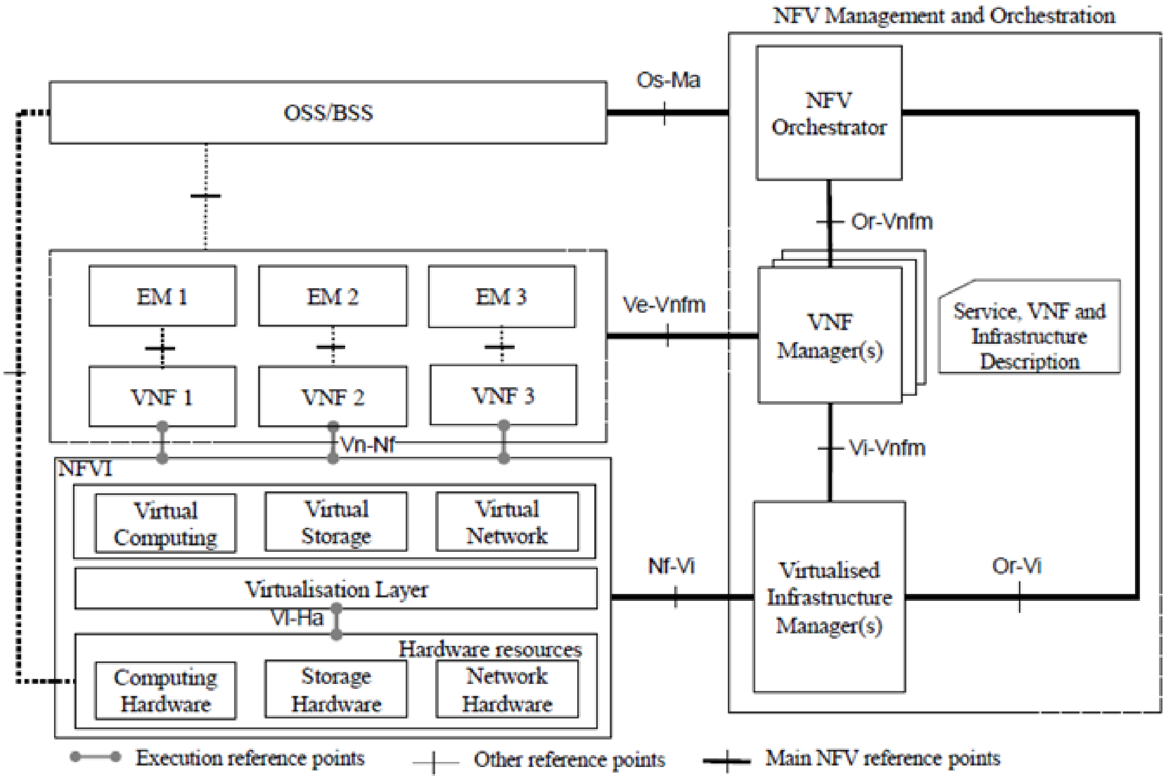


Figure 13. Architectural description of NFV. From [7].

The basic architectural functional blocks of NFV are the following [7]:

- Virtualized Network Function (VNF): The Virtualized Network Function refers to a standard network function, Non-virtualized prior to NFV introduction, which can be fully or partially virtualized. Typical examples might be the core network elements of LTE - EPC (Evolved Packet Core) such as the MME (Mobility Management Entity), the SGW

(Serving Gateway), or even the eNodeB through the C-RAN virtualization that shall be discussed in chapter 4 of this thesis.

- Element Management (EM): Refers to the management operations strategy of the VNFs.
- NFV Infrastructure (NFVI): VNFI refers to the common hardware resources (servers, storage) that host the NVFs and can be physically located in various data centers across a city or a country supporting a pool of resources for the NFVs. Routers, switches and wireless links interconnecting the main servers can be regarded as part of the NFV Infrastructure as well.

An additional set of routers, switches and wireless or wired links can operate as a NFVI interconnection point of the NFV network with other external networks that might be virtualized or not.

The virtualization layer has the role of abstracting and decoupling the virtualized network functions from the hardware resources under the aspect that each VNF shall use dedicated hardware resources so that the highest performance characteristics are assured, while it is also assured that the software running the NFV can at the same time run on multiple hardware resources. Through the latter virtual access to the underlying compute resources is ensured while standard actions like starting, stopping or migrating VMs can be deployed.

- Hypervisor software is able to manage several guest operating systems and enable consolidation of physical servers onto a virtual stack on a single server. CPU, RAM, and storage are flexibly allocated to each VM via software deployments [5].
- Virtualized infrastructure managers: VNFI managers provide resource management for the hypervisors, allocate resources to NVFs and provide fault management capability for the NFVI.
- NFV Manager: Provides the NFV lifecycle management and either multiple VNF managers operating per VNF, or one handling all VNFs can be deployed.
- NFV Orchestrator: Provides the orchestration and management of NFVI and combined with the VNF manager and VNF Infrastructure manager consist the NFV Management and Orchestration (MANO) layer. This layer provides connectivity and interacts with other virtualized networks or standard network infrastructures. Open stack protocols and Software Defined Networking (SDN) functions can be deployed on the orchestrator layer while it is the connection point to the Operations Support System and Business Support System (OSS / BSS) of a service provider or operator.

The following example depicts the role of the Orchestration and Management layer (MANO) for a possible scenario where a Service Provider provides NFVs to two different enterprises and an operator. Through MANO, different policies can be applied for each customer through the orchestration layer leading to high flexibility and common use of hardware resources [8].

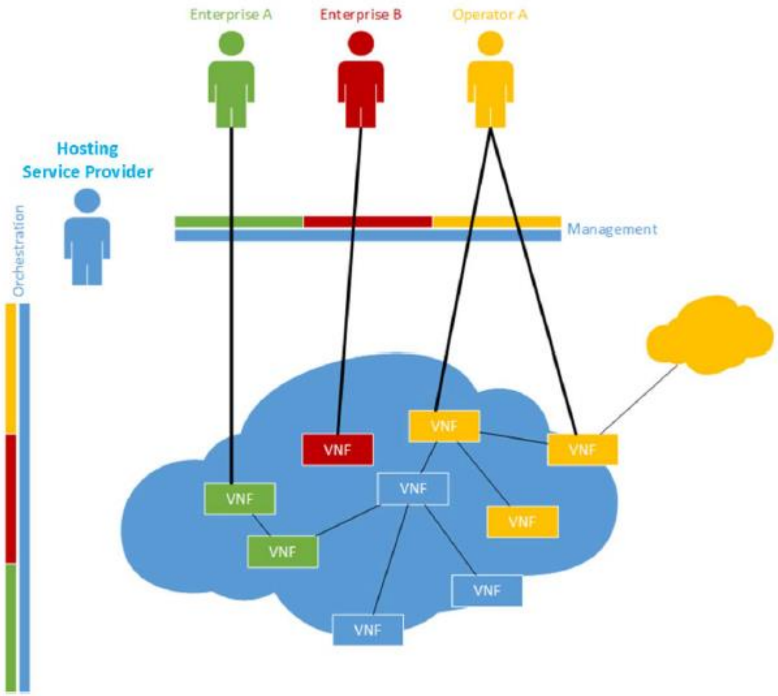


Figure 14. Sharing of the same infrastructure through NFV. From [8].

The orchestrator layer provides the parallel management of NFVI (Network Functions Virtualization Infrastructure) and NFVs, while the concurrent handling of the latter leads to the creation of virtual network services that can collaborate through NFV forwarding graphs (NFV-FG) and virtual switches (Vswitches) [5], [8].

The services can be further optimized per case (according to different traffic requirements scenarios) while NFV-FGs can provide high scalability since only graph update is needed in a Vswitch after a new VM is created [5], [8].

1.3 NFV Use cases and deployment scenarios

Several Use cases have been defined by ETSI, and as NFV scenarios have started being commercially deployed more and more cases might emerge. The following are the initial use cases examined by ETSI [8]:

Use Case 1 - Network Function Virtualization Infrastructure as a Service: The specific use case provides a solution for mapping the Cloud Computing Service Models IaaS and NaaS as Network Function Virtualization Infrastructure elements.

Use Case 2 - Virtual Network Function as a Service (VNFaaS): Application of virtualization to the enterprise so that the operator provides services and the enterprise consumes predefined resources.

Use Case 3 - Virtual Network Platform as a Service (VNPaaS): Addition to case 2 in the aspect that the enterprise can host VNFs instances on its own.

Use Case 4 - VNF Forwarding Graphs: Definition of packet sequence routing.

Use Case 5 - Virtualization of Mobile Core Network (EPC) and IMS: This use case describes the virtualization design concerns of the of the mobile packet core and IMS network elements and services.

Use Case 6 - Virtualization of the Mobile Base Station: Description of virtualization of the mobile RAN through standard IT servers.

Use Case 7 - Virtualization of the Home Environment: Description of CPE virtualization.

Use Case 8 - Virtualization of CDNs (vCDN): Description of virtualization for content delivery networks (CDNs).

Use Case 9 - Fixed Access Network Functions Virtualization: Description of virtualization of fixed network access infrastructure (an example might regard DSL cabinets) and description of co-location with wireless access nodes.

In this thesis we shall focus on use cases 5 and 6, and more specifically we shall refer to the virtualization solutions for the core network of LTE (EPC) as well as the virtualization of LTE Radio Access Network (E-UTRAN) and Mobile Base Station (eNodeB) through the C-RAN approach.

1.4 NFV Design Considerations

In order for NFV solutions to be acceptable to operators and keep the same quality constraints for the end network users with the existing architectures, the following design considerations must be taken into account [2]:

1. **Performance:** The performance of network services must be the same as with existing network services running on dedicated hardware, not leading to bottlenecks at all layers and keeping low latency characteristics. As an example, consider a scenario where NFVs providing the same service reside in different VMs, then the interconnection of the latter must ensure high bandwidth and low latency.
2. **Security:** Current security policies applied to network services must be able to operate at the same way in NFVs. The most important point as far as security is concerned is that the NFVI must be protected from the services delivered to the end users through firewalls included in the NFV solution architecture.
3. **Availability and reliability, Disaster recovery compliance:** It must be ensured that outages are within the same timeframes described by today’s SLAs (Service Level Agreements), while in case of a failover there must be a redundancy solution.

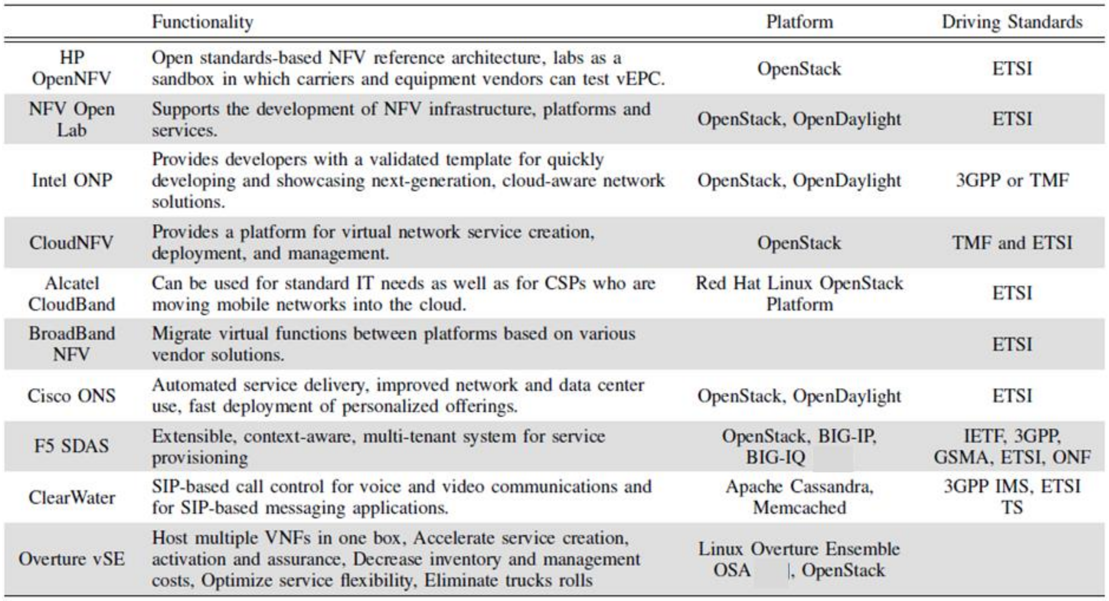
4. **Heterogeneous Support:** Currently operators have the option of sharing network elements and selecting among different vendors since all platforms can communicate through standardized interfaces. The same rationale must be able to be deployed with NFV through keeping open interfaces and ensuring interoperability of multiple vendors.
5. **Legacy systems support:** It is clear that the transition to NFV is not yet mature, since right now operators are still evaluating the solutions and very few commercial deployments exist. Due to this, it is expected that during the period hybrid NFV solutions being able to support current network architectures with legacy hardware and software systems shall prevail till full virtualization becomes a reality.

1.5 Vendor Specific NFV Implementations

A variety of NFV implementations have already been deployed by multiple vendors mainly in telecommunications field. Some examples are listed below:

- **CISCO Open network strategy:** This platform provides a solution to MANO NFV layer while the NFV manager deployed by CISCO is able to support third party NFVs as well. Additionally; CISCO developed a virtual EPC (LTE core network) solution which was recently deployed in NTT DOCOMO, largest mobile network provider in Japan and announced on 11th of March 2016 [44].
- **HUAWEI NFV Open Lab:** HUAWEI launched an open NFV lab aiming to test various deployment scenarios while prior to CISCO HUAWEI had deployed a virtual EPC (LTE core network) PoC (Proof of Concept) trial in NTT DOCOMO as well during 2014.
- **NEC:** NEC has already deployed a vEPC solution which shall be described later n this chapter.
- **HP Open NFV:** HP developed a solution based on the ETSI standards.
- **ClearWater:** ClearWater in cooperation with Metaswitch developed an open source implementation of IMS (IP Multimedia System) network.
- **Alcatel Lucent (ALU):** ALU in cooperation with RedHat developed a solution called Cloud Band, while a vEPC solution is developed as well.

The following table summarizes some of the industry projects on NFV up to this point:



	Functionality	Platform	Driving Standards
HP OpenNFV	Open standards-based NFV reference architecture, labs as a sandbox in which carriers and equipment vendors can test vEPC.	OpenStack	ETSI
NFV Open Lab	Supports the development of NFV infrastructure, platforms and services.	OpenStack, OpenDaylight	ETSI
Intel ONP	Provides developers with a validated template for quickly developing and showcasing next-generation, cloud-aware network solutions.	OpenStack, OpenDaylight	3GPP or TMF
CloudNFV	Provides a platform for virtual network service creation, deployment, and management.	OpenStack	TMF and ETSI
Alcatel CloudBand	Can be used for standard IT needs as well as for CSPs who are moving mobile networks into the cloud.	Red Hat Linux OpenStack Platform	ETSI
BroadBand NFV	Migrate virtual functions between platforms based on various vendor solutions.		ETSI
Cisco ONS	Automated service delivery, improved network and data center use, fast deployment of personalized offerings.	OpenStack, OpenDaylight	ETSI
F5 SDAS	Extensible, context-aware, multi-tenant system for service provisioning	OpenStack, BIG-IP, BIG-IQ	IETF, 3GPP, GSMA, ETSI, ONF
ClearWater	SIP-based call control for voice and video communications and for SIP-based messaging applications.	Apache Cassandra, Memcached	3GPP IMS, ETSI TS
Overture vSE	Host multiple VNFs in one box, Accelerate service creation, activation and assurance, Decrease inventory and management costs, Optimize service flexibility, Eliminate trucks rolls	Linux Overture Ensemble OSA |, OpenStack	

Table 3. Industry NFV projects. From [2].

Key points in industrial approaches to NFV

As far as the industry and vendors are concerned, it is clear that there are various approaches on the modelling of an NFV and each NFV entity, as described above. The examples listed above are indicative and right now several similar projects are examined by the industry leading vendors. Moreover it seems that the majority of the vendors are focused on the deployment of vEPC cases. The main questions that arise are the following:

1. Which are the network functions that can run in data centers and which can run in network end nodes?
2. Which network functions can be fully transformed on NFVs?
3. Which are the basic hybrid solutions including traditional network structures and NFV deployments that might lead to full virtualization?
4. What are the types of NFVIs that can support NFVs?
5. What are the interoperability aspects of NFV?
6. Can all vendor specific solutions up to now be transformed to NFVs?

1.6 Research projects on NFV

T-NOVA NFV project

T-NOVA is an integrated Project on NFV, co-funded by the European Commission (7th Framework Programme, Grant Agreement no. 619520) with duration of 36 months (January 2014 - December 2016).

In more detail, T-NOVA project aims at designing and implementing an integrated management architecture that includes an Orchestrator platform, for the automated provision, management, monitoring and optimization of Virtualized Network Functions over Network/IT infrastructures.

T-NOVA initially examined the NFV architectures from industry such as CloudNFV, HP OpenNFV, Intel / Tieto, ALU CloudBand, architectures from research projects such as Mobile Cloud Networking, CONTENT, NetIDE, UNIFY and ETSI NFV ISG specifications that we also refer to in current thesis and proposed an architecture fully compliant with the ETSI structures as far as the modules and interfaces are concerned [11]. The T-NOVA architecture is presented below and it should be noted that all interfaces included have the same operation rationale with ETSI.

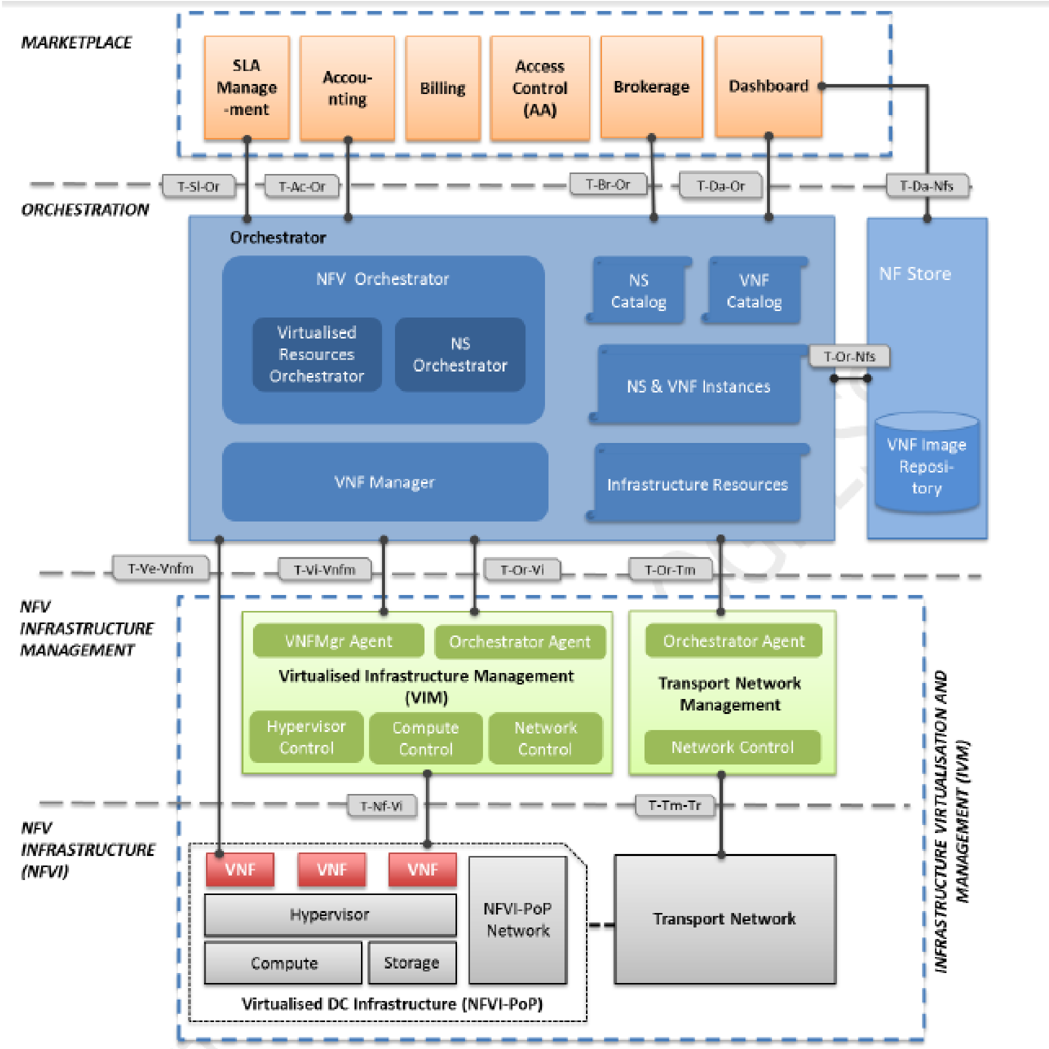


Figure 15. T-NOVA proposed architecture. From [11].

The T-NOVA architecture is based on the following basic components:

- The NFV Infrastructure (NFVI) including all the physical and virtual elements supporting the services that are deployed.
- The NFVI Infrastructure (NFVI) Management layer including all the infrastructure management entities.
- The Orchestration layer and the Marketplace layer that includes the modules that support business related operations and it can be regarded as an addition to the standard ETSI structure.

Mobile Cloud Networking (MCN) project

Another very interesting project that correlates NFV, SDN and Cloud Networking, is the Mobile Cloud Networking (MCN), funded by the European Commission with duration of 36 months starting from November 2012. The target of the project was to investigate and propose solutions on the convergence of Cloud Computing with the mobile communications industry and to provide cloud computing solutions to mobile operators under the aspect of transforming their current network element modules to cloud computing functions as well as virtualizing the latter. Due to this, the main areas of interest cover the following:

- RAN - Radio Access Network.
- EPC - Evolved Packet Core.
- Services (IMS: IP Multimedia Subsystem).
- Operational Support Systems (OSS) including Provisioning, Monitoring and SLA management.
- Business Support Systems (BSS) including billing and charging.

We shall refer to the design approach proposed by MCN for the EPC in chapter 3 where an EPC as a service (EPCaaS) is described.

