Project Number:	**CELTIC / CP7-011**
Project Title:	Mobile Networks Evolution for Individual Communications Experience – MEVICO
Document Type:	D (Internal)

Document Identifier:	D 1.2
Document Title:	**Architecture Design Release 2 Documentation**
Source Activity:	WP 1
Main Editor:	Ivan Froger – Didier Becam
Authors:	László Bokor, Khadija Daoud Triki, Zoltán Faigl, Philippe Herbelin, Jiejin Hou, Benoit Lemoine, Tero Lötjönen, Edgardo Montes de Oca, Pascal Pagnoux, Jose Costa-Requena, Tapio Suihko
Status / Version:	Released 1
Date Last changes:	12.21.11
File Name:	MEVICO D1.2 Architecture Design v1.2

Abstract:	Based on the evolution of traffic and network usage, MEVICO proposes an enhanced mobile architecture for the LTE (Long Term Evolution) and LTE-Advanced. The architectural requirements related to different aspects (e.g. mobility, scalability, security…) are identified. These requirements allow identifying the related architecture challenges (network, applications, services…). MEVICO proposes technology solutions which address these challenges.

Keywords:	Mobile Broadband, traffic evolution, traffic properties, scenarios, use cases

Document History:	
10.15.2010	Document creation
07.22.2011	V1.1 approved at the WP1 Level
12.21.2011	V1.2

Table of Contents

Table of figures

Authors

Partner	Name	Phone / Fax / e-mail	
AALTO	Jose Costa-Requena		
		Phone:	+358 9 470 26257
		e-mail:	jose.costa@aalto.fi
France Telecom	Ivan Froger		
		Phone:	+33 1 45 29 49 65
		e-mail:	ivan.froger@orange.com
France Telecom	Didier Becam		
		Phone:	+33 2 96 05 32 02
		e-mail:	didier.becam@orange.com
France Telecom	Khadija Daoud Triki		
		Phone:	+33 1 45 29 54 57
		e-mail:	khadija.daoud@orange.com
France Telecom	Philippe Herbelin		
		Phone:	+33 1 45 29 45 87
		e-mail:	philippe.herbelin@orange.com
France Telecom	Jiejin Hou		
		Phone:	+33 1 45 29 89 79
		e-mail:	jiejin.hou@orange.com
France Telecom	Benoit Lemoine		
		Phone:	+33 2 96 05 04 99
		e-mail:	benoit.lemoine@orange.com
France Telecom	Pascal Pagnoux		
		Phone:	+33 1 45 29 40 70
		e-mail:	pascal.pagnoux@orange.com
CEA	Michael Boc		
		Phone:	+33 1 69 08 39 76
		e-mail:	Michael.Boc@cea.fr
CEA	Alexandru Petrescu		
		Phone:	+33 1 69 08 92 23
		e-mail:	Alexandru.Petrescu@cea.fr

CEA	Mounir Kellil		
		Phone:	+33 1 69 08 22 74
		e-mail:	Mounir.Kellil@cea.fr

MIK	Zoltán Faigl		
		Phone:	+36 1 463 2499
		e-mail:	zfaigl@mik.bme.hu

MIK	László Bokor		
		Phone:	+36 1 463 2499
		e-mail:	bokorl@hit.bme.hu

Montimage	Edgardo Montes de Oca		
		Phone:	+33 5 380 3577
		e-mail:	edgardo.montesdeoca@montimage.com

Nokia Siemens Networks	Tero Lötjönen		
		Phone:	+358 40 5747842
		e-mail:	tero.lotjonen@nsn.com

Nokia Siemens Networks	Pekka Wainio		
		Phone:	+358 40 5811211
		e-mail:	pekka.wainio@nsn.com

Nokia Siemens Networks	Wolfgang Hahn		
		Phone:	+4989515924122
		e-mail:	wolfgang.hahn@nsn.com

Turk Telekom	Ahmet Serdar Tan		
		Phone:	+90 212 3099975
		e-mail:	ahmetserdar.tan@turktelekom.com.tr