The following figure depicts the basic architecture components and areas of interest:

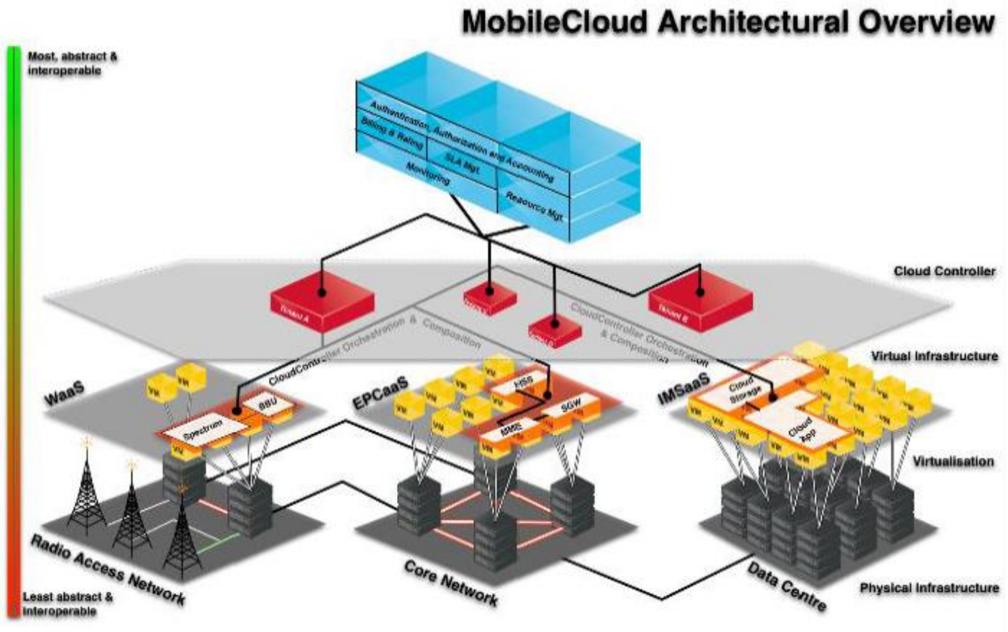


Figure 16. MCN Mobile Cloud Architectural overview. From [12].

Chapter 2: LTE Core and Radio Access Network

In order to analyze virtualization scenarios and propose solutions and examples, background knowledge on the Radio Access Network (RAN) and the core network is definitely needed. This chapter describes the evolution from 2G (GSM) to 3G (UMTS) and 4G (LTE) and describes the main functions of core (EPC) and radio network of LTE (E-UTRAN), since these elements are very good candidates for virtualization.

It is critical to understand the main functionalities of each network element, so that virtualization targets are clearly set. We shall focus on LTE networks since virtual EPCs (LTE core network) as well as virtual eNodeBs are already commercially deployed and operators and vendors understand that there shall be a clear benefit from virtualizing these elements due to their flat IP based architecture, simplicity of interfaces and CAPEX and OPEX savings that virtualization can provide.

2.1 Evolution of GSM to UMTS

The main disadvantages of the 2G GSM architecture for mobile Internet connections were the low data rates, which could not support satisfactory QoS services, and the need for several separate connections for every data transfer link. The solution to the data rate limitations and Internet impairments was the creation of the General Packet Radio Service (GPRS), a derivative of 2.5G networks that added an IP backbone among the mobile core transmission network and led to the development of 3G networks such as the Universal Mobile Telecommunication System (UMTS) [6].

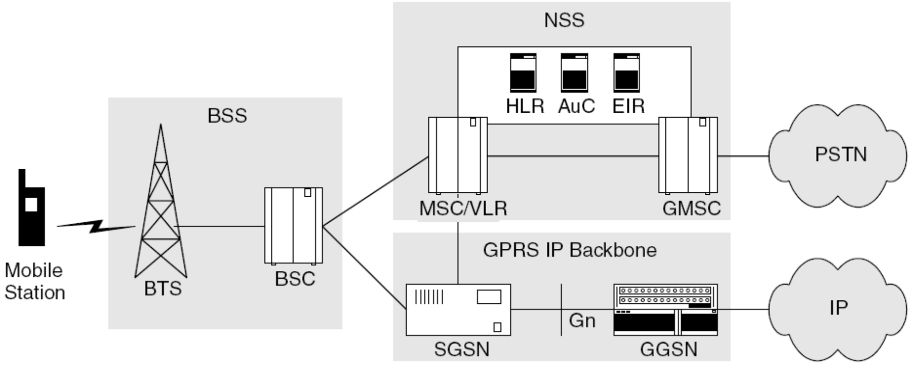


Figure 17. GPRS General Architecture. From [13].

The main elements that provide interconnection with packet switched data networks like the Internet are the Serving GPRS Support Node (SGSN), which is the counterpart of the MSC/VLR for the IP part of the network, and the Gateway GPRS Support Node (GGSN), which is the counterpart of the GMSC in the IP domain connecting the mobile network circuit switched domain with other fixed or mobile networks that might belong or not to the same operator [13], [14].

The IP protocol is used for communication among the SGSN and the GGSN and the standard interface between them is called Gn. The use of IP between these counterparts of the network is the enhancement that provides data services to the mobile user.

UMTS release 99 (R99) was initially developed after several advancements in the GPRS network and was superior from the existing architectures because of the use of Code Division Multiple Access (CDMA) as a multiple access technique [13]. The UMTS Radio Access Network (UTRAN) consists of one or more Radio Access Subsystems (RNS) which further include a Radio Network Controller (RNC) that controls the Node B's while the core network is serving GSM and UTRAN too. The GSM and UMTS network architecture as well as their common interfaces are depicted below:

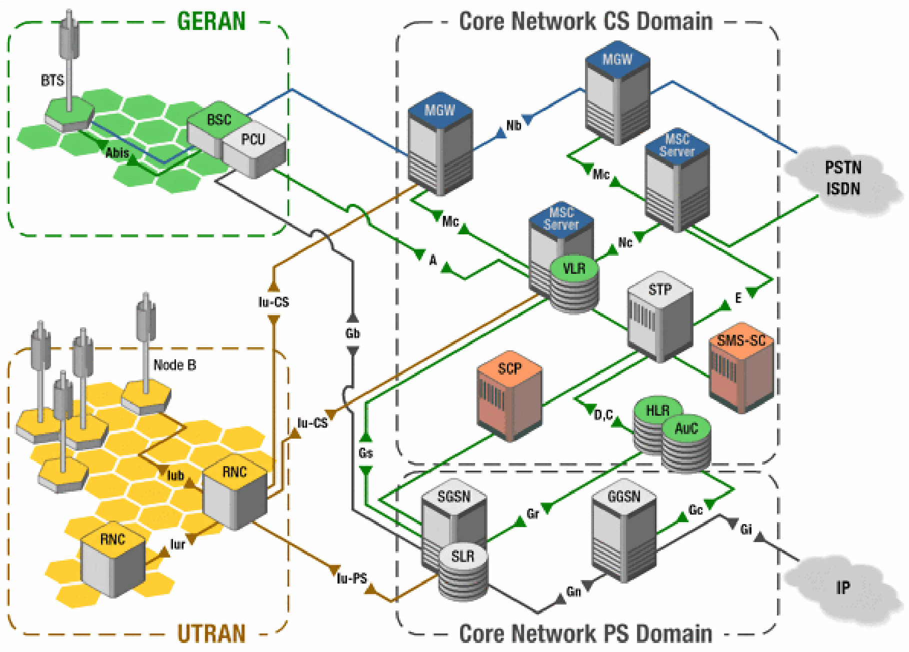


Figure 18. GERAN and UTRAN architecture and interfaces. From [15].

2.2 The LTE network

LTE evolution and all – IP rationale

After UMTS deployment, several enhancements led to the release of LTE standard (Long-Term Evolution or referred as 4G) developed by 3GPP (3rd Generation Partnership Project) which consists of the radio part (E-UTRAN) and SAE (System Architecture Evolution). The first release was Release 8, while now we are in Release 12 [16].

Figure 19. LTE and WCDMA evolution. From [16].

The main element of SAE is EPC (Evolved Packet Core), which provides simplified architecture, all-IP network characteristics, higher throughput and low latency if compared to the UMTS core network while as new technologies in the RAN (Carrier Aggregation, MIMO) and core network are introduced, data rates evolution keeps increasing.

Compared with UMTS, in LTE many functions have been migrated and aggregated to fewer modules and interface simplicity is clear.

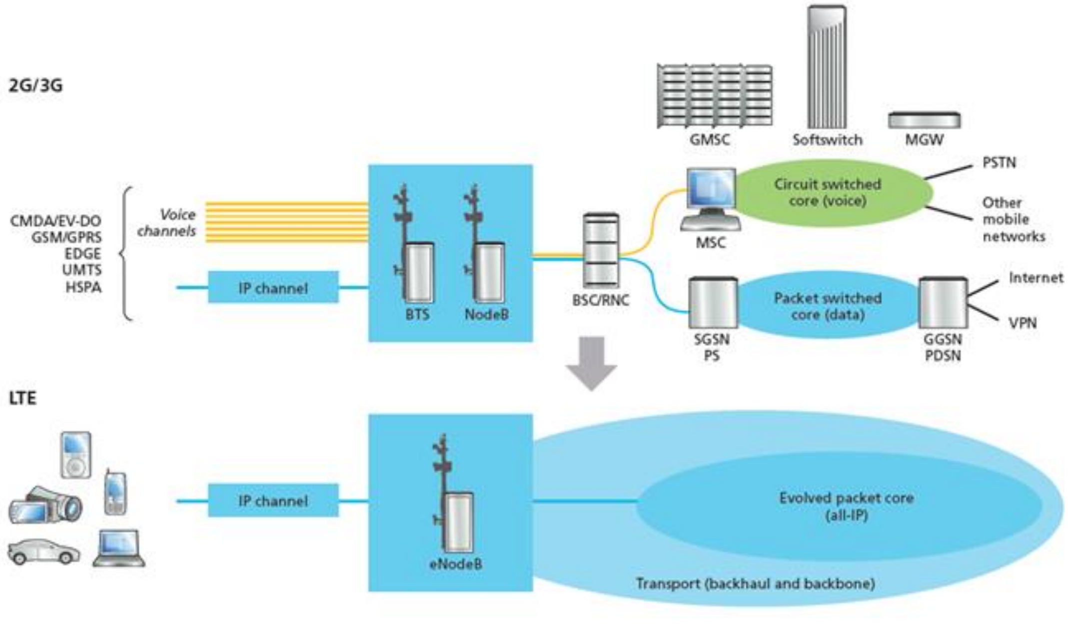


Figure 20. GSM/UMTS/LTE comparison. From [22].

LTE is designed to support mainly Packet Switched (PS) services, contrary to the standard GSM/UMTS networks supporting basically CS (Circuit Switched) traffic, while Internet Protocol (IP) connectivity and mobility between end users and the packet core is ensured through Radio Bearers which are IP data streams with a given set of Quality of Service (QoS) characteristics. Based on this fact, LTE provides seamless connectivity between the end user equipment (UE) and the Packet Data Network (PDN) without disruptions during user mobility [18].

Evolved Packet Core (EPC)

The main network elements of EPC that consist the core network of LTE are the following [20]:

1. Serving Gateway (S-GW).
2. PDN Gateway (P-GW).
3. Mobility Management Entity (MME).
4. PCRF Policy Control and Charging Rules Function (PCRF)
5. Home Subscriber Server (HSS)

The following figures depict the EPC network elements basic architecture and the interconnection architecture of LTE with UMTS, IMS and FEMTO networks.

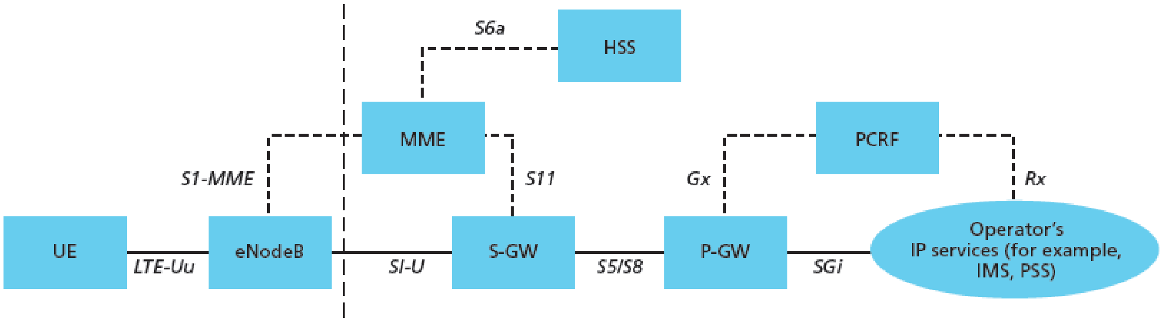


Figure 21. LTE basic architecture [A]. From [22].

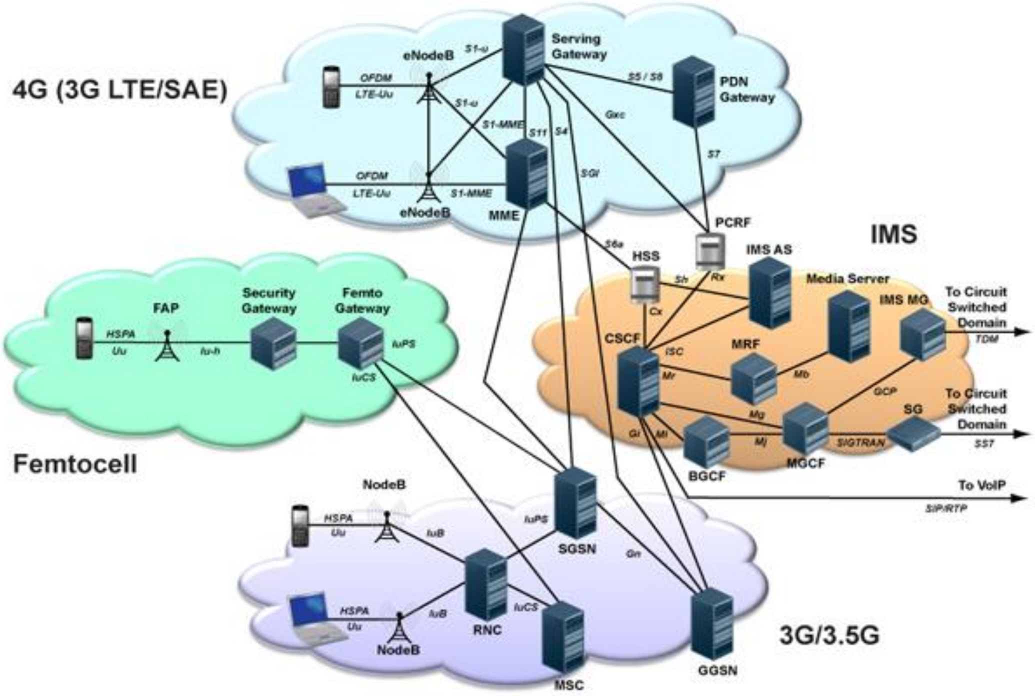


Figure 22. Interconnection of LTE with other networks.

Serving Gateway (S-GW)

The Serving Gateway has the role of the mobility agent for the UE and provides data buffering in cooperation with the paging function of MME, while it is also the basic interconnection point with other 3GPP networks such as UMTS.

All IP packet traffic passes through the S-GW that divides the core network with the radio access part and serves as the mobility anchor as a UE moves and connects with different e-NodeBs, in the sense that traffic is routed through that element during intra system mobility or mobility with external networks such as GSM or UMTS. Finally, lawful interception and basic charging operations can be deployed through the S-GW [18], [19], [20], [22].

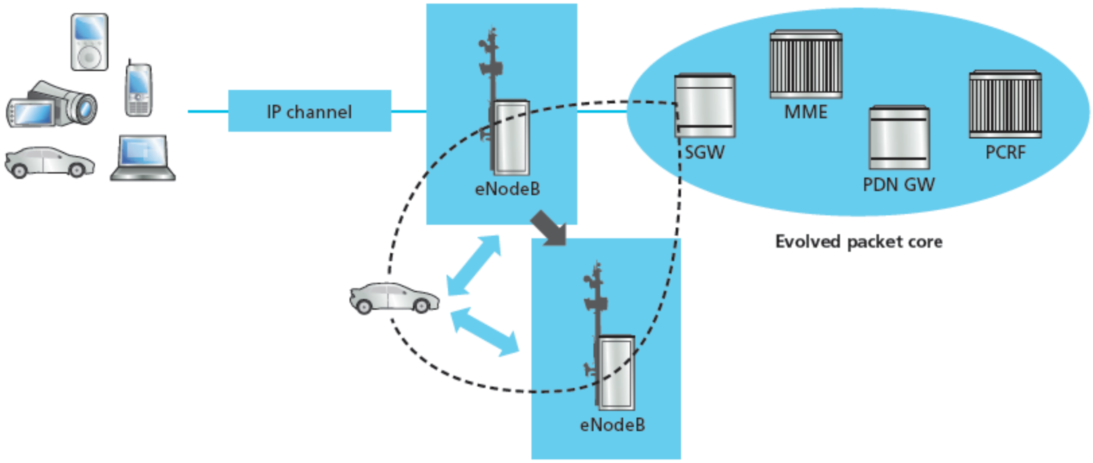


Figure 23. SGW role. From [18].

PDN Gateway (P-GW)

Packet Data Network (PDN) Gateway is assigned to handle the IP address allocation for the UE and the QoS profiles that can be set through the PCRF, acting as the mobility agent with non-3GPP networks such as WiMaX. By deploying downlink IP packet filtering, different traffic flow templates can be assigned per user and various guaranteed bit rate profiles (GBR) can be assigned to the end users [18], [19], [20], [22].

NAT functionality in PDN Gateway

PDN can support Network Address Translation (NAT) functionality, but it is a design decision that the operator must take in order to migrate NAT functions to the end routers [21]. The two possible approaches are shown below:

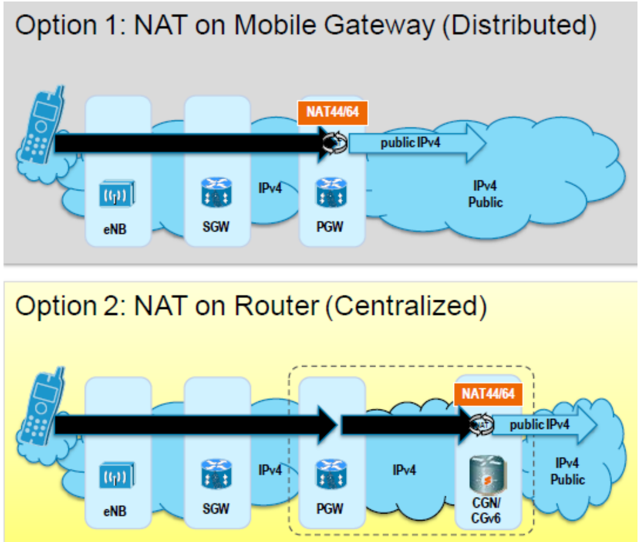


Figure 24. NAT deployment approaches in PDN. From [21].

Control Plane and Signaling Plane

In LTE, as well as in earlier UMTS networks, traffic is divided in control (signaling) plane and user (data) plane. The former regards the signaling communication controlling the bearer data flows while the latter regards the user data traffic that includes the communication among two or more user peers [18], [19], [20], [22].

As already mentioned, LTE has all-IP architecture and the IP presence in both control as well as user plane is present in all network modules.

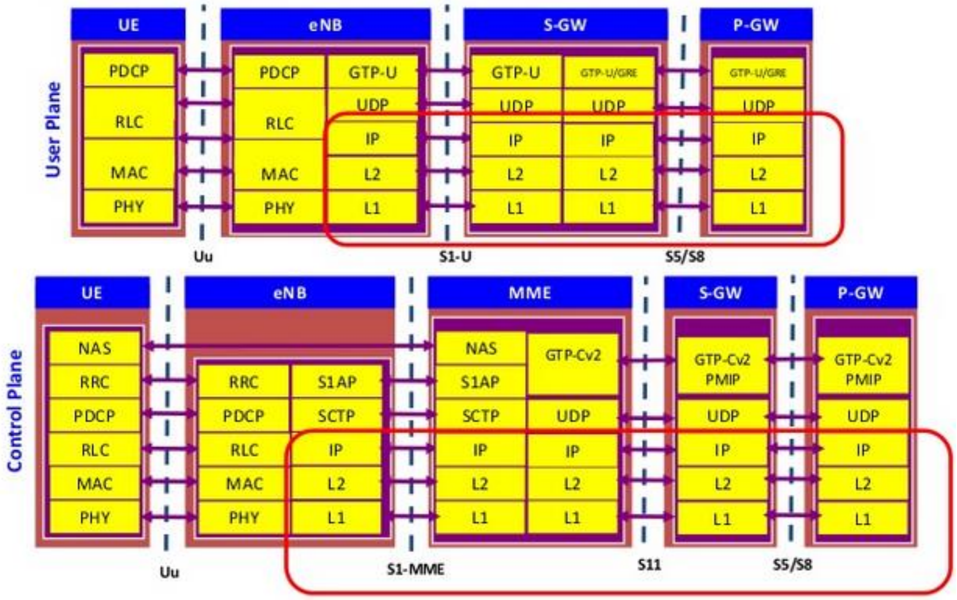


Figure 25. Control and User Plane in LTE [A].

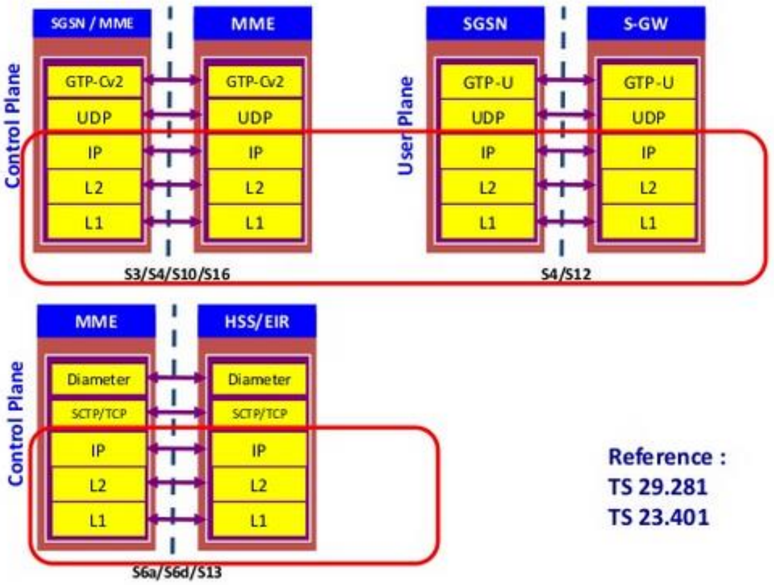


Figure 26. Control and User Plane in LTE [B].

User data plane traffic is carried over virtual connections called service data flows (SDFs) while SDFs are transferred over data bearers with specific QoS (Quality of Service) characteristics. SDFs are usually aggregated so that they can be transferred through one or more bearers [18], [19], [20], [22].

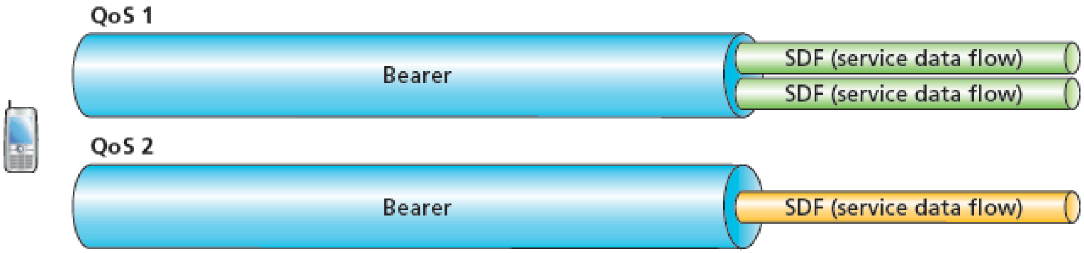
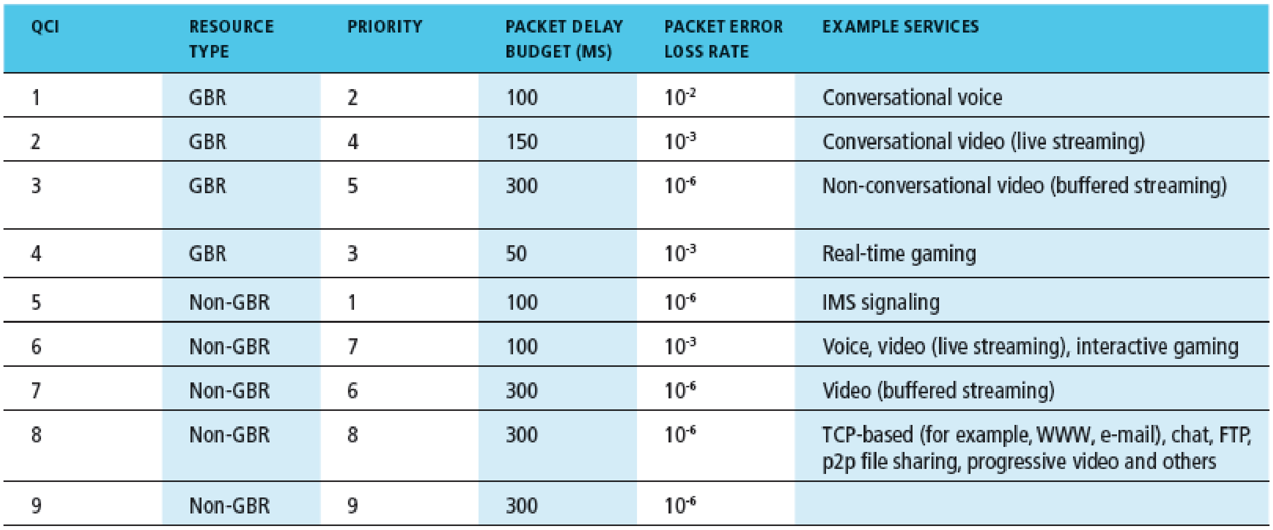


Figure 27. SDF and Bearers in LTE. From [19].

Each bearer has an associated Quality Class Identifier (QCI) which in turn has a specific priority and packet loss rate and each QCI class defines the manner in which the eNodeB handles the bearer.



QCI	RESOURCE TYPE	PRIORITY	PACKET DELAY BUDGET (MS)	PACKET ERROR LOSS RATE	EXAMPLE SERVICES
1	GBR	2	100	10^{-2}	Conversational voice
2	GBR	4	150	10^{-3}	Conversational video (live streaming)
3	GBR	5	300	10^{-6}	Non-conversational video (buffered streaming)
4	GBR	3	50	10^{-3}	Real-time gaming
5	Non-GBR	1	100	10^{-6}	IMS signaling
6	Non-GBR	7	100	10^{-3}	Voice, video (live streaming), interactive gaming
7	Non-GBR	6	300	10^{-6}	Video (buffered streaming)
8	Non-GBR	8	300	10^{-6}	TCP-based (for example, WWW, e-mail), chat, FTP, p2p file sharing, progressive video and others
9	Non-GBR	9	300	10^{-6}	

Table 4. LTE QCI classes. From [18].

Each bearer defining a data path among the UE and the Packet Data Network has the following segments:

- Radio bearer between UE and eNodeB.
- Data bearer between eNodeB and SGW (S1 bearer).
- Data bearer between SGW and PGW (S5 bearer).

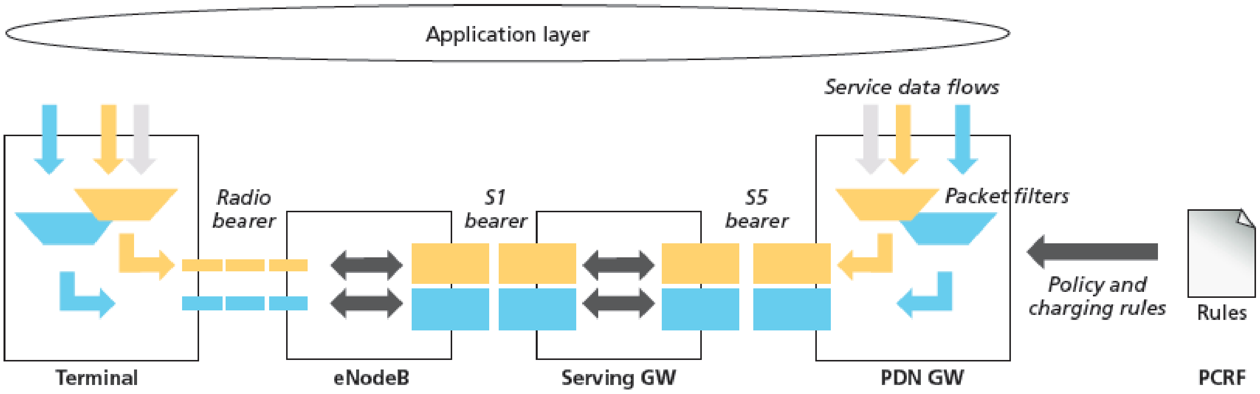


Figure 28. Radio, S1 and S5 bearers in LTE. From [20].

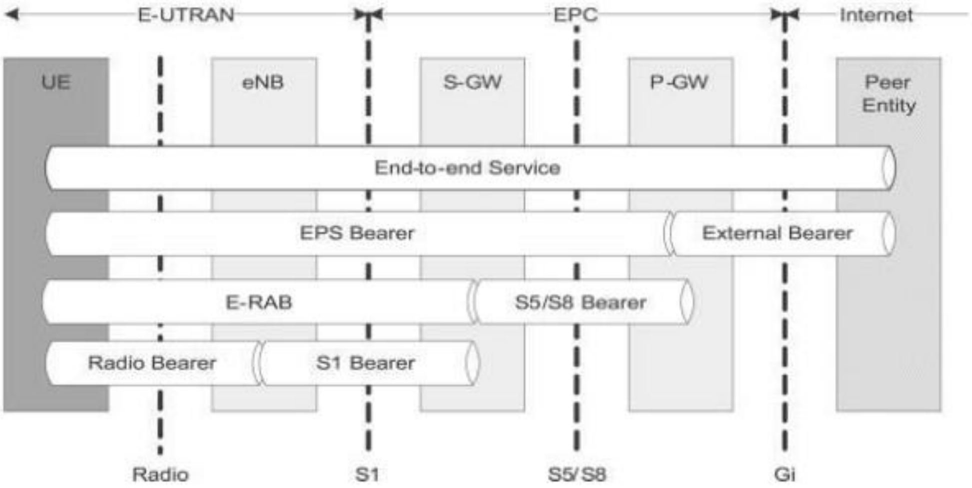


Figure 29. Bearers per LTE network element.

The primary role of PGW and SGW with respect to the bearer and traffic management is that the PGW defines the QoS classes while SGW manages the bearers.

Mobility Management Entity (MME)

The Mobility Management Entity (MME) is responsible for handling the User Equipment (UE) control and signaling functions, network resource sharing, mobility management under the aspect of paging, handover and roaming, bearer management and signaling control of handovers to GSM/UMTS networks, since it selects the gateway to external networks [18], [19], [20], [22]. Furthermore, it should be noticed that all UE control plane functions are managed by the MME.

An MME is able to handle thousands of eNodeBs contrary to the older GSM/UMTS RNC/SGSN architectures. In more detail it is assigned to handle the following basic actions:

- Access Control, authentication and authorization
- Radio Resource Management (RRM)
- Mobility Management
- Roaming with other networks
- Tracking Area Management and subscriber
- Traffic Load Balancing Between S-GWs
- Security operations handling.

The protocols used for the communication of the UE with the Core Network are called Non Access Stratum (NAS) protocols and the connection management processes defined for LTE are similar to UMTS networks.

MME Pool operation

MMEs can operate in a pool spread across different geographical locations leading to redundancy, higher capacity due to load sharing capability, better software upgrade handling and single point of failure avoidance. This operation can be possibly examined as a virtualization case as well.

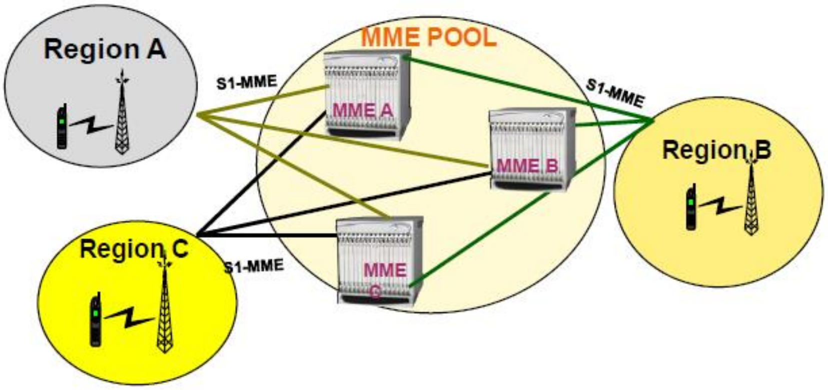


Figure 30. MME Pool in LTE.

Home Subscriber Server (HSS)

The basic role of HSS is to store the subscription data of the user, including QoS profiles and lists of PDN (Packet Data Networks) that a user can connect. These lists are the Access Point Names (APN) including the DNS naming or the IP addresses needed for a PDN to connect.

At this point we should note out that the initial bearer QoS parameters are assigned to the UE by the MME according to the user information fetched by the HSS. Moreover, authentication mechanisms such as an Authentication Center generating keys for each user can be deployed in HSS.

Policy Control and Charging Rules Function (PCRF)

PCRF was developed in order to enable and expand the Policy and Charging Control (PCC) for non-3GPP networks such as WiFi or Wimax access to LTE. Additionally, the basic role of PCRF is to set the QoS parameters and communicate them to the SGW.

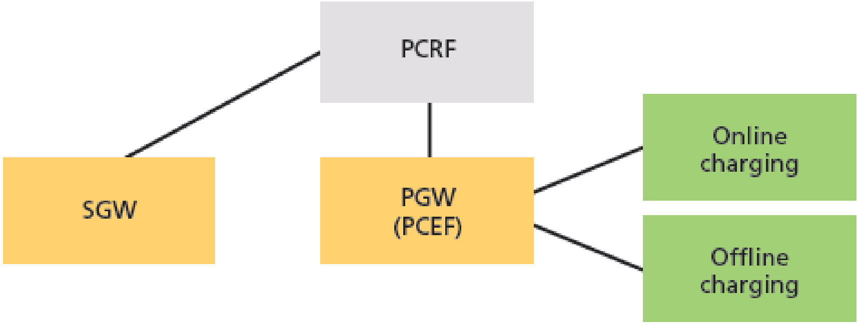


Figure 31. PCRF interconnection. From [22].

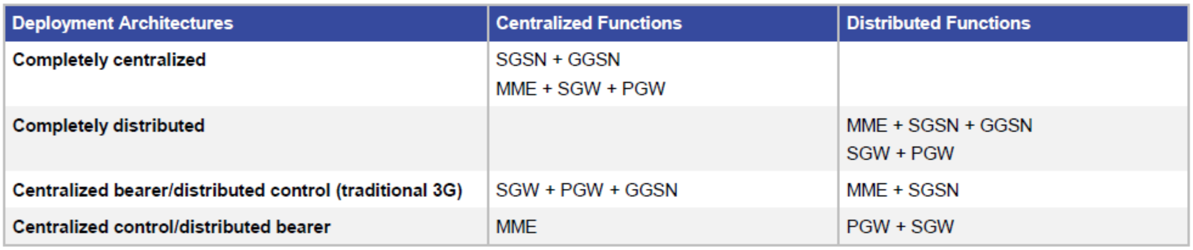
2.3 EPC design and deployment strategies

EPC elements architecture can be deployed in a centralized, semi-centralized or non-centralized manner according to the operator's policy. The centralization of core network elements and their combination in one hardware element is the root of virtualization, since multiple operations can be deployed as a software solution or as a service.

The following typical EPC architecture examples combine network elements operations and de-centralize operations [21]:

- **Combined MME and SGSN:** This approach can be followed in case of existing GSM/UMTS networks where the MME and SGSN communicate for the needs of mobility management and create high signaling load between them and towards other elements such as the HSS and MSC. Due to this a possible combined solution reduces the load and limits the latency between nodes while cost reduction is feasible through this approach.
- **Combined MME, SGSN and SGW:** Even higher cost reduction and performance advancement can be achieved by adopting this solution leading up to 80% CAPEX savings.
- **Combined SGW and PGW:** This solution enables traffic offloading and reduces dramatically backhauling communication costs.

The following table summarizes the deployment scenarios of combined and collocated network elements in EPC [21]:



Deployment Architectures	Centralized Functions	Distributed Functions
Completely centralized	SGSN + GGSN MME + SGW + PGW	
Completely distributed		MME + SGSN + GGSN SGW + PGW
Centralized bearer/distributed control (traditional 3G)	SGW + PGW + GGSN	MME + SGSN
Centralized control/distributed bearer	MME	PGW + SGW

Table 5. EPC design approaches and functions placement [A]. From [21].

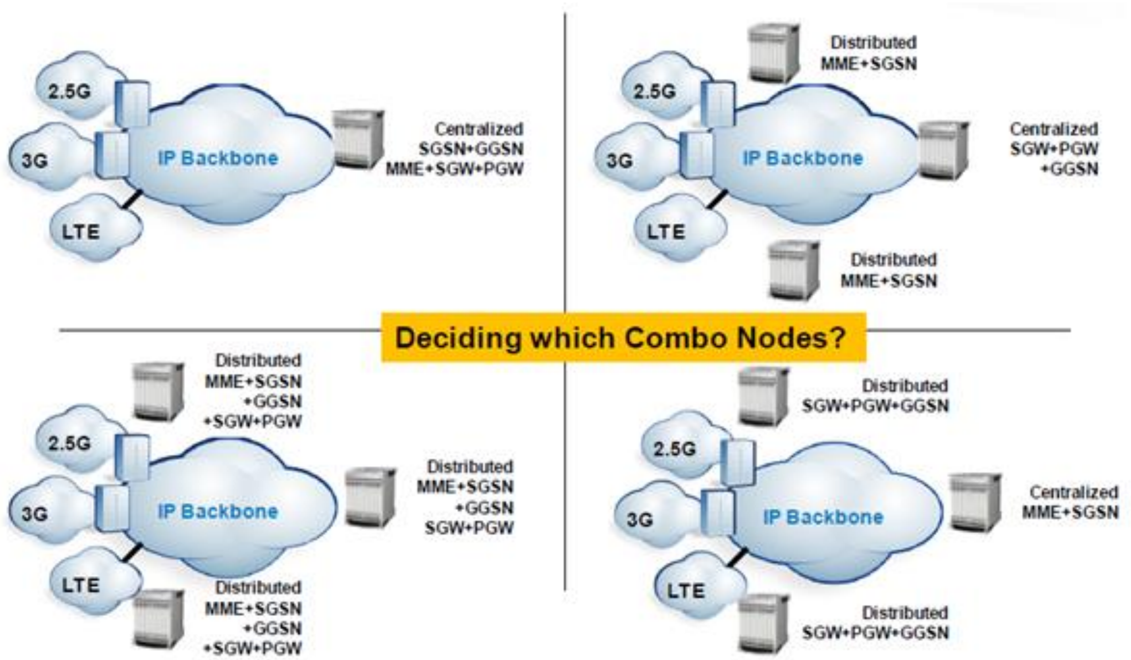


Figure 32. EPC design approaches and functions [B]. From [21].

2.4 LTE Radio Access Part (E-UTRAN)

The eNodeBs are the main elements of the Radio Access Part (RAN) part of LTE (often called E-UTRAN) and the architecture is flat since many operations which exist in the RNCs in UMTS have been migrated to the eNodeBs, thus there is no separate module controlling the eNodeBs. The rationale behind the removal of RNC as a node is the target to minimize the number of nodes and avoid diversity for dedicated user traffic [18].

The X2 interface is used to interconnect the eNodeBs and support handovers, while the S1 interface is used to interconnect the eNodeBs to the MME and to the SGW. The former interface is called S1-MME while the latter S1-U. It should be noticed that S1 handover is also feasible in case of X2 unavailability. eNodeBs and the UEs communicate through Access Stratum (AS) protocols [18].

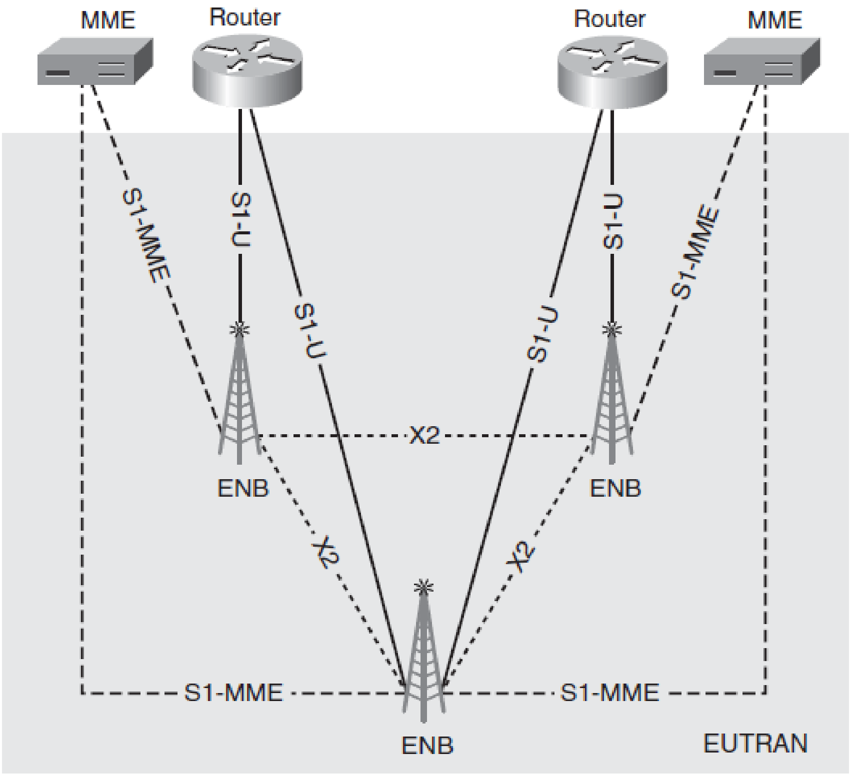


Figure 33. EUTRAN architecture. From [24].

The basic E-UTRAN functionalities are described below:

- **Radio resource management (RRM):** Management of bearers, admission control (Evaluation of sufficient resources for connection and bearer setup), handover and mobility control, resource scheduling and resource allocation.
- **Header Compression:** Compressing packets (header part) for latency sensitive applications such as VoIP.
- **Security:** Encryption of information.
- **EPC Interconnection:** Handling signaling and bearers flow towards and from the MME and SGW.

The basic elements of an eNodeB are the Baseband Unit (BBU), which performs the digital signal processing, monitoring of the eNodeB and acts as the interconnection point with MME, and the RRU (Remote Radio Unit) or RRH (Remote Radio Head), that mainly performs the baseband to RF signal transform and vice versa. In the majority of cases, the BBU and the RRU are interconnected through a Common Public Radio Interface (CPRI) cable.

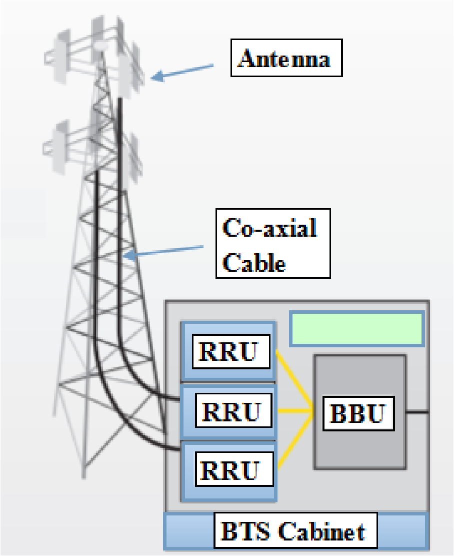


Figure 34. Baseband Unit to Radio Units interconnection.

Chapter 3: LTE EPC (Evolved Packet Core) Virtualization aspects

In the specific chapter we present the main drivers for the virtualization of EPC, as well as the basic architectures proposed by research and industry.

3.1 NFV in the mobile core -The vEPC case and design approaches

EPC (Evolved Packet Core) for LTE networks is an interesting virtualization case, since EPCs elements can break down in multiple VNFs, but special attention must be paid to the provision of user plane services, virtualization must not create a performance trade-off for the end users.