AVEA	Engin ZEYDAN		
		Phone:	+90 216 987 6386
		e-mail:	engin.zeydan@avea.com.tr

AVEA	Çağatay EDEMEN		
		Phone:	+90 216 987 6386
		e-mail:	cagatay.edemen@avea.com.tr

Alcatel-Lucent	Jean-Luc Lafragette		
		Phone:	+33130772738
		e-mail:	jean-luc.lafragette@alcatel-lucent.com

Alcatel-Lucent	Erick Bizouarn		
		Phone:	+33130772724
		e-mail:	erick.bizouarn@alcatel-lucent.com

Ericsson AB	Rashmi Purushothama		
		Phone:	+46 10 715 5964
		e-mail:	rashmi.purushothama@ericsson.com

Ericsson AB	Jörgen Andersson		
		Phone:	+46 10 719 7013
		e-mail:	jorgen.andersson@ericsson.com

Ericsson AB	Conny Larsson		
		Phone:	+46 10 714 8458
		e-mail:	conny.larsson@ericsson.com

VTT	Suihko Tapio		
		Phone:	+358 09 451 1111
		e-mail:	Tapio.Suihko@vtt.fi

Executive Summary

This document defines the Architecture Design Release 2 for the Celtic MEVICO project. The scope and the context of MEVICO project are summarized to recall the business drivers. An overview of Evolved Packet Core (EPC) architecture and the related network elements are described. The evolution of the network traffic and usage scenarios are the guidelines of MEVICO to generate a more efficient mobile architecture for the LTE (Long Term Evolution) and LTE-Advanced radio access systems of 3GPP.

The MEVICO architecture is based on requirements related to different aspects (usage and operational, performance, network management, mobility, scalability, reliability, availability, security, charging, energy efficiency, traffic management).

These requirements allow identifying Architecture Challenges related to the different topics (network topology, mobility, network transport and management, traffic management, applications and services).

This release of document includes technology solutions and maps them to the different architecture views and relation with other proposals. This release is not containing yet the further analysis of technology solution benefits and architecture evolution recommendations.

List of terms, acronyms and abbreviations

Generally the 3GPP used terms are used in this document [1].

Clarification of used terms in the document

Access Point Name	In 3GPP, Access Point Name (APN) is a reference to the Gateway GPRS Support Node (GGSN) or Packet Data Network Gateway (P-GW) to be used. In addition, Access Point Name may, in the GGSN or P-GW, identify the packet data network and optionally a service to be offered [2]
Application agnostic group communications	The group communications will include a variety of multimedia application types so the solution that enables the group communications shall be application-agnostic.
Busy Hour	In a communications system, the sliding 60-minute period during which occurs the maximum total traffic load in a given 24-hour period.
Connected subscription	A subscription that has one IP address assigned to enable always-on feature.
Device	A physical entity with communications interface that requires an active subscription to networking infrastructure to establish a connection. There is an endless list of devices e.g. smart phones and other mobile phones, laptops with USB dongle or integrated wireless interfaces, vehicular network with several multimedia devices, home network with sensors, actuators, home devices such as picture frame, Video-on-Demand players, Home GWs, etc., vehicular devices such as in-car multimedia player, game console, etc., other devices associated to the user such as personal sensors, body network, etc.
Dynamic resource allocation	The network shall dynamically reconfigure providing additional bandwidth to traffic demands.
Fixed broadband data connection	Wireline connection enabling speed >1Mbps per user.
Hyperconnectivity	Use of multiple means of communication, such as email, instant messaging, telephone, face-to-face contact and Web 2.0 information services. Also a trend in computer networking in which all things that can or should communicate through the network will communicate through the network.
Offloading	The traffic offloading in this document means routing away the traffic originating from the EPS/mobile network/mobile device onto some other network such as WLAN.
macroscopic traffic management	It includes all mechanisms with the primary objective to improve efficient usage of network resources. Parameters for optimization describe traffic patterns without detailed knowledge of individual flow attributes.
microscopic traffic management	It is associated with all mechanisms with the primary objective to improve performance of individual flows based on application type, user profile and other policy related information.
Mobile broadband data connection	Wireless connection enabling speed >256kbps per user and wide user mobility. Technologies include CDMA2000 EV-DO, WCDMA/HSPA, LTE, Mobile WiMAX, and TD-SCDMA.
Mobility type support	Host mobility (a host changes its point-of-attachment), user mobility (user moves from one host to another) and session mobility (old session is restored when the user moves to a new host) shall all be supported e.g. via aggregation of mobility protocols or a single protocol.
Moving network	The network and its subsequent mobility protocol(s) must support network mobility i.e. moving networks such as bus, cars, aircraft, PAN, etc.
Multi-homed Devices	Terminals with several interfaces up that and allow mobility between any IP address currently bound to the device. Multihoming is a technique that allows to be connected to several networks; it can be used to avoid the single point of failure for the network connectivity. Most of the time, the implementation is realized through use of multiple interfaces