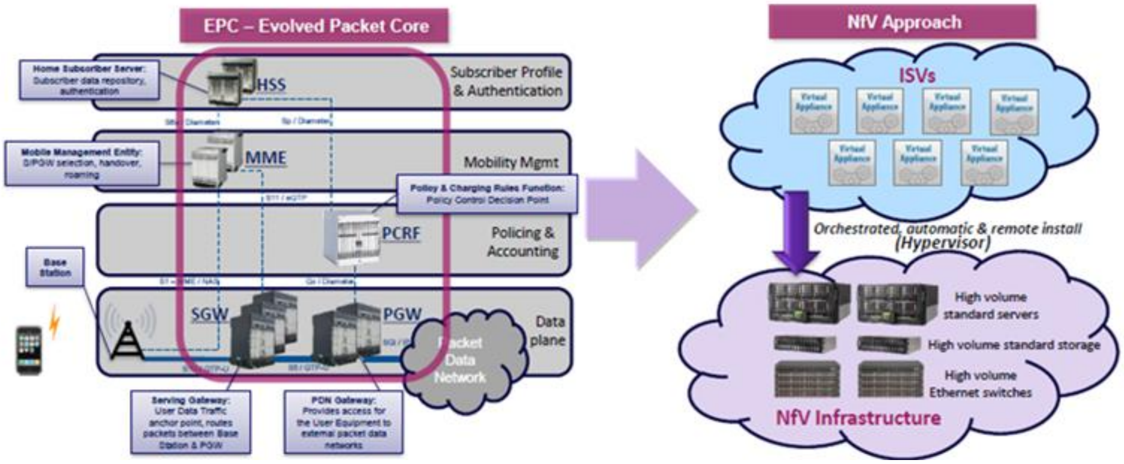


Figure 35. Moving from traditional EPC to the virtual EPC. From [25].

NFV is the optimum solution to the core network issues of the mobile operators that derive mainly from the unpredictable traffic generated. NFVs can replace standard EPC structures without sacrificing user experience, while fast expansion and network scalability is feasible as well. Additional drivers for NFV core network deployments might include:

- Vo-LTE (Voice Over LTE)
- New mobile services (M2M)
- IP edge nodes flexible expansion
- MVNO (Mobile Virtual Network Operators) deployments
- Core and RAN network sharing
- IMS expansion.

Apart from the pre-mentioned drivers, emergency handling is an interesting vEPC scenario since congestion cannot be avoided with current LTE deployments. An example is listed below and regards an earthquake that occurred in Japan where congestion was clear. With the usage of vEPC, resources from other nodes could have been used to avoid such a scenario.

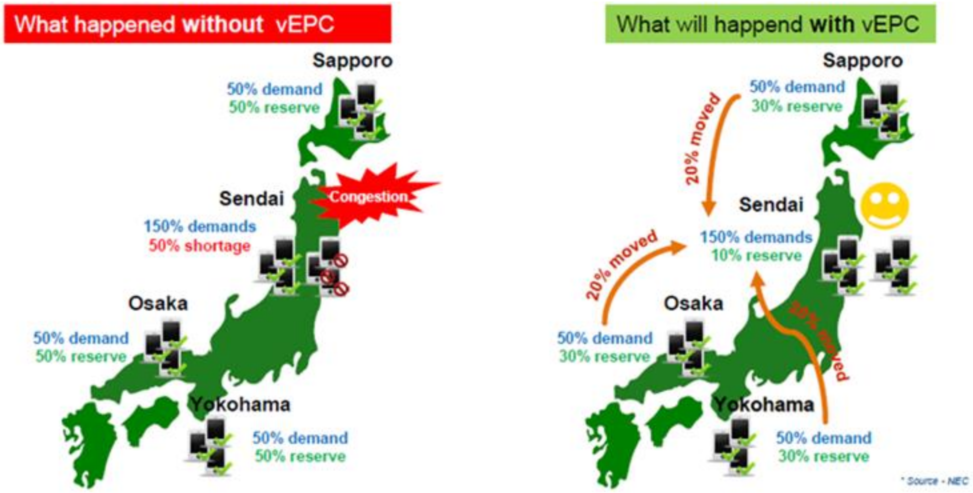


Figure 36. EPC versus vEPC performance in traffic surges (Earthquake). From [25].

The following deployment scenarios that might create high capacity demands can be considered as ideal for vEPC application scenarios [25].

Figure 37. vEPC possible use case scenarios. From [25].

In the EPC case, the elements that are good candidates for virtualization are the following:

- Mobility Management Entity (MME)
- Policy and Charging Rules Function (PCRF)
- Authentication, Authorization and Accounting (AAA)
- Home Subscriber System (HSS)
- Serving Gateway (S-GW)
- Packet Data Network Gateway (P-GW)

The Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW) can be separately virtualized based on the fact that traffic bottlenecks occur in the user plane, thus leading to separate control per customer.

Additionally, the Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW) are designed for achieving high throughput of user packets, while MME/SGSNs have fewer requirements as far as throughput and capacity is concerned but stricter requirements in terms of processing latency. Due to this, the control plane entities are more suitable for being virtualized under the aspect of taking of advantage of the high availability of computing resources in the Cloud. However, under specific use cases user plane functions can be virtualized as well.

3.2 Research on vEPC, architectures and LTE elements modelling

vEPC research work - The EPC as a service approach (EPCaaS)

Possibly the most interesting research activity related with the virtualization of EPC is the MCN (Mobile Cloud Networking) research project. The specific project led to the development of multiple architectural models for EPC as a service (EPCaaS) [26]. The two basic architecture approaches described are the following:

- **Full virtualization:** With this approach all control and user plane functions are virtualized and deployed in VMs.
- **Partial virtualization:** With this approach only control plane functions are virtualized and deployed in VMs while data plane user traffic is forwarded by high performance hardware switches. ***Additionally the MCN Orchestration framework manages control plane VMs, but only controls the forwarding of user plane on the hardware switches, for example using SDN (Software-defined Networking) [26].*** The following figure depicts the pre-mentioned rationale:

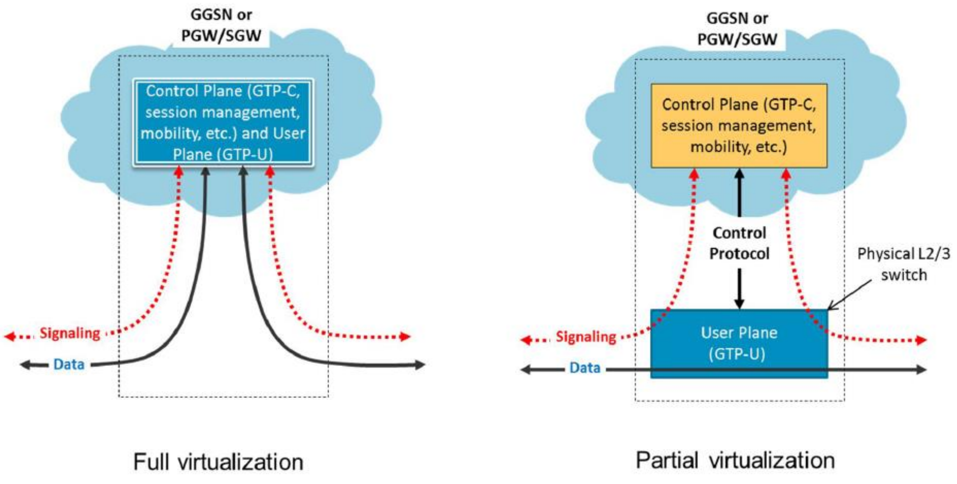


Figure 38. Full versus partial Virtualization approach. From [26].

During the project the key enablers and areas that software was developed for are the following [26]:

1. **Cloudification enablement software:** Aiming to enable the deployment of modules that support the operation of EPCaaS on a cloud infrastructure. The specific modules include:
    - The EPCaaS Service Orchestrator (EPCaaS SO).
    - The EPCaaS Service Manager (EPCaaS SM).
    - The Mobility and Bandwidth Prediction as a Service (MOBaaS).
    - The dynamic configuration mechanism for the software EPC components (EPCaaS Config).
2. **Functions supporting cloudification:** Aiming to deploy software components that realize the parallel running of EPC elements over the cloud.
3. **EPC software that can run on for cloud deployments:** Possibly the most important deployment regards the design and adaptation of the EPC elements and their translation to software including the standard functions that each LTE network element already provides including load balancing and interface communication.

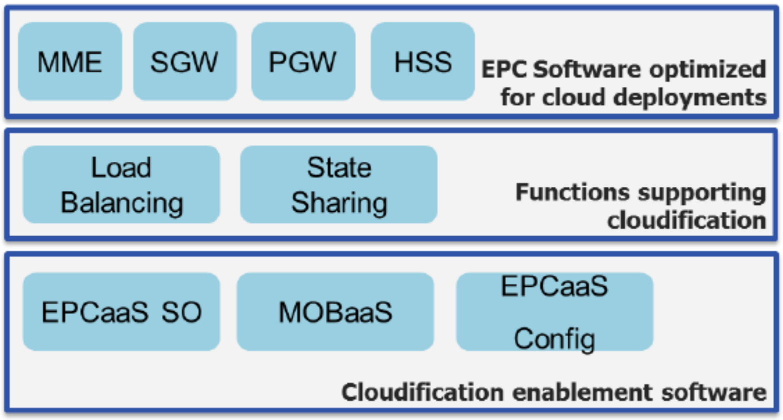


Figure 39. Key software enablers. From [26].

The architectural rationale is that a service manager (SM) can provide an interface for various service instances of a certain kind while in order to create, configure, orchestrate and manage the Service Instances (SI), a Service Orchestrator (SO) is used [27]. SIs can be created after the SO informs the Cloud Controller (CC) and after a SI is operational the end user gets an interface. Moreover, management interfaces are accessed through the service instance (SI) and services can be classified in atomic and composed with the former not being able to be further decomposed. [27].

EPCaaS can be a part of MCN services and the collaboration with the cloud controller is depicted below.

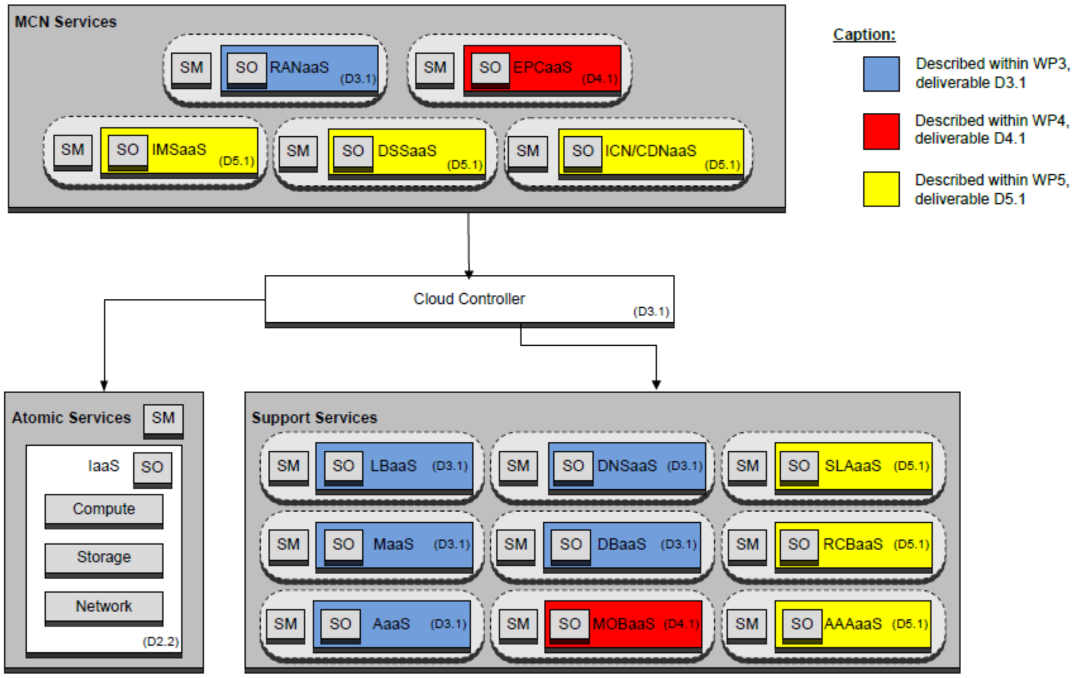


Figure 40. Cloud Controller and MCN services. From [27].

Additionally, as far as the EPCaaS deployment is concerned the following figure depicts one of the four basic architectures developed by MCN based on 1:1 mapping of EPC network elements in SIs (Service Instances).

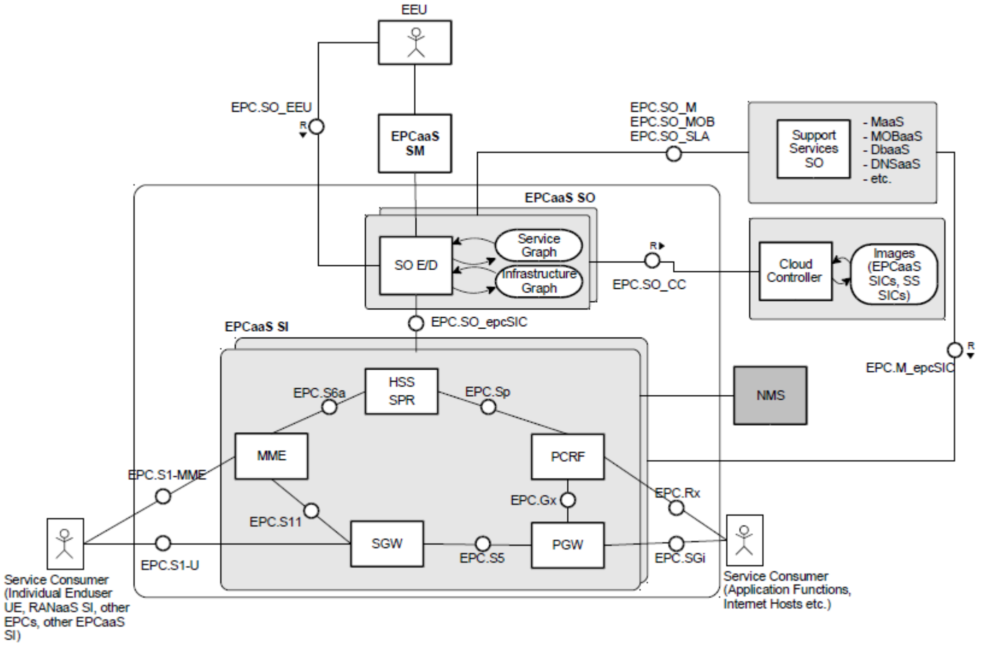


Figure 41. 1:1 mapping of EPC elements in Service Instances. From [27].

Modelling examples of LTE network elements

It is clear that the conceptual modelling of LTE-EPC modules under the aspect that each module represents a software class is one of the most important processes in order to proceed with vEPC architectural design and software development. The following modelling example comes from IBM's Network Connectivity and Inventory Model (NCIM) topology database which is used from IBM Tivoli Network Manager. This product enables network discovery, device monitoring, topology visualization, and root cause analysis (RCA) capabilities. Each LTE interface represented through a sub-class can inherit common attributes and characteristics from a parent LTE Interface as shown below:

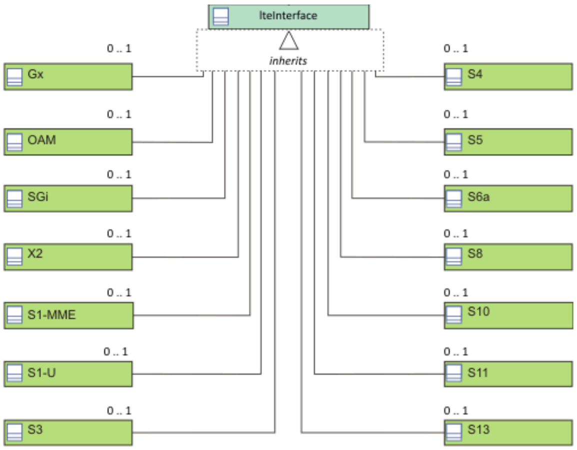


Figure 42. LTE interfaces modelling and inheritance. From [28].

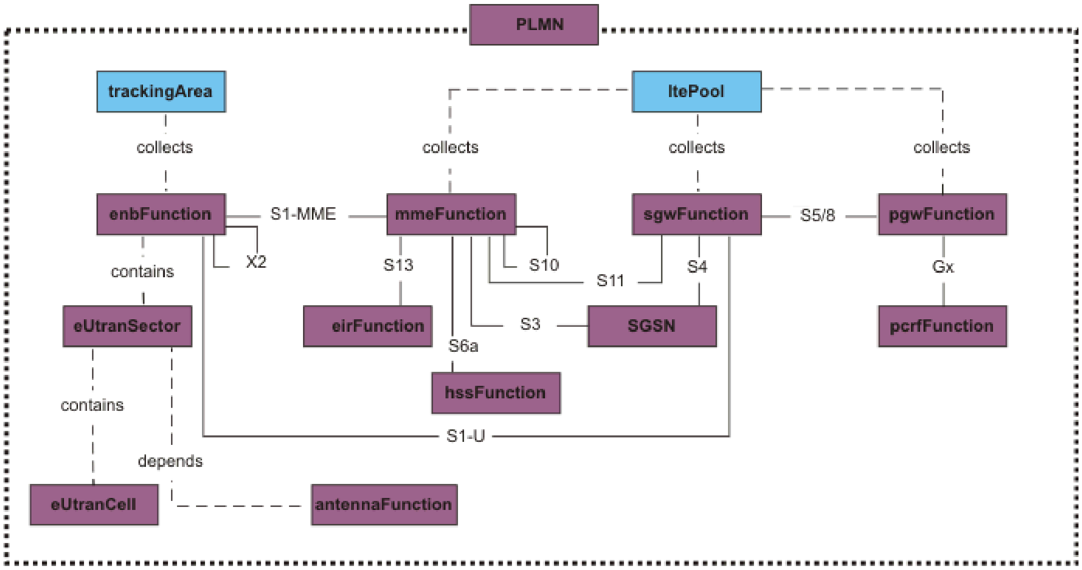


Figure 43. EPC and EUTRAN network elements modelling (Basic UML diagram). From [28].

MME can be represented by the following UML class diagram:

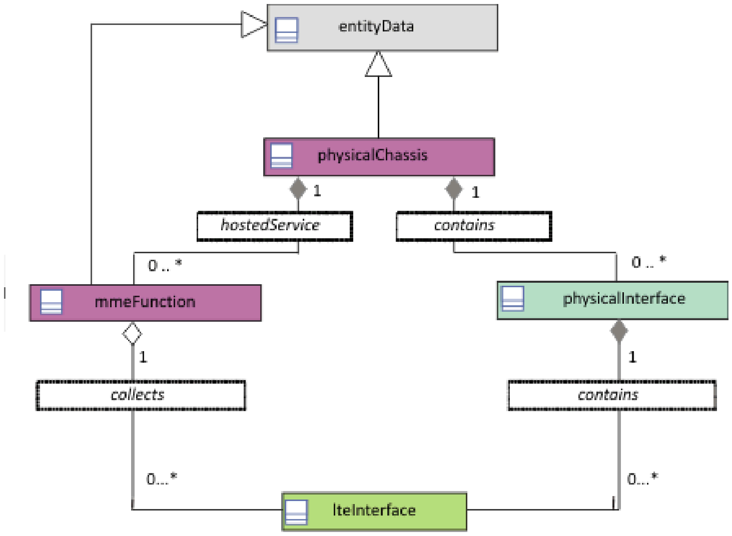


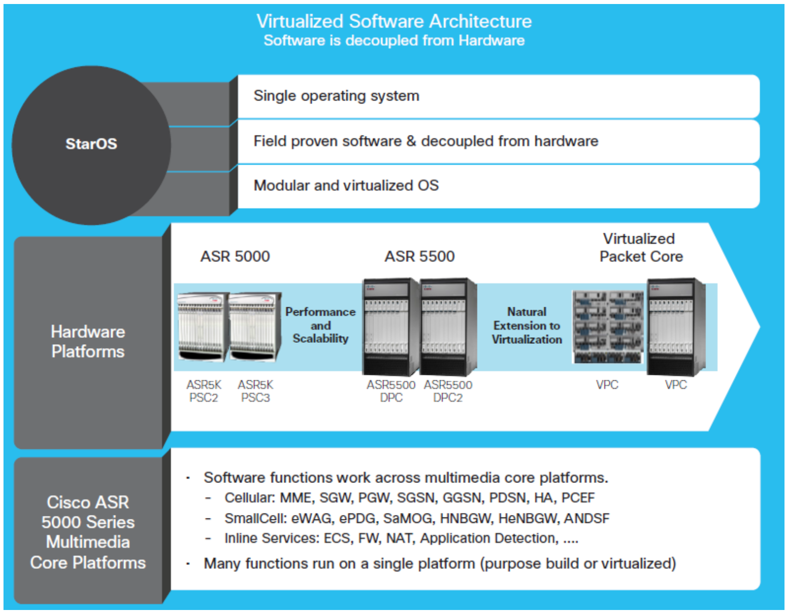
Figure 44. MME class diagram. From [28].

3.3 Examples of vEPC deployments by vendors

In the following paragraphs we present some indicative examples of commercial vEPC deployments coming from leading vendors around the world.

CISCO Quantum Virtualized Packet Core (VPC)

CISCO developed a key solution called Quantum vPC (QvPC) for the needs of packet solution virtualization that is able to support all packet core services of GSM/UMTS/LTE/WiFi and small cells technologies. The virtual packet core (VPC) software runs on top of virtualized StarOS software, which was already deployed since 2007 in order to provide scalable and distributed services, agility in new functions and resource optimization, combining by this manner NFV and SDN [30], [31]. The solution is hypervisor and hardware independent and can run on the already existing CISCO ASR 5000 Series platforms.



- Software functions work across multimedia core platforms.
  - Cellular: MME, SGW, PGW, SGSN, GGSN, PDSN, HA, PCEF
  - SmallCell: eWAG, ePDG, SaMOG, HNBGW, HeNBGW, ANDSF
  - Inline Services: ECS, FW, NAT, Application Detection, ...
- Many functions run on a single platform (purpose build or virtualized)

Figure 45. CISCO Quantum QvPC solution. From [31].

The European Advanced Testing Center (EANTC) was assigned by CISCO to perform additional testing on QvPC solution of S-GW and P-GW. Two UCS B200 blade servers running two hypervisors (KVM QEMU version 1.4.0 and VMware ESXi 5.1.0) where used in cooperation with ASR 5000 platform. Additionally, the SPIRENT LANDSLIDE platform was used in order to emulate UMTS/LTE users, the NodeB/eNodeB, the SGSN and the MME [30].

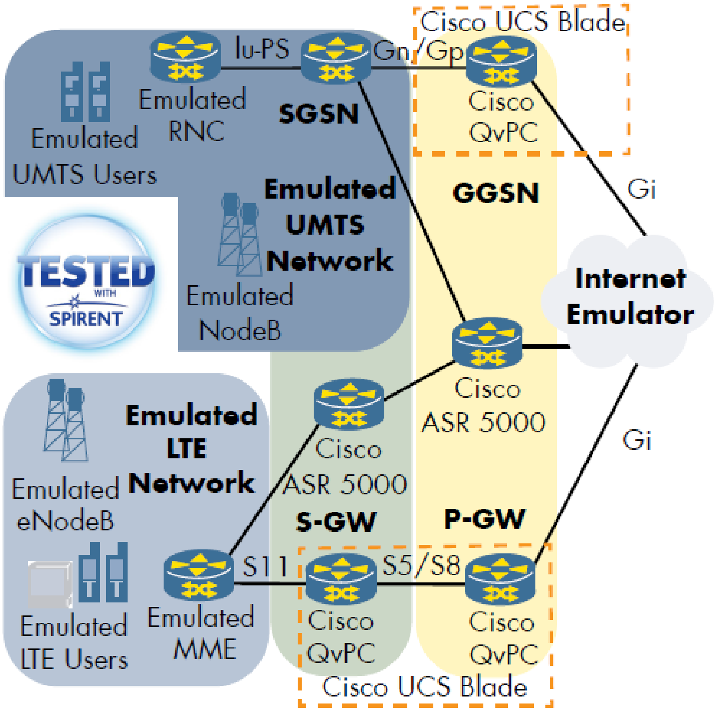


Figure 46. CISCO Quantum QvPC solution test setup. From [30].

During the test procedure, the existing configuration of ASR 5000 was copied to the two QvPC instances in order to check if the specific instances can support the same calls. The only change applied regarded the IP addresses that obviously could not be the same [30].

Additionally, commands that could run on a standard mobile network such as “show ip interface”, “show p-gw service” where executed in both ASR5000 and QvPC and returned the same output leading to the outcome that QvPC behaves exactly in the same way that the standard hardware ASR 5000 does, while the same command set can be used in both cases [30].

High availability testing also took place, to check that in case a CPU of the system fails, then traffic is transferred to the alternative host.

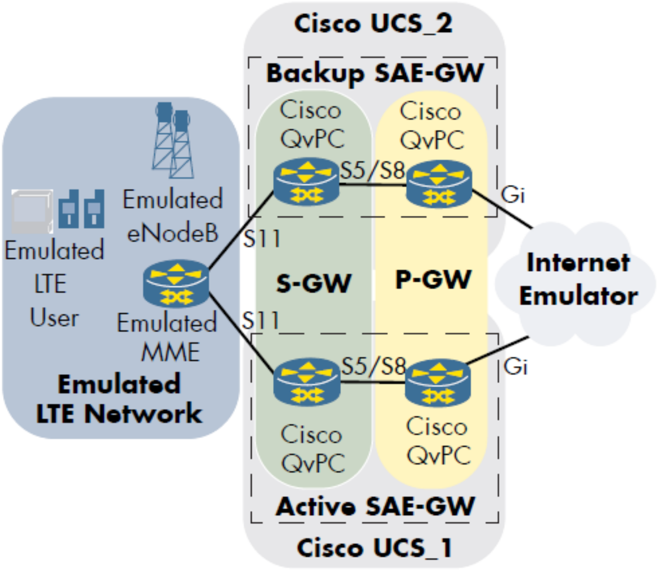


Figure 47. CISCO Quantum QvPC High availability testing. From [30].

During the high availability testing, 10.000 users where emulated and where sent data traffic while when one of the two UCS servers was switched off, no session was lost. Additionally, it took no more than 26 seconds to move all traffic from one server to the other. As a final testing outcome, it can be said that the virtualized solution behaved exactly as a standard hardware solution does in terms of high availability and load balancing.

ALCATEL vEPC deployment

The following vEPC deployment example comes from ALCATEL (before merging with NOKIA) where in the packet and service delivery domain a service router (Alcatel-Lucent 7750), acting as a Mobile Gateway provides virtual SGW, PGW and GGSN network functions.

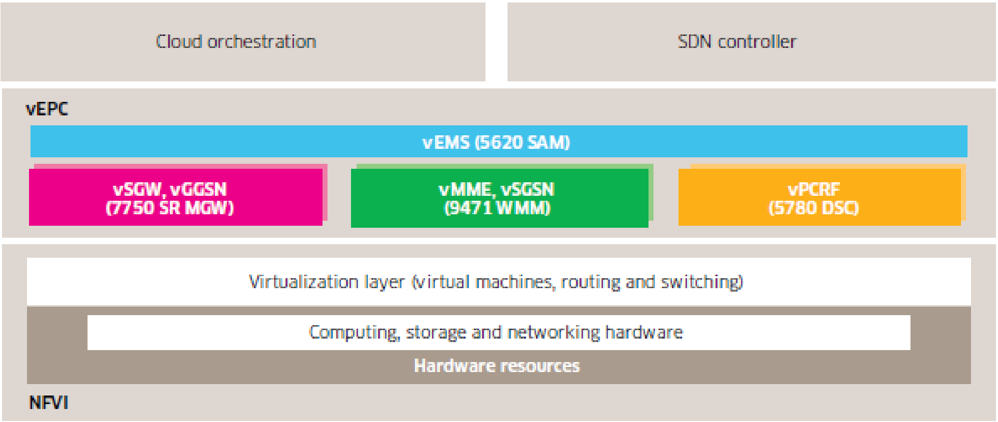


Figure 48. ALCATEL vEPC architecture. From [32].

The signaling and mobility management part is provided by the Alcatel-Lucent 9471 Wireless Mobility Manager (WMM) that virtualizes the MME and SGSN operations and as far as policy and charging is concerned the solution uses the Alcatel-Lucent 5780 Dynamic Services Controller (DSC) that deploys the virtual PCRF functions. The management of the vEPC is based on the Alcatel-Lucent 5620 Service Aware Manager (SAM) enabling backhaul, mobile core and radio access network (RAN) monitoring.

NEC vEPC deployment

Another example solution for vEPC is provided by NEC. With the specific solution Mobility Management Entity (MME), Packet Gateway (P-GW) and Serving Gateway (S-GW) are designed to run in a virtual machine (VM) providing virtualized functions and operating in an isolated manner so that performance is not affected. All VMs are interconnected while commercial off-the-shelf (COTS) servers are used as a hardware that hosts the vEPC solution.

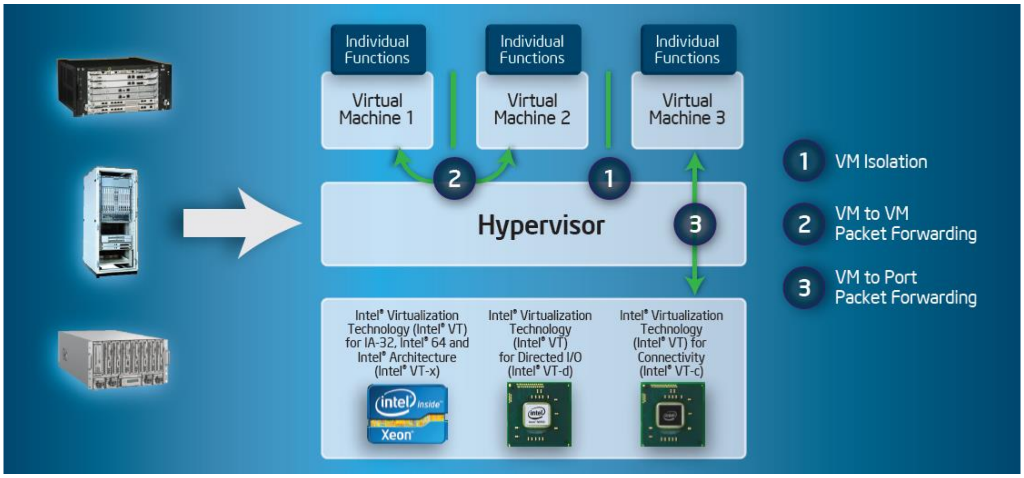


Figure 49. NEC vEPC solution architecture [A]. From [33].

NEC uses the Carrier Grade HyperVisor (CGHV) module that provides disk space, CPU, memory and network resources handling in order to orchestrate the multiple VMs operations, providing load balancing between them and avoiding any operation conflicts. Additionally, log collection and VM management is deployed through CGHV centrally, leading to minimized troubleshooting time in case of failovers, while low latency and jitter are ensured for Dataplane traffic.

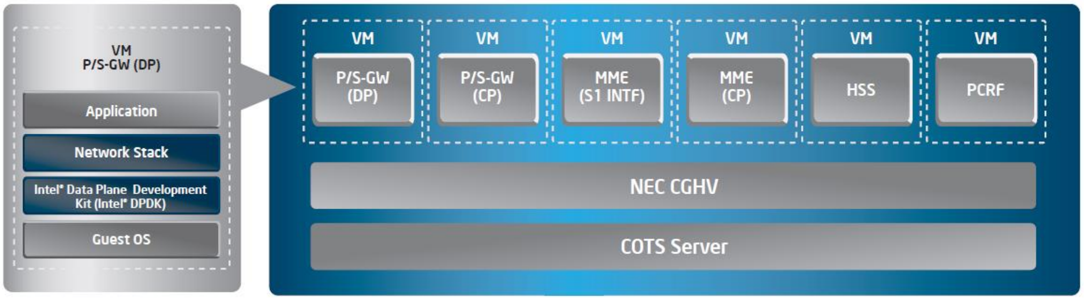


Figure 50. NEC vEPC solution architecture [B]. From [33].

The hypervisor provides an open API (Application Programming Interface) so that multiple virtualized applications can run through this and each virtualized network element is an aggregation of multiple VMs with each of them allocated individually to the control plane or the data plane of the element. Through this technique different capacity characteristics can be provided to different network nodes.

Additionally, the role of the SDN controller is to create and manage the VMs according to the different traffic characteristics and as a next step assign them to the appropriate server dynamically according to the resource availability. Finally, VM interconnection is ensured through virtual nodes in the SDN controller.

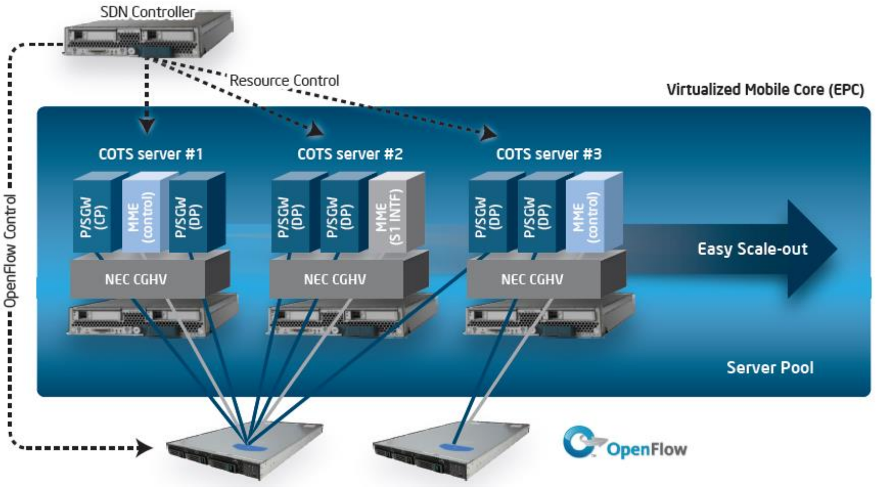


Figure 51. NEC vEPC solution architecture [C]. From [33].

Chapter 4: C-RAN

This chapter refers to the idea of virtualization in the radio access part of mobile networks realized through the C-RAN approach. We present the main reasons that led to the need for C-RAN deployments, including a CAPEX and OPEX analysis, and the basic deployment architectures up to the time of speaking.

4.1 Virtualization in the mobile access network - The C-RAN approach.

Drivers for virtualization in the radio access part

Recent total cost of ownership (TCO) research has indicated that mobile operators start facing difficulties in expanding their Radio Access Network (RAN) infrastructure due to the fact that the revenue growth is not increasing at the same rate as the cost of RAN equipment does. At the same time. RAN expansion cannot be avoided since the subscriber's demand for new services, higher rates and capacity is constantly increasing [35].

The standard RAN expansion until now is based on the acquisition of new sites or the swapping of pre-exiting deployments with newer equipment coming from the same or a different vendor. Each Base Station (BS) covers a specific area while system capacity is always restricted by interference concerns, especially for the case of LTE where interference (SINR degradation) creates non-satisfactory user experience and lower throughput. CAPEX and OPEX (CAPEX is related to network infrastructure build while OPEX is mainly related to network operation and management) are constantly increasing with the current RAN expansion policy, since site acquisition cost, equipment cost, power consumption, planning and management costs cannot be trespassed while the latter costs are aggregated since a separate investment must be made for every site individually.

Single RAN (SRAN) solutions adopted during the last years helped CAPEX and OPEX savings, since sites already operating in one technology (GSM, UMTS, LTE) were equipped with the missing technology by adding new baseband units (BBU) or baseband unit processing cards for the new technology (usually LTE), as well as antenna systems and other RF passive devices. However, the cost remains high even with this solution. Furthermore, the fact that equpment / solutions usually come from different vendors creates management / planning difficulties and extra complexity.

Finally, as RAN technologies evolve and the path towards 5G from LTE (4G) is the next step, challenges related to bandwidth, throughput, latency and coverage arise, which in turn shall lead to the need for extra hardware investments.

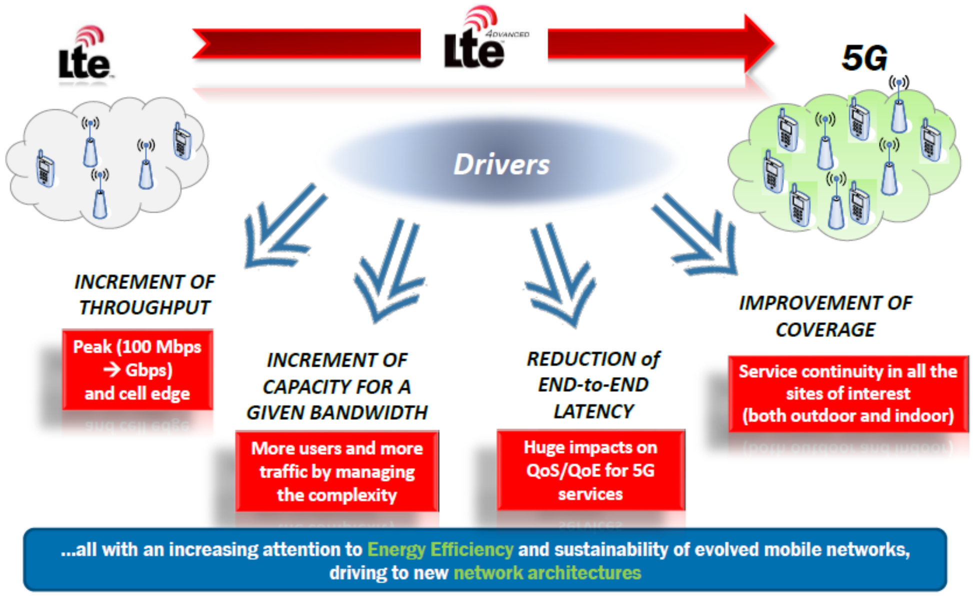


Figure 52. Main drivers for the transition from 4G (LTE) to 5G.

It is clear that operators apart from the cost related issues and the fact that an energy sustainable policy must be adopted, are expected to satisfy the upcoming traffic explosion, adopt multi-standard solutions, simplify interfaces and provide new services. The adoption of C-RAN as the virtualization solution for the radio access part can fully support the pre-mentioned demands.

The C-RAN approach

C-RAN (Cloud-RAN), or Centralized-RAN, was first proposed as an architecture by China Mobile Research Institute in April 2010 providing a unified centralized and cloud computing solution for 2G, 3G, 4G networks [35]. It should be noted that the first commercial C-RAN solution was deployed just two months before the presentation of our thesis project during February 2016 [43].

C-RAN can provide the following benefits to mobile operators:

- CAPEX and OPEX savings for mobile operators.
- Power efficiency and energy consumption savings.
- Support of multiple standards and interfaces.
- Spectral efficiency and capacity improvement.
- Provision of a platform for the development of new services.
- Interference and load balancing handling.

The C-RAN solution is mainly centralized since it can provide Base Band (BB) pool processing for a set of base stations and can reduce equipment and management costs by aggregating into one single point the handling of Remote Radio Head units (RRH). Higher spectrum efficiency and interference management can be provided centrally, while its Cloud Computing capabilities based on virtualization enable several management features including resources allocation, power consumption and utilization [35], [39].

The comparison between current RAN deployments and C-RAN is described by the following figure:

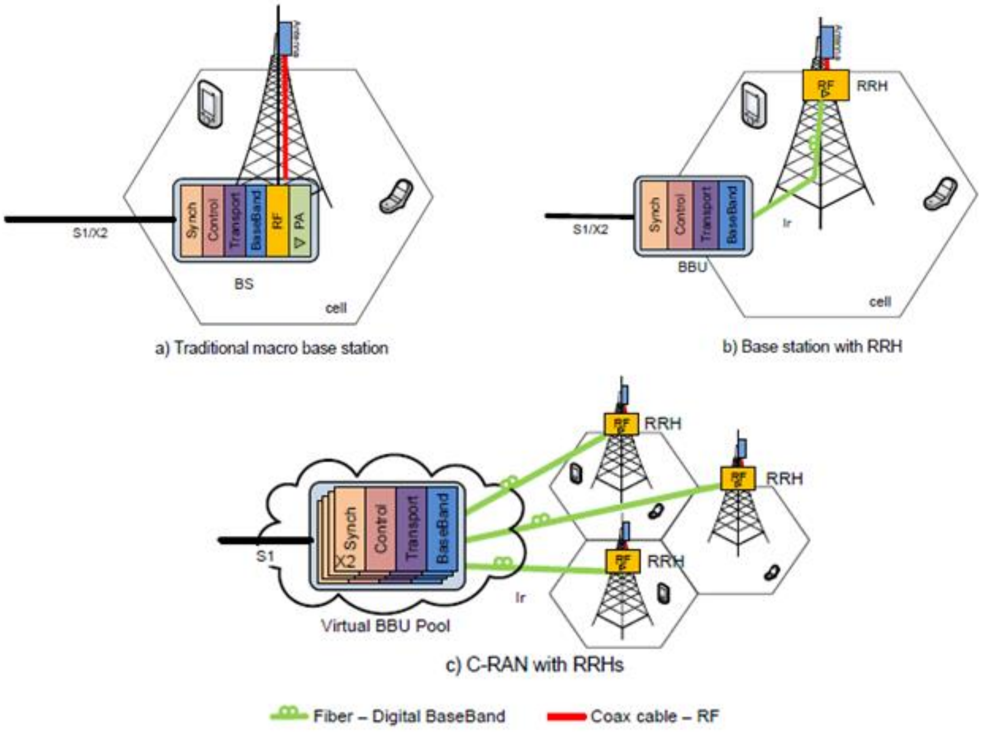


Figure 53. Traditional RAN versus Virtualized RAN [B]. From [39].

A traditional RAN approach must follow the next basic steps so that it is transformed gradually to a C-RAN approach:

1. Centralization of DU and cloud DU approach.
2. Pooling of DU and centralized resource sharing.
3. RAN Virtualization.

DU centralization aims to gather basic RAN functions that reside in the BBU of each site to one location, where a single BBU or a BBU stack shall serve multiple Radio Units. Pooling of DU provides the ability of load balancing between different DUs and the ability to handle traffic peaks in a centralized manner. RAN virtualization refers mainly to the virtualization of services offered.

The C-RAN portfolio can include various application scenarios not only related to macro cells but Pico, micro and WiFi as well as shown below. However, the scenarios and deployment cases that we shall present later in this chapter are focused on LTE technology.

C-RAN Key benefits

The key benefits from C-RAN can be summarized as follows [35], [36], [37]:

1. **OPEX and CAPEX savings:** It is clear that C-RAN architectures introduce OPEX and CAPEX savings, mainly due to the fact that BTS equipment rooms are eliminated, separate licensing is not need per site, site visits are less, power consumption is less, upgrades are less and deployed only centrally, while capacity and coverage planning can be deployed with a more efficient manner.
2. **Traffic Offloading, Quality of Service (QoS) centralized handling and centralized Traffic Load Balancing:** Through the C-RAN, core network traffic can be offloaded since back-haul traffic can be moved to one central node, core network traffic can be utilized more efficiently and latency to the end nodes is improved. Additionally, service quality can be handled in a central manner and different Quality of Service Profiles can be set.
3. **Efficient capacity and feature control handling:** Due to the centralization, signaling can be shared among different sites while scheduling and resource handling for UMTS and LTE sites can be deployed more efficiently since Channel Elements (CE for UMTS) and Physical Resource Blocks (PRBs for LTE) optimization per case can be deployed centrally. Additionally, RAN vendors have deployed a big variety of RAN features that can be handled per case from one single node (BBU Pool).
4. **Sustainable Environmental Power Savings – Green RAN**: With C-RAN, power needs for a base station are reduced and extra site resources such as air-conditioning or monitoring equipment costs in BTS equipment rooms are reduced as well.
5. **Interference handling and coverage increase**: Interference among different cells can be handled in a central manner, while higher density in RRUs can be provided.

C-RAN Use cases and RAN sharing

In the same way that a NFV deployment can be shared among different enterprises as referred in chapter 1, C-RAN can enforce RAN sharing deployments in the sense that two or more operators share the infrastructural part of the radio access network or even the spectrum. Network sharing has been 3GPP standardized and various deployments exist already in UK, Greece, France and other countries in order to gain CAPEX and OPEX reduction benefits and easier site acquisition [40], [37]. Various C-RAN use cases for RAN sharing exist that mainly utilize the use of DU pool however we shall refer to the two most frequent.

Radio sites sharing

Through this scenario the operators can share the DU pool infrastructure and keep the same site collocated sites, however interference issues might rise.