Provider Edge	Provider Edge devices is standard layer 2 (L2) Ethernet, which is paring with the Customer Edge (CE) through a User-Network Interface (UNI)
Scalability	Scalability in a network is the ability to handle growing amount of load or users. The network is able to total throughput when increasing the load.
Small cell	Small Cells are low-powered in-building or outdoor radio access nodes that operate in licensed and unlicensed spectrum that have a range from 10 meter upto few kilometers. Types of small cells include micro, pico and femto cells, distributed radio systems with remote radio heads and WiFi hotspots. Small cells are used by mobile operators to extend the wireless service coverage and/or increase network capacity, both indoors and outdoors
Subscriber	A Subscriber is an entity (associated with one or more users) that is engaged in a Subscription with a service provider. The subscriber is allowed to subscribe and unsubscribe services, to register a user or a list of users authorized to enjoy these services, and also to set the limits relative to the use that associated users make of these services. [1]
Subscription	A subscription describes the commercial relationship between the subscriber and the service provider. [1]
Network topology	Network topology represents the layout of the interconnection between network elements e.g. routers, switches or other communication elements.
User	End user, an entity, not part of the (3GPP) System, which uses (3GPP) System services.[1]
User Equipment (UE)	In 3GPP System, allows a user access to network services. A User Equipment can be subdivided into a number of domains, the domains being separated by reference points. Currently the User Equipment is subdivided into the UICC domain and the ME Domain. The ME Domain can further be subdivided into one or more Mobile Termination (MT) and Terminal Equipment (TE) components showing the connectivity between multiple functional groups [1]. In this document UE and Mobile Device are used parallel.
vertical handovers	Vertical handover is a handover between two different radio access technologies and that do not share the same radio infrastructure. For example, a handover between 3G and Wifi is a vertical handover, but a handover between GPRS and HSDPA is not a vertical handover, it remains a horizontal handover. Usual handover with the same radio access technology is horizontal handover.

List of abbreviations

3GPP	3rd Generation Partnership Project, based on GSM Technology
AAA	Authentication, Authorization and Accounting
AKA	Authentication and Key Agreement
ALTO	Application Layer Transport Optimization
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
APN	Access Point Name
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat-reQuest
BAT	Bulk Analysis Tool
BER	Bit Error Rate
BS	Base station
BTS	Base Station
CAPEX	Capital Expenditure
CBS	Committed Burst Size
CDN	Content Delivery Network
CES	Customer Edge Security
CET	Carrier-Ethernet Transport
CIR	Committed Information Rate
CMIP	Common Management Information Protocol
CN	Core Network
CoMP	Coordinated Multi-Point
CSCF	Call Session Control Function
DDMM	Distributed and Dynamic Mobility Management
DHCP	Dynamic Host Configuration Protocol
DHT	Distributed Hash Table
DL	Downlink
DMA	Distributed Mobility Anchoring
DNS	Domain Name Server
DPI	Deep Packet Inspection
DWDM	Dense Wavelength Division Multiplexing
E2E	End-to-end
EAP-SIM	Extensible Authentication Protocol - Subscriber Identification Module
EBS	Excess Burst Size
EIR	Excess Information Rate
eNB	Evolved Node B (eNodeB)
EPC	Evolved Packet Core
ePDG	Evolved Packet Data Gateway (ePDG)
EPS	Evolved Packet System
ETSI	European Telecommunications Standards Institute
E-UTRAN	Evolved UMTS Terrestrial Radio Access Network
EVC	Ethernet Virtual Connection
FI	Future Internet
FTTA	Fiber To The Antenna
Gbps	Giga Bit Per Second
GBR	Guaranteed Bit Rate
GGSN	Gateway GPRS Support Node