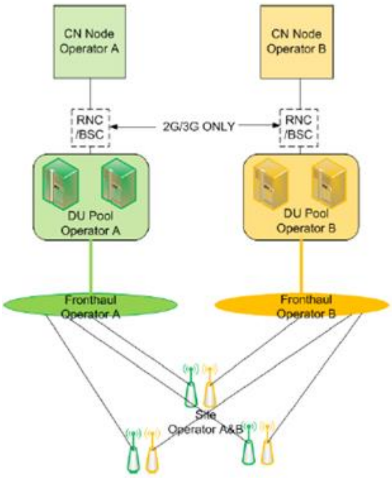


Figure 54. Radio sites sharing, From [37].

Fronthaul and Radio Units sharing

This scenario leads to more efficient transport utilization characteristics since the two operators share the radio units and fronthaul, but not the same spectrum and PLMN characteristics. The radio unit in this case must support wider bandwidth, and more DU pools must be supported if compared to the first use case [37].

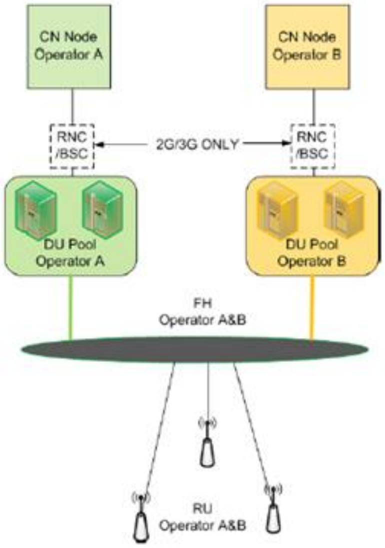


Figure 55. Radio Units and Fronthaul sharing. From [37].

4.2 C-RAN Architectures, deployment scenarios and key design concerns

C-RAN Architectures and design approaches

C-RAN is the solution to the deficiencies discussed above and the next generation approach to the standard BTS structures as we know them today, since virtualization is adopted in the RAN part. The possible architectures are initially based on the assignment and sharing of the functionalities of the BBU (Baseband Unit) and Remote Radio Head (RRH) (referred as Remote Radio Unit (RRU) as well).

There are two basic approaches: In the first approach full centralization is provided, since the baseband unit provides layer 1, layer 2 and layer 3 functionalities while in the second approach the RRH provides the BBU functionalities leading in partial centralization. It should be noticed that in the second solution the BBU still provides functions such as the interface point with the core network as well as main control and clock [37].

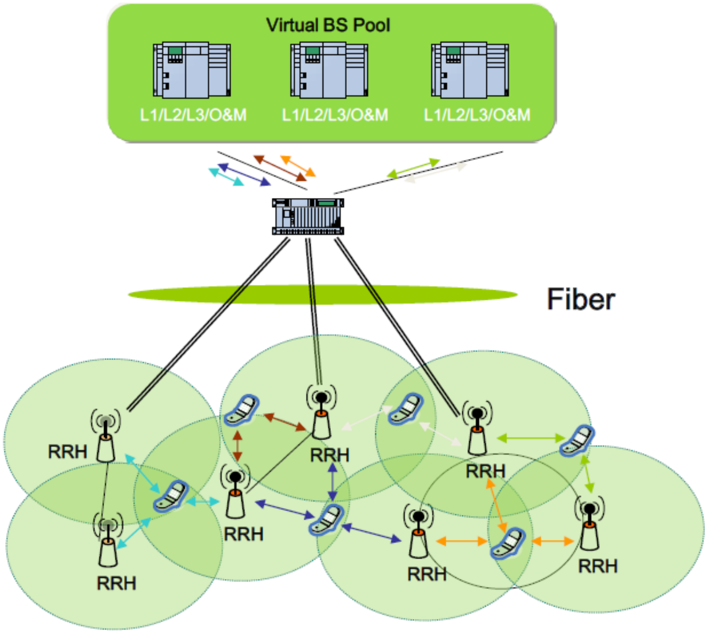


Figure 54. Full centralization C-RAN Solution. From [35].

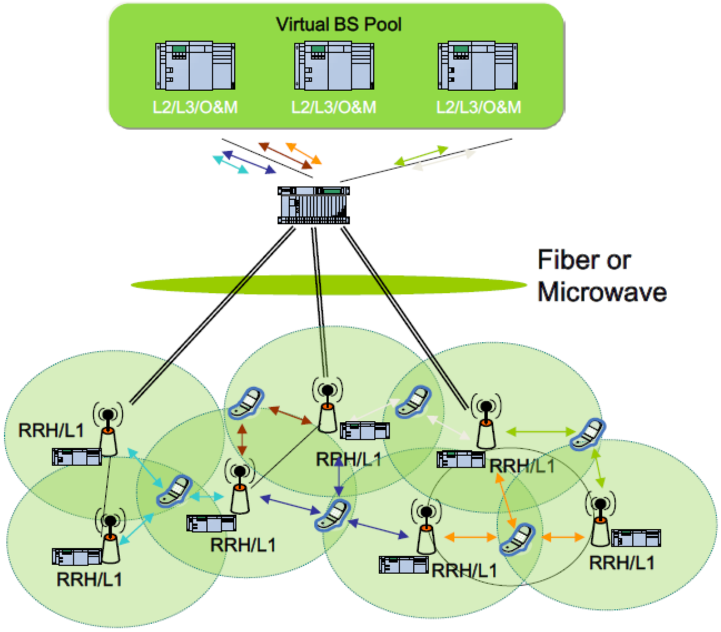


Figure 55. Partial centralization C-RAN solution. From [35].

With both architectures operators can expand the network by just adding more and more RRUs to the central BBU pool or reconfigure specific sites or cells that have increased capacity demands.

The traditional strong relation of the BBU with the RRU where the latter coexist in the same site does not exist anymore since a RRU might belong to different BBUs and the functionality of a “Virtual BTS” is introduced. This point correlates C-RAN with Network Function Virtualization (NFV).

A fully centralized C-RAN can provide better capacity and upgrade capabilities, since resource handling is easier through a central node and SDR (Software Defined Radio) can have a significant role in the network through only one central BBU, however the disadvantage of high bandwidth requirement exists.

On the other hand, partial centralization has the advantage of splitting network functionalities such as (L2, L3 scheduling) while bandwidth allocation between baseband units can be handled in a more efficient manner. The main disadvantage of this solution is that a small number of sites with minimum equipment cannot be avoided [35].

The following example regards a typical C-RAN deployment where a Front End interface (FE) is used to separate the communication synchronous traffic domain with the asynchronous traffic IT domain while a centralized BBU Pool handles the RRUs and according to the traffic load carrier processing can be assigned to different BBUs.

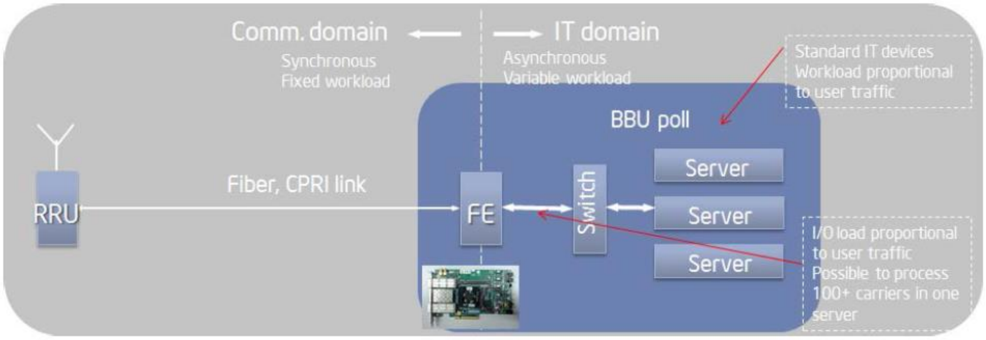


Figure 56. Front End interface role. From [37].

Antenna RF data from RRU are passed to the FE interface for processing through a CPRI link and after the cell level physical processing, data are passed to the BBU pool through a switch for user level processing. In the downlink channel the opposite procedure takes place. The Front End interface corresponds to interface II of the following figure.

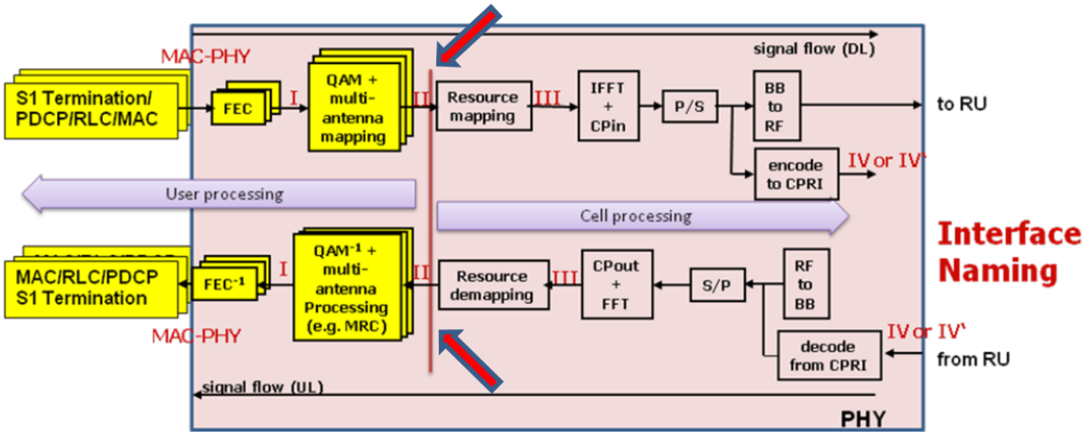


Figure 57. Interface II corresponds to Front End (FE) interface. From [37].

Since the basic target of C-RAN is to decouple and centralize basic functions of the BBU and Radio Unit, it is clear that the discussion on virtualization regards mainly the placement of interfaces and functions that in the traditional approaches reside either in the RRU or the BBU.

The key approach is to start migrating and investigating which interfaces can be transferred to the BBU and what can be the impact mainly in latency and performance operations if this transfer can be deployed in a BBU pool rationale.

The following figure indicates the case where the MAC interface is migrated to the central BBU while the PHY interface remains at the Radio Units, thus the MAC-PHY interface is split.

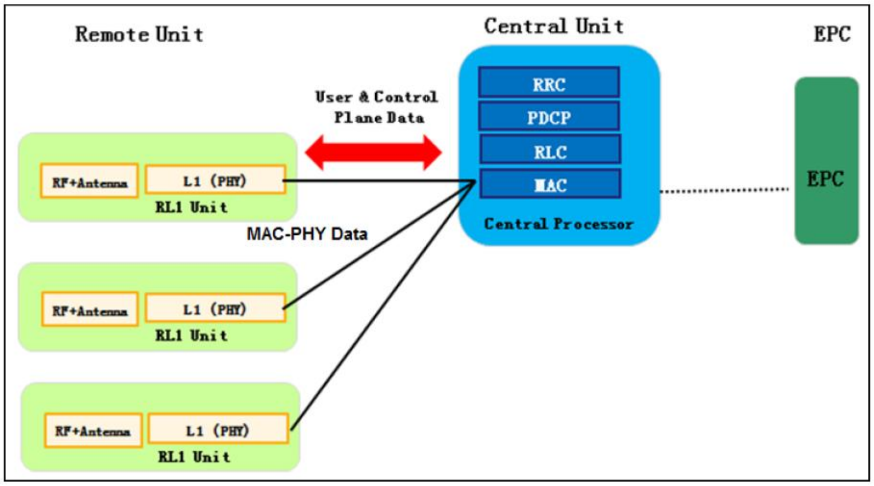
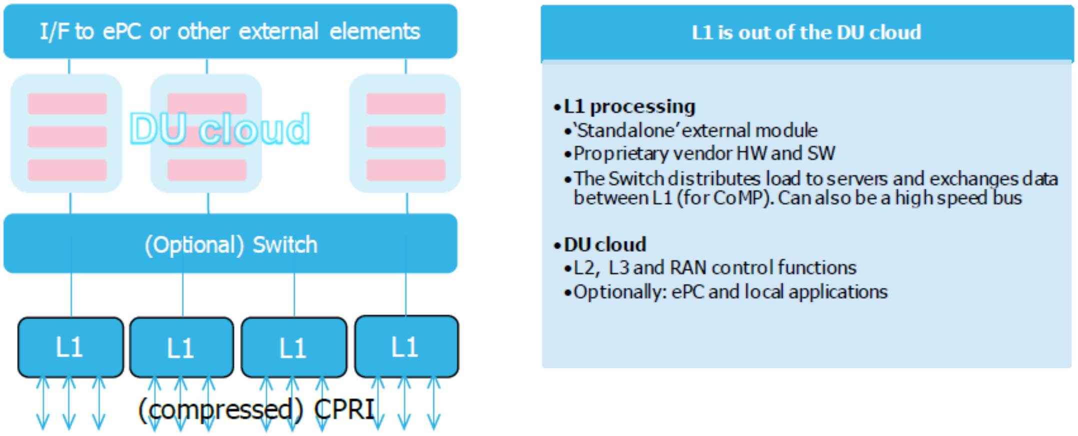


Figure 58. MAC interface functionality migration to the central BBU. From [37].

The rationale of starting the centralization from the decoupling of MAC-PHY interface is that the baseband part is a user processing traffic load dependent interface and can be separated from the cell processing part.

Layer 1 functions placement

As previously discussed, one of the most critical design decisions regards the placement of Layer 1 Functions (L1). In the following C-RAN deployment scenario the L2 and L3 functions are deployed in the DU pool while L1 functions are deployed out of the DU cloud in hardware that can support multi-vendor vendor proprietary protocols. Additionally, a switch provides the interconnection between the DU pool and the outer L1 hardware.



- **L1 processing**
  - 'Standalone' external module
  - Proprietary vendor HW and SW
  - The Switch distributes load to servers and exchanges data between L1 (for CoMP). Can also be a high speed bus
- **DU cloud**
  - L2, L3 and RAN control functions
  - Optionally: ePC and local applications

Figure 59. L1 out of the DU cloud. From [37].

However, further abstraction of functions is feasible since some L1 functions can be migrated to the DU cloud as well as shown below:

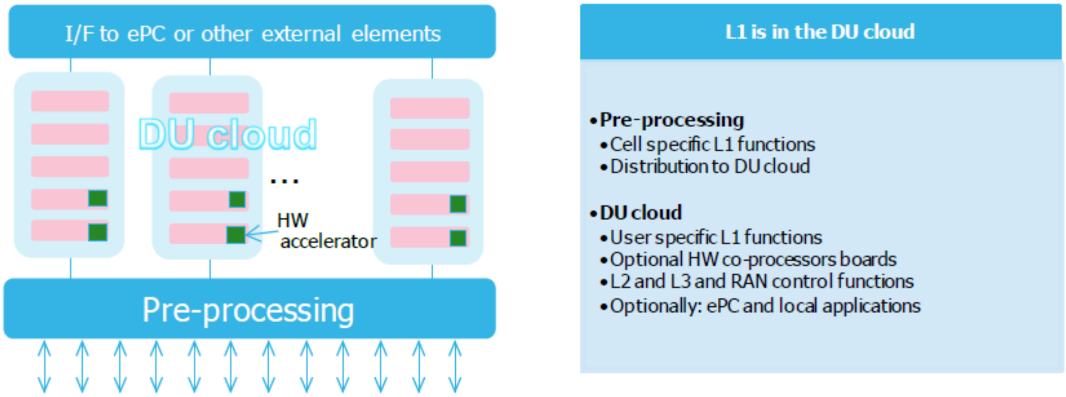


Figure 60. L1 in the DU cloud with hardware accelerators. From [37].

C-RAN Testing and deployments to operators

China Mobile in cooperation with IBM, ZTE, Huawei, Intel, Datang Mobile, France Telecom Beijing Research Center, Beijing University and China Science Institute, jointly developed a C-RAN prototype supporting GSM/TD-SCDMA/TD-LTE with the structure that is shown below [38]:

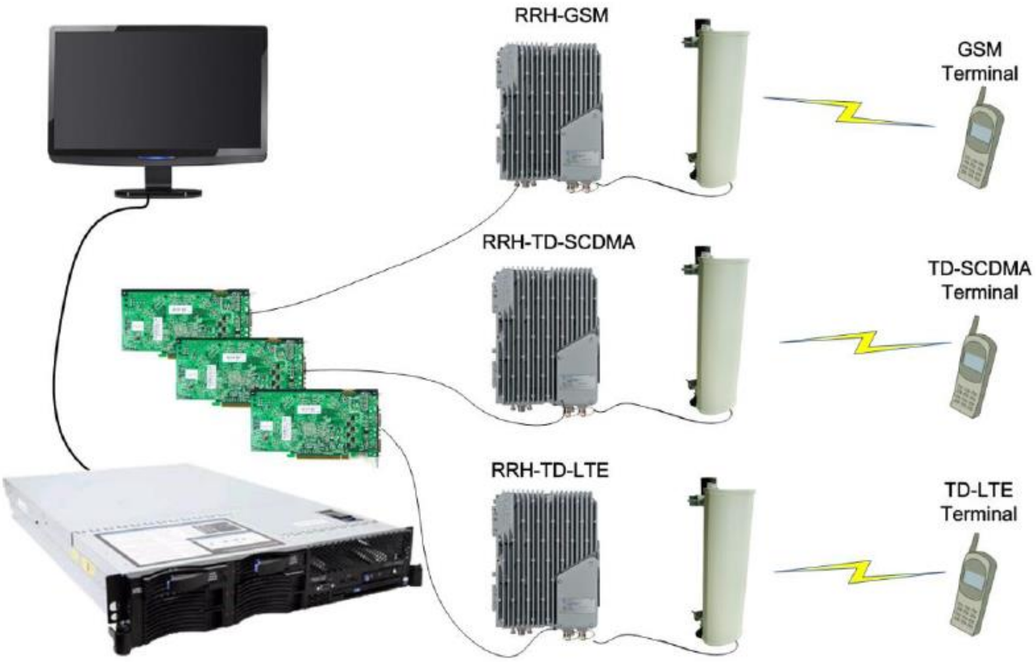


Figure 61. CHINA MOBILE C-RAN test setup. From [38].

A server is used to perform the BBU functions including Layer 1, 2 and 3 on GSM and TD-SCDMA, and Layer 1 processing on TD-LTE while the MAC scheduler was not deployed at the time the test was deployed leading to only one user testing. A CPRI interface card is used per technology leading to separate synchronization between the server and the RRUs since real time synchronization constraints exist for the RRUs.

The team managed to implement a software based 3GPP release 8 TD-LTE physical layer supporting 20Mhz bandwidth, 2x2 MIMO downlink, 1x2 SIMO uplink and 64QAM/15QAM/QPSK modulation with satisfactory timing and delay characteristics since the frames where processed with 1ms TTI (Transmission Time Interval). However, it should be noticed that the standard IT servers are not designed to meet the real time processing requirements needed for a standard RRU and real time traffic for the case of LTE, thus the basic functions where deployed through FPGA hardware based elements [38].

The main issue that rises for the case of LTE is that a hardware accelerator might be needed in order to handle the physical layer functions since the outcome from the testing and prototype is that at least 10 CPU cores are needed to process a case with 8-antennas 20MHz TD-LTE carrier physical layer processing. The role of a hardware accelerator can be shown below:

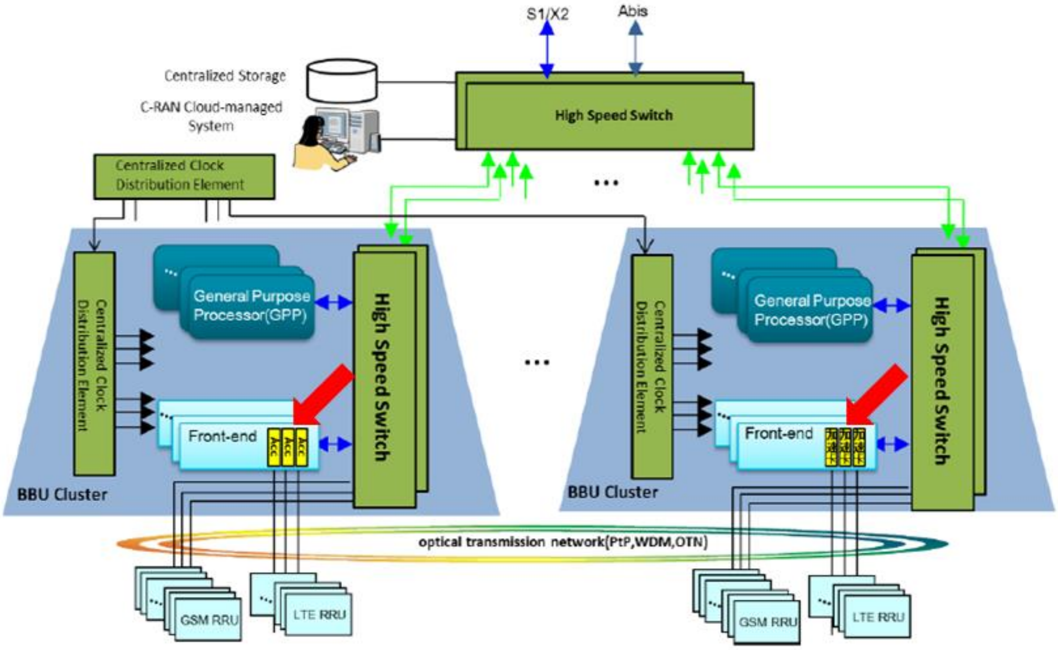


Figure 62. C-RAN architecture and hardware accelerator. From [35].

The next test performed in China Mobile was focused on the front haul transmission characteristics for the interconnection of LTE RRUs with a BBU pool and a max distance of 10km between them. According to the specific test the uplink CPRI signaling was passed via grey light to the (HSN-8100) multiplexer and to the hub multiplexer (HSN-8300) via WDM rings with colored light. The hub multiplexer reverts back the colored light to grey light and passes the signals to the BBU pool. The opposite process takes place for the downlink transmission [38], [39]. The next figure depicts the test structure.