GHz	Giga Hertz
GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
GUTI	Globally Unique Temporary ID
GW	Gateway
HeNB	Home eNB
HetNet	Heterogeneous Network
HIP	Host Identity Protocol
DEX	Diet Exchange (HIP DEX AKA)
HNP	Home Network Prefix
HO	Handover
HSPA	High-Speed Packet Access
HSPA+	Evolved HSPA (3GPP release 7, including I-HSPA)
HSS	Home Subscriber Server
HTTP	Hypertext Transfer Protocol
HW	HardWare
I-CSCF	Interrogating-CSCF
ID	Identifier
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IFOM	IP Flow Mobility
I-HSPA	Internet HSPA
IKEv2	Internet Key Exchange, version 2
IM	Instant Messaging
IMS	IP Multimedia Subsystem
IMT	International Mobile Telecommunications
IMT-A	IMT Advanced
IP	Internet Protocol
IPsec	Internet Protocol Security
IS	Intermediate System
IS-IS	Intermediate System to Intermediate System
ISP	Internet Service Provider
IT	Information Technology
ITU	International Telecommunication Union
KPI	Key Performance Indicator
L2	Layer 2
L3	Layer 3
LAN	Local Area Network
LFN	Local Fixed Node
LSP	Label-Switched Path
LTE	Long Term Evolution
LTE-A	LTE Advanced
LMA	Local Mobility Anchor
LU	Location Update
M2M	Machine-to-Machine
MAC	Media Access Control, a low layer protocol
MAG	Mobile Access Gateway
MASE	Media Aware Serving Entity

MBH	Mobile BackHaul
MBR	Maximum Bit Rate
MCCS	Multi-Criteria Cell Selection
MIH	Media Independent Handover
MIMO	Multiple Input Multiple Output
MIP	Mobile IP
MLB	Mobility Load Balancing
MME	Mobility Management Entity
MN	Mobile Node
MNO	Mobile Network Operator
MTM	Microscopic Traffic Management
mP4P	Mobile P4P
MPLS	Multi-Protocol Label Switching
MPLS-TP	MPLS Transport Profile
MPTCP	Multi-Path TCP
MPTCP-Pr	MultiPath TCP - Proxy
MR	Mobile Router
MRO	Mobility Robustness Optimization
mRVS	mobile Rendezvous Server
m-SCTP	mobile-SCTP
MSO	Multimedia Streaming Optimizations
MTC	Machine-Type Communications
NB	Node B
NB-IFOM	Network-based IP Flow Mobility
NEMO	Network Mobility
NETCONF	Network Configuration Protocol
NG	Next Generation
NIMTC	Network Improvements for Machine-Type Communications
NMIP	Not Mobile IP
NW	Network
O&M	Operations & Maintenance
OC	Optical Carrier
OPEX	Operational Expenditure
OTT	Over The Top
P2P	Peer-to-Peer
P4P	Proactive Network Provider Participation for P2P
PBB-TE	Provider Backbone Bridge Traffic Engineering
PBM	Policy-based Management
PCC	Policy and Charging Control
PCRF	Policy Control and Charging rules function
PE	Provider Edge
PDN	Packet Data Network
P-GW	Packet Data Network (PDN) Gateway
PMIP	Proxy Mobile IP
PMIP-RO	Proxy Mobile IP – Route Optimisation
PoP	Point of Presence
PPP	Point-to-Point Protocol
PS	Packet Switched