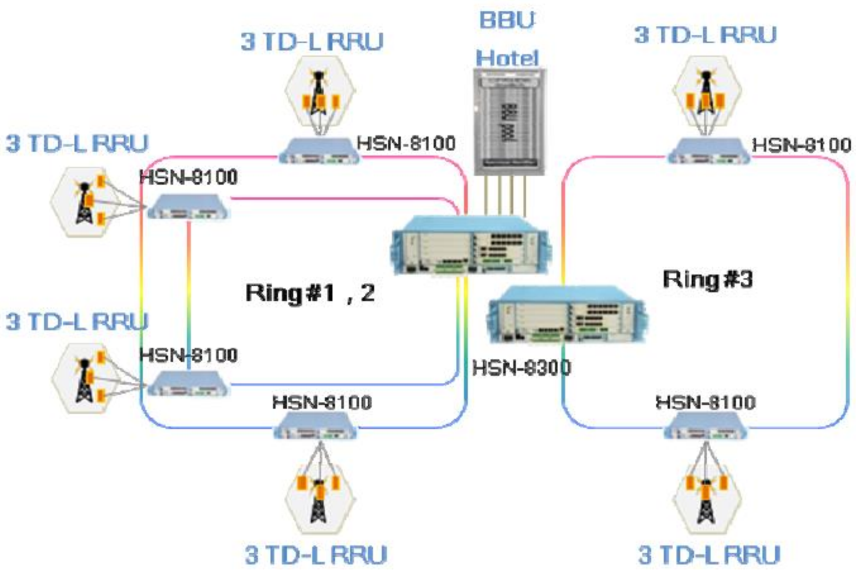


Figure 63. Test scenario of C-RAN with WDM rings. From [38].

C-RAN Deployment concerns

One of the basic challenges that rises from C-RAN is the front-haul interconnection of BBU and RRU through high speed Common Public Radio Interface (CPRI) signaling with focus on very low latency, jitter and satisfactory utilization characteristics. CPRI was first developed as an internal interface for interconnecting the NodeB/eNodeB internal interfaces or the BBU with the RRU. A typical CPRI interface is shown below:

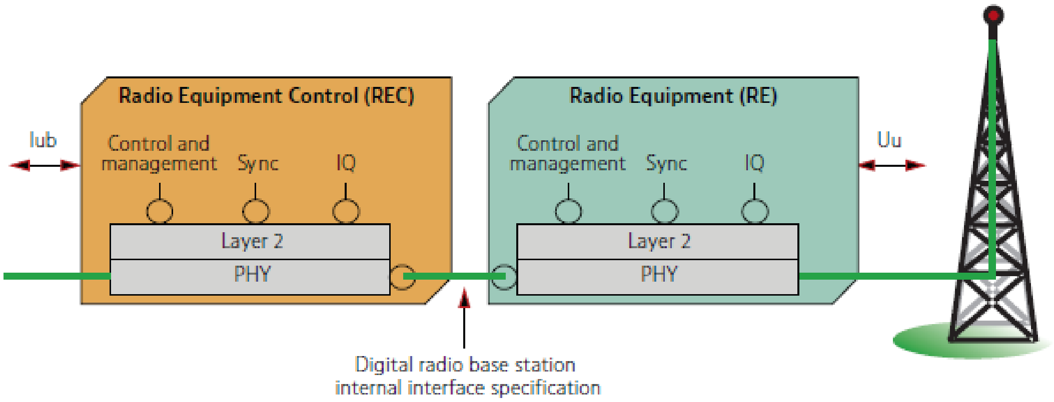


Figure 64. CPRI interface. From [40].

The basic approaches of BBU to RRU interconnection are the following:

1. CPRI over dark fiber.
2. CPRI over WDM-PON (Wavelength Division Multiplexing-Passive Optical Network).
3. CPRI over WDM (Wavelength Division Multiplexing).
4. CPRI over OTN (Optical Transport Network).

The first solution (CPRI over dark fiber) is not efficient due to the fact that no wavelength division multiplexing (WDM) technology was adopted thus leading to a non appropriate solution due to low utilization characteristics [38], [39], [40].

CPRI over WDM-PON is not yet a mature solution and CPRI over OTN introduces delay thus CPRI over WDM seems to be the most mature and efficient solution even if there is a clear cost increase in the network. Through the specific solution multiple CPRI signals (around 80 channels) can be transmitted over a single fiber, the adoption of a ring topology can provide redundancy up to 16 end nodes and their separation distance can be up to 20 kilometers. Furthermore, the delay must be below 120 to 200 micro minutes [38], [39], [40].

4.3 CAPEX and OPEX analysis for C-RAN

CAPEX and OPEX analysis of RAN investments prior to C-RAN

C-RAN and virtualization shall provide a cost effective solution to the standard CAPEX and OPEX expenditures since power, equipment, planning, resources and maintenance activities shall be optimized and a centralization approach shall be followed.

Power consumption savings

As RAN expands more and more, standalone BTS are introduced to the network, thus expenses related to power consumption are increased dramatically leading to higher OPEX and environmental impact.

The next figure shows that the BTS equipment is leading the power consumption in a single site, since inside the BTS room power is 50% consumed by the BTS (baseband, radio, transmission modules) while the rest regard the air-conditioning and supporting equipment [34], [35].

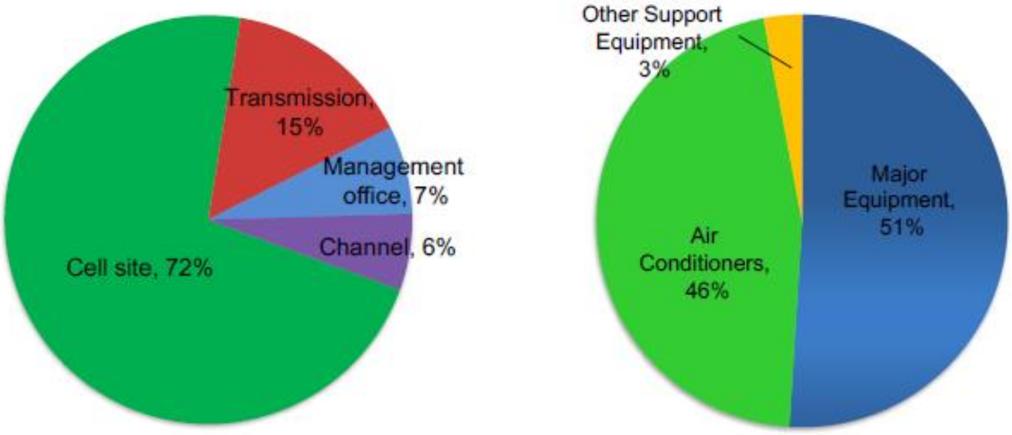


Figure 65. Base Station power consumption. From [35].

The obvious solution of reducing the number of sites is not applicable, since there shall be a big impact on capacity and coverage thus up to the time of speaking vendors developed special features for BTS power handling (Enabling/disabling on a demand basis), while renewable energy manufacturer companies provide turn-key solutions. However, the latter appear to have high cost and full autonomy cannot be ensured 100% leading to the need for standard power lines as well [34], [35].

Due to the above, it is clear that mobile operators and service providers must gradually migrate to alternative solutions. C-RAN and virtualization can be positioned as a solution to the power consumption problem.

CAPEX and OPEX Analysis and their impact to TCO (Total Cost of Ownership)

Cost analysis in major operators in China has shown that almost 80% of the CAPEX is spent on RAN investments and site related activities including site reconfiguration, construction, rent, acquisition, planning or even civil works. Additionally, as new technologies are gradually released, costs keep increasing since a base station must be re-equipped with modules supporting the newer access technologies. This was clear during the first 3G deployments where sites operating only in 2G had to be reconfigured so that 3G is supported as well and continues with the LTE network expansions of the operators [34], [35].

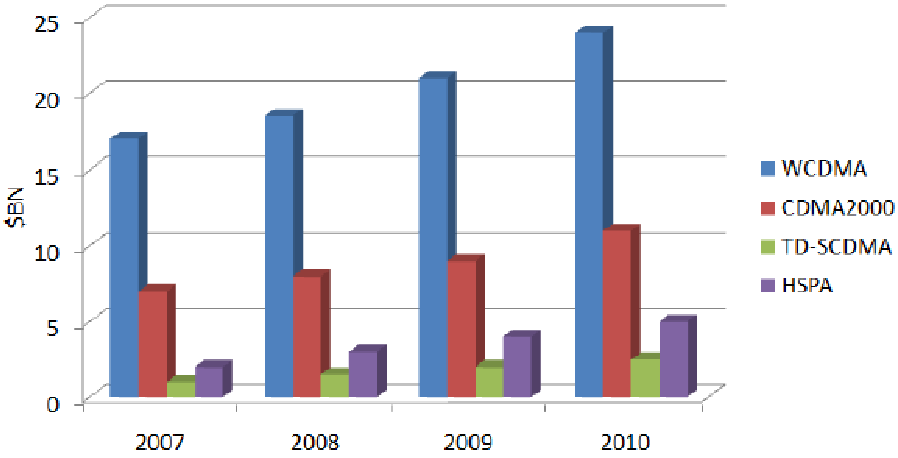


Figure 66. CAPEX per technology. From [35].

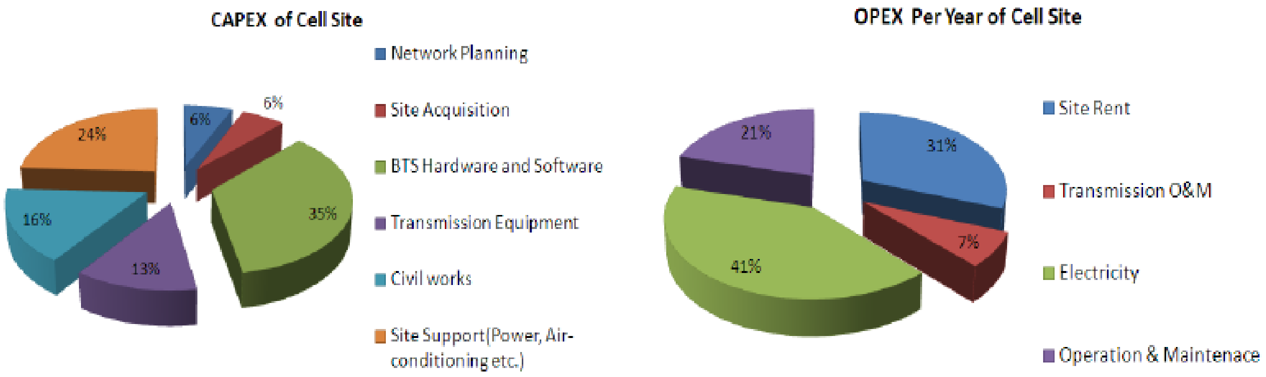


Figure 67. CAPEX and OPEX analysis for cell site. From [35].

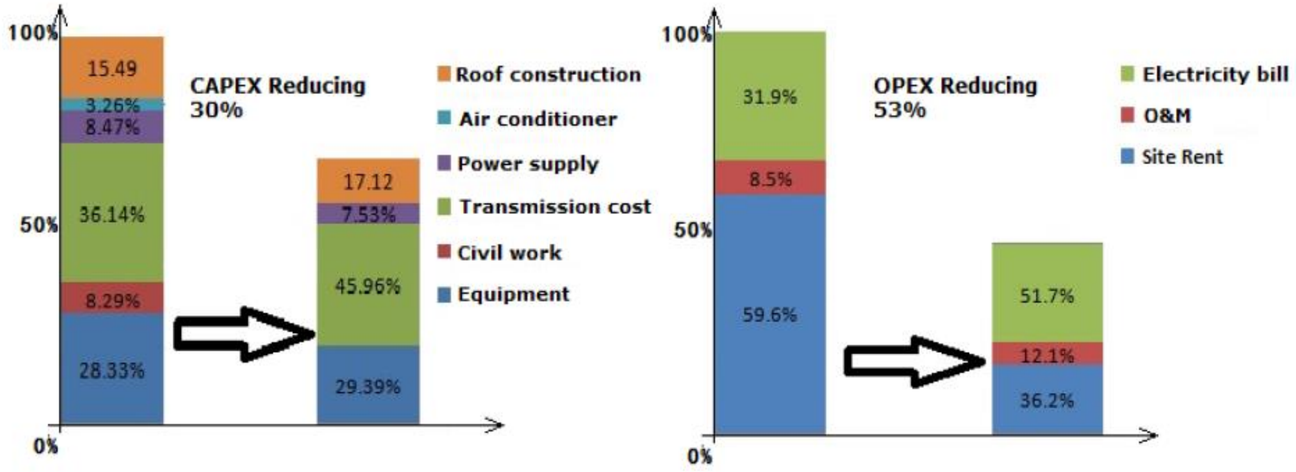


Figure 68. Traditional RAN versus C-RAN CAPEX and OPEX savings. From [35].

OPEX has a higher impact in the TCO since studies have shown that for a 7-year deployment OPEX involves more than 60% of the TCO while the rest goes to CAPEX.

TCO decrease might be achieved by reducing the number of RAN sites, however as already pre-mentioned such an action might lead to poor coverage and capacity degradation issues.

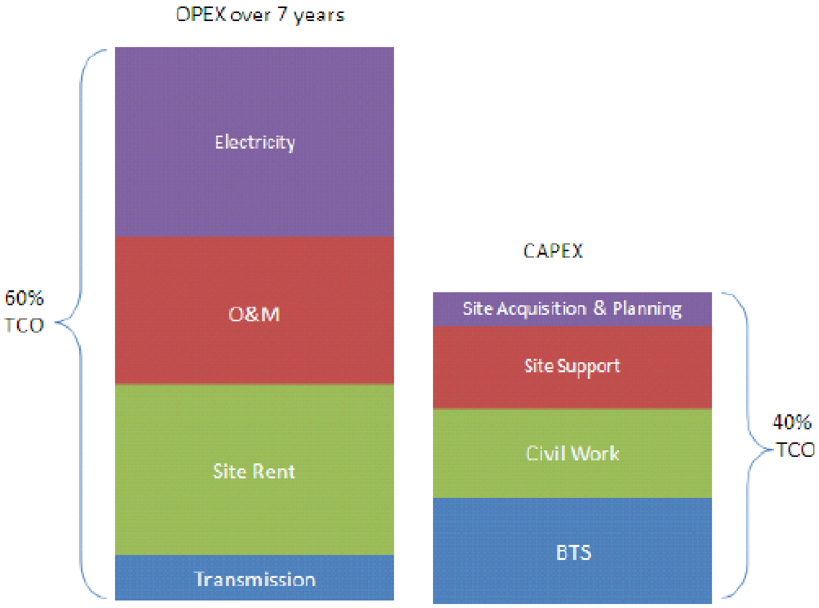


Figure 69. TCO analysis of a site. From [35].

RAN architecture	Base Station equipment	Air conditioning	Switching Supply	Storage Battery	Transmission System	Total
Traditional	0.65 KW	2.0 KW	0.2 KW	0.2 KW	0.2 KW	3.45KW
C-RAN	0.55KW	0	0.2KW	0.10KW	0	0.85KW

Table 6. Power consumption of Traditional RAN versus C-RAN. From [35].

4.4 C-RAN and C-SON

We believe that if the centralized virtualization approach of baseband units through C-RAN is combined with C-SON (Centralized Self Optimizing / Organizing Networks) even better performance benefits can come. SON is a software solution that leads to the automation of network optimization, planning and parameters configuration processes through the collection of real network KPIs and performance counters. Typically, there are two approaches on SON systems deployment. The centralized and the decentralized where SON functions run at the end nodes, however we shall refer to the centralized one since it is the one most commonly deployed at this time.

In the centralized solution SON runs as a service (SONaaS) on standard servers (Usually a servers cluster depending on the network scale) hosting the SON applications and the database of the system.

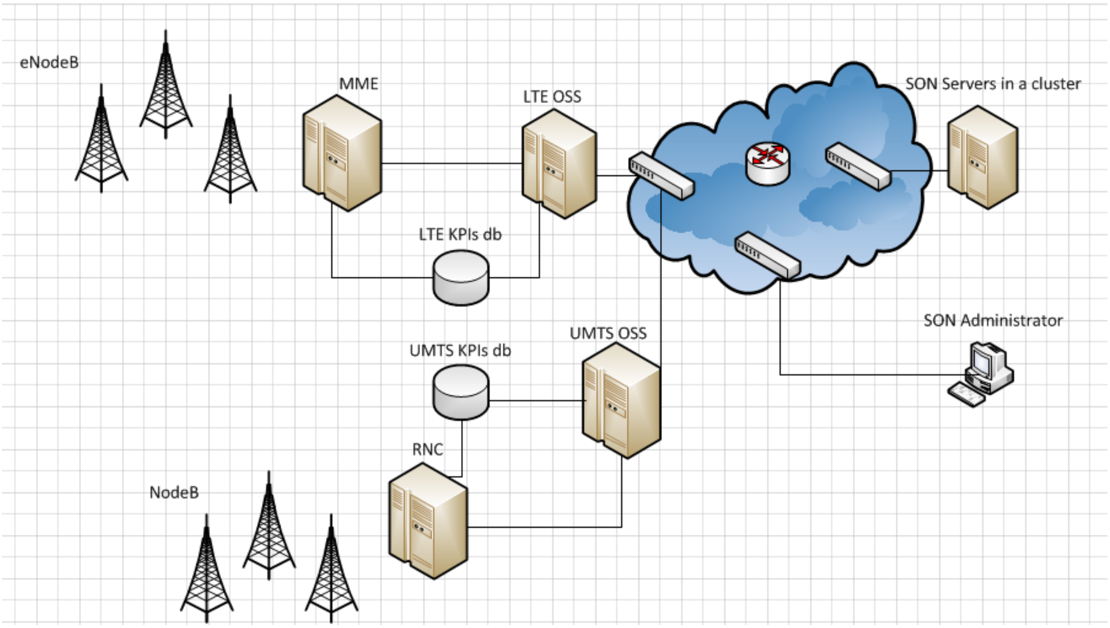


Figure 70. Typical Centralized SON deployment.

SON system interacts with the OSS in the same rationale that at this time a user configures through the OSS one or more sites through scripts. More specifically the SON system collects network KPIs and according to the algorithm used for each application creates an executable plan for the OSS. The following figure depicts the SON framework operation:

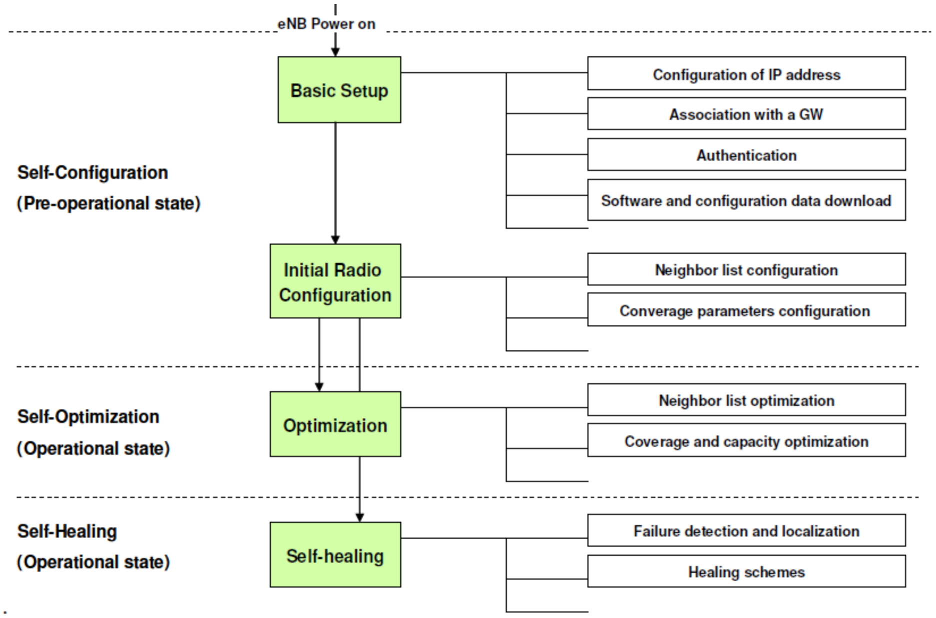


Figure 71. Basic SON activities. From [41].

Various applications for all technologies (GSM, UMTS, LTE) covering neighbor relation optimization, Load Balancing, Interference Management, PCI and Scrambling Code optimization etc. exist and it should be noticed that all of them are based on the KPIs collection.

With the rise of vEPC and especially C-RAN, it is possible that additional scenarios and use cases for C-SON systems might appear. The rationale behind this proposal is that C-SON systems shall be able to monitor in a more centralized manner metrics and parameters related to the Baseband Units hardware related operation and possibly new SON applications focusing on hardware management can be developed. An example might be the efficient channel elements assignment through C-SON applications which is a well-known issue for operators.

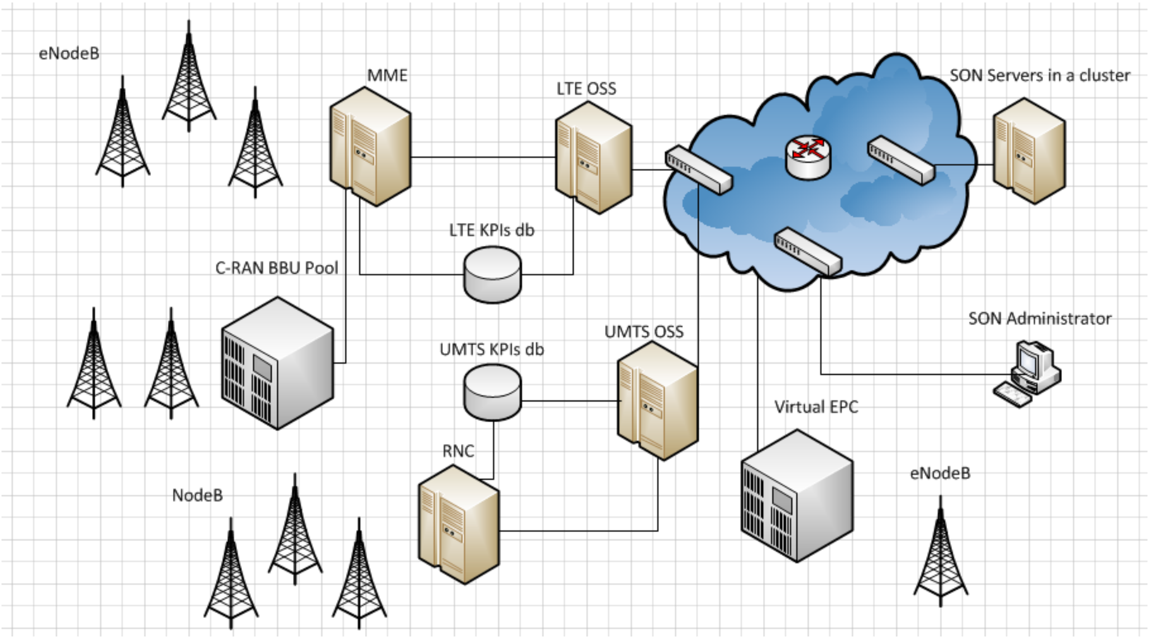


Figure 72. C-SON operation in parallel with vEPC and C-RAN BBU Pool.

4.5 C-RAN and NFV for 5G

5G is the next step after 4G - LTE networks aiming to provide the following performance characteristics if compared to LTE [45], [46], [47]:

- Ultra High Speed and Low Latency by achieving 10Gb/s peak data rates and less than 5 ms end to end latency.
- 99.999% Reliability.
- Supporting network connectivity for vehicles moving with 500km/h.
- Capabillity to be deployed and provide service in 90 minutes.
- Higher energy effirciency by lowering the consumption needs to 10% of current consumption of 4G-LTE networks.
- Supporting high user density areas (Up to 1 M devices connected per square meter).

- Support of more than 1 Trillion connected IoT devices.
- Accuracy of outdoor terminal location less than 1 meter.

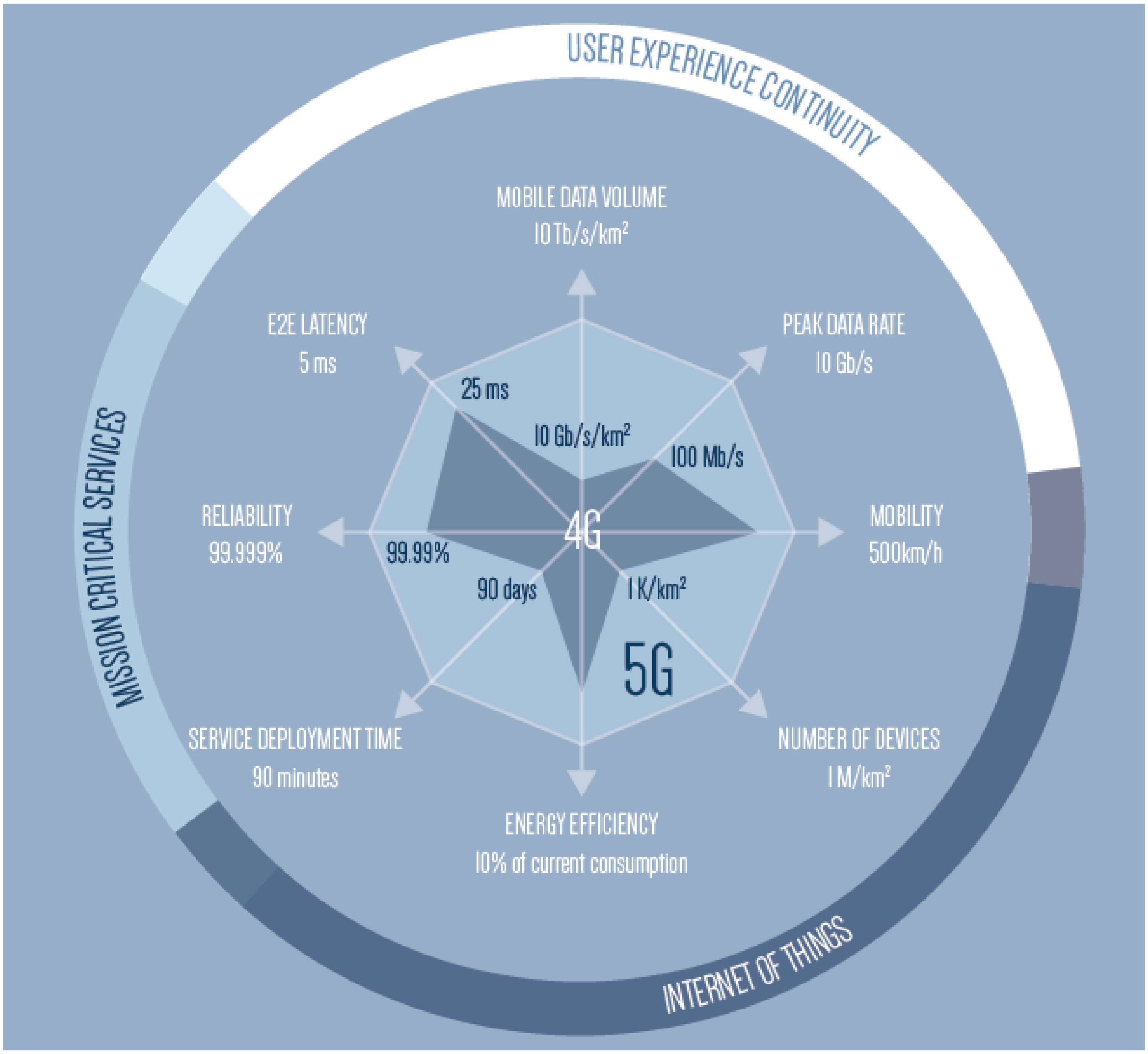


Figure 73. Radar diagram of 5G versus 4G. From [45].

Even if there is no standardization for 5G yet and the first commercial 5G networks are expected after 2020, the first trial architectures that appeared are software driven while virtualization in radio access part through C-RAN and virtualization in the core access network are the basic key components. C-SON and Big Data Analytics technologies are also key contributor since they shall provide the expected network intelligence to 5G networks [45], [46], [47].

As an example SK Telecom (South Korea) which is one of the worldwide leading operators trialing 5G networks has already structured an architecture model based on NFV in the mobile core and C-RAN in the mobile access part as shown below:

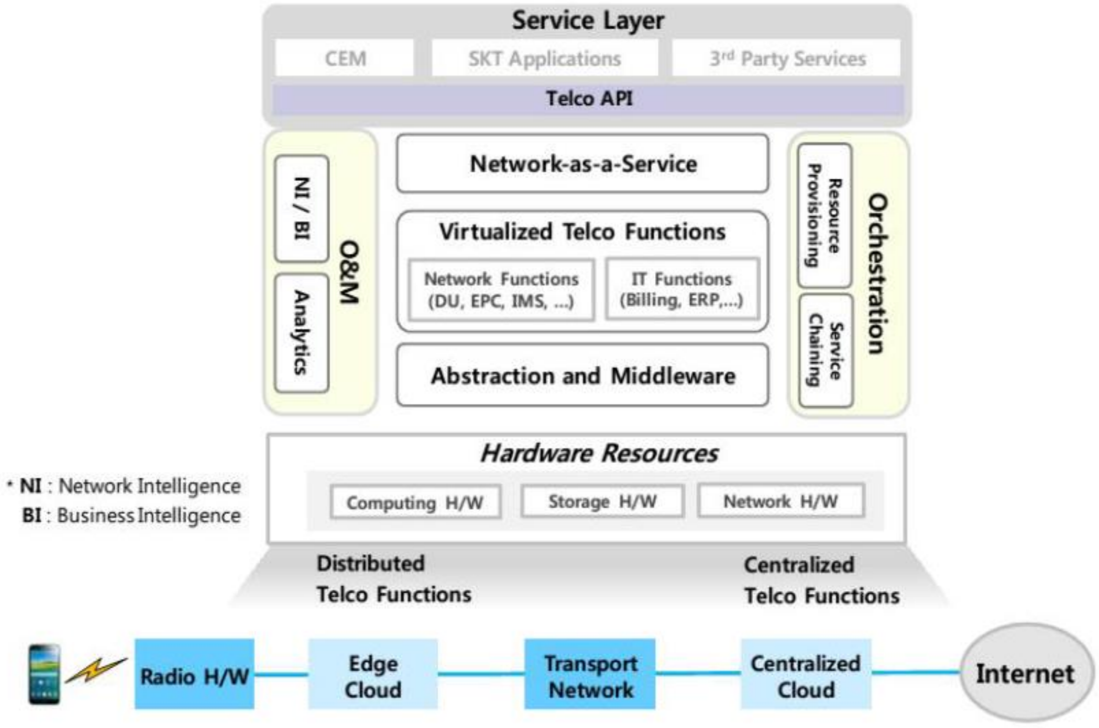


Figure 74. Indicative 5G network architecture based on NFV. From [46].

Conclusion

Network Function Virtualization (NFV) main target is to decouple the dependency between hardware and software and enable the development of network solutions by using multi-vendor hardware that can effectively cooperate with multi-vendor software platforms. NFV solutions are ideal for telecommunication operators and service providers since it is clear that the data traffic explosion cannot be handled with the standard traditional network infrastructural architecture and at the same time high OPEX and CAPEX savings can be gained.

Due to multiple research projects and basically due to the information technology and telecommunications industry demands, NFV together with IoT (Internet of Things), SDN (Software Defined Networking), SON (Self-Organized/Optimized Networks) and 5G are predicted to shape the future with NFV being the key enabler for the deployment of the latter technologies.

Moreover, NFV is tightly coupled with SDN with the latter having less stringent requirements as far as real time processing is concerned, however their combination can lead to flexibility with the development of non-proprietary communication protocols and automation of network functions. The main difference between the two is that NFV aim at the decoupling of software with hardware while SDN aims at the separation of packets and interfaces from network control plane.

It is important to understand the key design considerations for NFV deployments which include:

1. **Performance requirements:** The same performance with the existing network services running on dedicated hardware must be ensured by keeping tight latency and high bandwidth profiles.
2. **Security:** Current security policies applied to network services must be able to operate at the same way in NFVs.
3. **Availability and reliability, Disaster recovery compliance:** In case of a failover redundancy solutions must exist, keeping the same SLA characteristics as standard equipment.
4. **Heterogeneous Support:** Multivendor approach and compatibility through the development of open interfaces.
5. **Legacy systems support:** Hybrid solutions must be supported including current network architectures, till NFV architectures become mature.

As far as the fixed telecommunications domain is concerned, standard switching, routing and load balancing functionalities can be replaced by NFV and for the mobile domain end to end solutions including LTE 4G EPC (Evolved Packet Core) and the RAN (Radio Access Network) can be deployed. Vendors have already set high interest on the development of Virtual EPC architectures through vEPC, while the Radio Access Network virtualization through C-RAN is possibly the next big thing after LTE/4G at least for the radio access technology domain through the approach of aggregating multiple radio units into one baseband processing pool.

The key point in Evolved Packet Core virtualization is to select among full or partial virtualization of the control plane and user plane. As an example Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW) can be separately virtualized based on the fact that traffic bottlenecks occur in the user plane, thus leading to separate control per customer.

Serving Gateway (S-GW) and Packet Data Network Gateway (P-GW) are designed for achieving high throughput of user packets, while MME/SGSNs have fewer requirements as far as throughput and capacity is concerned but stricter requirements in terms of processing latency. Due to this the control plane entities are more suitable for being virtualized under the aspect of taking of advantage of the high availability of computing resources in the Cloud. However, under specific use cases user plane functions can be virtualized as well.

The key rationale in C-RAN solution is that through the centralized Base Band (BB) pool processing a set of base station equipment and management costs can be reduced, since higher spectrum efficiency and interference management can be provided centrally, while its Cloud Computing capabilities based on virtualization enable several management features including resources allocation, power consumption and utilization. Traffic offloading, power savings and feature control can be handled centrally in a separation distance of more than 20km between the BBU pool and the radio units through high capacity fronthauling optical or microwave links.

The two basic approaches include the full centralization where the baseband unit provides layer 1, layer 2 and layer 3 functionalities while in the second approach the RRH provides some of the BBU functionalities.

A fully centralized C-RAN can provide better capacity however the disadvantage of high bandwidth requirement exists. Partial centralization has the advantage of splitting network functionalities such as L2, L3 scheduling and bandwidth allocation between baseband units can be handled in a more efficient manner, however the disadvantage of having sites with minimum hardware equipment supporting the pre-mentioned functionalities exists. It can said that the most important design approach concerns the placement of interfaces and functions that in the traditional approaches reside either in the RRU or the BBU, while the key point is to evaluate the impact on latency and performance according to the policy followed.

One of the basic challenges that rises from C-RAN is the front-haul interconnection of BBU and RRU through high speed Common Public Radio Interface (CPRI) signaling with focus on very low latency, jitter and satisfactory utilization characteristics. Based on tests we can conclude that CPRI over WDM is the most mature solution that can satisfy the latency and utilization characteristics.

Additionally C-RAN can be combined with C-SON (Centralized Self Optimizing / Organizing Networks) bringing even better performance benefits especially if baseband units centralized functions handling provided by the BBU pool can be embedded to C-SON software applications and algorithms.

Standards groups, operators and vendors are at this time working on NFV deployments including core and radio access network however it cannot be regarded as a fully mature technology. Some NFV solutions regarding mainly the core network are already developed by vendors, however the majority of operators worldwide are still at the stage of evaluation of such deployments since multiple PoCs (Proof of Concept) projects are ongoing. Issues regarding the evaluation of the network functions that can be fully transformed to NFVs, the placement of network functions (end nodes or data centers), interoperability between different NFVs and underlying hardware architecture (NFVI) are currently being examined.

Finally NFV applied in the mobile core and radio access is going to shape the basic 5G networks architectures and consists the key enabler for their development as recent trials confirm.

References

Chapter 1 references

[1] Wen, Tiwary, Le-Ngoc, "Wireless Virtualization", 2013, ISBN: 978-3-319-01290-2.

[2] Rashid Mijumbi, Joan Serrat, Juan-Luis Gorricho, Niels Bouten, Filip De Turck, Raouf Boutaba, "Network Function Virtualization: State-of-the-art and Research Challenges", IEEE Communications Surveys and Tutorials, 25-09-2015.

[3] ETSI Industry Specification Group (ISG) NFV, "ETSI, OpenStack Liason Statement: NFV Requirements," https://wiki.openstack.org , September 2014, NFV (14) 000154r2.

[4] Open Networking Foundation (ONF), "SDN Architecture. Issue 1. ONF-TR-502. https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical reports/ TR SDN ARCH 1.0 06062014.pdf, June 2014.

[5] Pierre Lynch, Michael Haugh, Liza Kurtz, Joseph Zeto, "Demystifying NFV in carrier networks", IXIA 2014.

[6] Vikram Nair, Vinod Kumar Gupta, "Software defined network and network functions virtualization", ARICENT 2014.

[7] ETSI GS NFV 002 V1.2.1 (2014-12), "Network Functions Virtualization (NFV), Architectural Framework".

[8] ETSI GS NFV 001 V1.1.1 (2013-10), "Network Functions Virtualization (NFV) Use Cases".

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

[10] HP Open-NFV http://h71028.www7.hp.com/enterprise/downloads/services/4AA51044ENW.pdf

[11] T-Nova project, "Network Functions as a Service over Virtualized Infrastructures", Deliverable 2.21, "Overall System Architecture and Interfaces".

[12] http://www.mobile-cloud-networking.eu/

Chapter 2 references

[13] Jeffrey Bannister, Paul Mather, Sebastian Coope, 'Convergence Technologies for 3G Networks IP, UMTS, EGPRS and ATM' Wiley 2004.

[14] Juha Kurhonen, 'Introduction to 3G Mobile Communications' Second Edition, 2003 ARTECH HOUSE, INC.

[15] http://tektronix.com/.

[16] 1G-2G-3G-4G Evolution, Qualcomm white paper “The Evolution of Mobile Technologies”, June 2014.

[17] CISCO LTE Design guide, “LTE Design and Deployment Strategies”, http://www.cisco.com/.

[18] Stefania Sesia (*ST-Ericsson, France)*, Issam Toufik (*ETSI, France)*, Matthew Baker (*Alcatel-Lucent)*, LTE – The UMTS Long Term Evolution-From Theory To Practice, Second Edition (2011).

[19] ALCATEL-LUCENT Strategic White Paper, “The LTE Network Architecture, a comprehensive tutorial”, 2009.

[20] 3GPP TS 23.401 V8.0.0 (2007-12). GPRS enhancements for E-UTRAN access release 8.

[21] CISCO White paper, “Top 10 Considerations for a Successful 4G LTE Evolved Packet Core Deployment”, 2013. http://www.cisco.com/

[22] ALCATEL-LUCENT Strategic White Paper, “Introduction to Evolved Packet Core”. 2009.

[23] https://www.nsnam.org/docs/models/html/lte-design.html#fig-epc-topology

[24] Mark Grayson, Kevin Shatzkamer, Scott Wainner, “IP Design for Mobile Networks” CISCO Press 2009.

Chapter 3 references

[25] British Telecom Research and Technology, “Why are operators interested in Virtualization and what does this mean for KPIs?”

[26] MobileCloud Networking Consortium 2012 – 2015, D4.2 Deliverable, “First Mobile Network Cloud Software Components”. 8th May 2014.

[27] MobileCloud Networking Consortium 2012 – 2015, D2.2 Deliverable, “Overall Architecture Definition, Release 1”. 30th October 2013.

[28] https://www.ibm.com/support/knowledgecenter/SSSHRK_4.2.0/itnm/ip/wip/overview/.

[29] https://www.nsnam.org/

[30] EATC CISCO Quantum Functionality test report, http://www.eantc.de/.

[31] http://www.cisco.com/c/en/us/solutions/service-provider/virtualized-packet-core/index.html

[32] ALCATEL-LUCENT Strategic White Paper, “The journey to Packet Core Virtualization, Evolving the Packet Core to an NFV/SDN Architecture”.

[33] http://www.nec.com/en/global/solutions/tcs/vepc/

Chapter 4 references

[34] ACG Research, Total Cost of Ownership Study Virtualizing the Mobile Core, July 2015.

[35] China Mobile, “C-RAN, The Road Towards Green RAN “White Paper Version 3.0, December 2013).

[36] NGMN (Next Generation Mobile Networks) Alliance, “Suggestions on potential solutions to C-RAN”, Version 4.0, January 2013.

[37] NGMN (Next Generation Mobile Networks) Alliance, “Further study on critical C-RAN technologies”, Version 1.0, March 2015.

[38] Sheng Liu, Ran Duan, Jinri Huang, Yunbo Li, Lei Wang, and Han Li, China Mobile Research Institute, “Cloud-radio access network with the aid of enabling front-haul technologies”.

[39] Aleksandra Checko, Henrik L. Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, Michael S. Bergery and Lars Dittmann, “Cloud RAN for Mobile Networks - a Technology Overview” IEEE Communications survey & tutorials, 2014.

[40] VIAVI White paper, “Cloud RAN deployment with CPRI Fronthaul technology”.

[41] Sujuan Feng, Eiko Seidel, NOMOR Research, Germany, “Self-Organizing Networks (SON) in 3GPP Long Term Evolution”, 2008.

[42] ERICSSON mobility report, https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf , June 2016.

[43] http://www.huawei.com/en/news/2016/2/Full-Outdoor-WDM-4G-C-RAN.

[44] https://newsroom.cisco.com/press-release-content?articleId=1753087.

[45] ‘5G Vision, the 5G Infrastructure Public Private Partnership, The next generation of communication networks and services’. https://5g-ppp.eu/.

[46] SK Telecom 5G White Paper, ‘SK Telecom’s View on 5G Vision, Architecture, Technology and Spectrum’, Network Technology R&D Research Center, SK Telecom, 2014.

[47] ERICSSON White paper, ‘5G Systems’, January 2015.

List of Acronyms

1	AAA	Authentication, Authorization and Accounting
2	API	Application Programming Interface
3	APN	Access Point Names
4	AS	Access Stratum
5	BBU	Baseband Unit
6	CAPEX	Capital Expenditure
7	CE	Channel Element
8	COTS	commercial off-the-shelf
9	CPRI	Common Public Radio Interface
10	C-RAN	Cloud-RAN or Centralized RAN
11	C-SON	Centralized Self Optimized or Organized Networks
12	EPC	Evolved Packet Core
13	EPCaaS	EPC as a service
14	EPCaaS SM	The EPCaaS Service Manager
15	EPCaaS SO	The EPCaaS Service Orchestrator
16	G-MSC	Gateway Mobile Switching Center
17	GPRS	General Packet Radio Service
18	GSM	Global System for Mobile Communications
19	HSS	Home Subscriber Server
20	IMS	IP Multimedia System
21	IoT	Internet of Things
23	KPI	Key performance Indicator
24	LTE	Long-Term Evolution
25	M2M	Mobile to Mobile
26	MAC	Medium Access Control
27	MANO	NFV Management and Orchestration module
28	MCN	Mobile Cloud Networking
29	MIMO	Multiple Input Multiple Output
30	MME	Mobility Management Entity
31	MOBaaS	The Mobility and Bandwidth Prediction as a Service
32	MSC	Mobile Switching Center
33	MVNO	Mobile Virtual Network Operators
34	NAS	Non Access Stratum
35	NAT	Network Address Translation
36	NFV	Network Function Virtualization
37	NFVI	NFV Infrastructure
38	OPEX	Operating Expense
39	OSS/BSS	Operations and Business Support Systems

40	OTN	Optical Transport Network
41	PCC	Policy and Charging Control
42	PCI	Physical Cell ID
43	PCRF	Policy Control and Charging Rules Function PCRF
44	P-GW	PDN Gateway
45	PoC	Proof of Concept
46	PON	Passive Optical Netwrok
47	PRBs	Physical Resource Blocks
48	QAM	Quadrature Amplitude Modulation
49	QCI	Quality Class Identifier
50	QoS	Quality of Service
51	QPSK	Quadrature Phase Shift Keying
52	RAN	Radio Access Network
53	RNC	Radio Network Controller
54	RRH	Remote Radio head
55	RRM	Radio resource management
56	RRU	Remote Radio Unit
57	SAE	System Architecture Evolution
58	SDF	Service Data Flow
59	SDN	Software Defined Networks
60	SDR	Software Defined Radio
61	S-GW	Serving Gateway
62	SI	Service Instance
63	SIMO	Single Input Multiple Output
64	SINR	Signal to Interference Ratio
65	SM	Service Manager
66	SO	Service Orchestrator
67	SON	Self Organized or Self Optimized Networks
68	TCO	Total Cost of Ownership
69	TD-SCDMA	Time Division Synchronous Code Division Multiple Access
70	TTI	Transmission Time Interval
71	UML	Unified Modelling Language
72	UMTS	Universal Mobile Telecommunications System
73	UTRAN	UMTS Radio Access Network
74	VM	Virtual Machines
75	VNF-FG	VNF Forwarding Graph
76	VoIP	Voice Over IP
77	Vo-LTE	Voice Over LTE
78	WDM	Wavelength Division Multiplexing

June 2016, Athens

Athens University of Economics and Business

MSc in Information Systems (I.S)

Supervisor: George Xylomenos

Andreas G. Papidas