QCI	QoS class identifier
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Technology
RFC	Request For Comments
RLF	Radio Link Failure
RNC	Radio Network Controller
ROF	Radio Over Fiber
RPC	Remote Procedure Calls
SA	Security Association
SAE	System Architecture Evolution, LTE’s core network architecture
SAIL	Scalable and Adaptive Internet Solutions
SCTP	Stream Control Transmission Protocol
SDH	Synchronous Digital Hierarchy
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway
SIM	Subscriber Identity Module
SIMTC	System Improvements to Machine-Type Communications
SIP	Session Initiation Protocol
SNMP	Simple Network Management Protocol
SNR	Signal-to-Noise Ratio
SON	Self Organizing Network
SW	Software
TCP	Transmission Control Protocol
TDD	Time-Division Duplex
TEHO	Traffic Engineered Handovers
THP	Traffic Handling Priority
TM	Traffic Management
TRILL	Transparent Interconnection of Lots of Links
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telephone System
USB	Universal Serial Bus
VLAN	Virtual Local Area Network
VoD	Video-on-Demand
VoIP	Voice over IP
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
WDM	Wavelength-Division Multiplexing
WiFi	“Wireless Fidelity” a trademark of WiFi Alliance (IEEE 802.11 certified devices)
WiMAX	Worldwide Interoperability for Microwave Access (IEEE 802.16 standard)
WLAN	Wireless Local Area Network
WMN	Wireless Mesh Network
XML	Extensible Markup Language

References

3GPP standards

[1] 3GPP, “Vocabulary for 3GPP Specifications”, TR 21.905, release 10.

[2] 3GPP, “General Packet Radio Service (GPRS); Service description; TS 23.060, release 10.

[3] 3GPP, “Network architecture”; TS 23.002, release 10.

[4] 3GPP, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access”; TS 23.401, release 10.

[5] 3GPP, “Architecture enhancements for non-3GPP accesses”; TS 23.402, release 10.

Publications

[6] European Commission: A Digital Agenda for Europe (May 2010): http://ec.europa.eu/information_society/digital-agenda/documents/digital-agenda-communication-en.pdf. Digital Agenda web page: http://ec.europa.eu/information_society/digital-agenda/index_en.htm

[7] Cisco Visual Networking Index: Hyperconnectivity and the Approaching Zettabyte Era (June 2010): http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html

[8] Celtic project IPNQSIS (IP Network Monitoring for Quality of Service Intelligent Support), http://www.celtic-initiative.org/Projects/Celtic-projects/Call7/IPNQSIS/ipnqsis-default.asp

[9] Perfecting policy control: Ericsson white paper 284 23-3158 Uen; May 2011; http://www.ericsson.com/res/docs/whitepapers/WP-e2e-policy-control.pdf

[10] MEDIEVAL (MultiMEDia transport for mobIlE Video AppLications, http://www.ict-medieval.eu/)

1. Introduction

This document defines the Architecture Design Release 1 Documentation for the Celtic MEVICO project. This is the D1.1 Report, part of the Work Package 1.

This study is conducted in 2010 and 2011. The content is contributed by the whole MEVICO project consortium.

1.1 Scope of the document

This document consists of the future mobile network traffic scenarios and usages that MEVICO architecture shall support.

The goal of this document is to analyze the trends and scenarios to enable defining the requirements for the future mobile network architecture.

1.2 Structure of the document

The current section 1 introduces the scope and the context of MEVICO project. The section 2 describes the “Network Traffic and Usage Scenarios”. The section 3 identifies the “Architecture Requirements”. The section 4 describes the “Architecture Challenges”. Based on these challenges, the section 5 proposed “Technology Solutions”. The choice of MEVICO “Architecture Approaches” is described in section 6.

1.3 Business drivers for the MEVICO project

Affordable, truly accessible mobile broadband has matured with HSPA (High-Speed Packet Access), HSPA+ (3GPP release 7, including I-HSPA), and LTE (Long Term Evolution) in the near future. It has blurred boundaries between mobile/fixed and voice/data for end-users, operators and application developers.

Mobile data traffic is expected to grow faster than the fixed Internet for the coming years and with the same rate as fixed Internet in the long term. Radio access and core network must be scaled to accommodate the expected traffic growth, especially if we consider limited revenue growth. It will lead to access and core networks cost pressure.

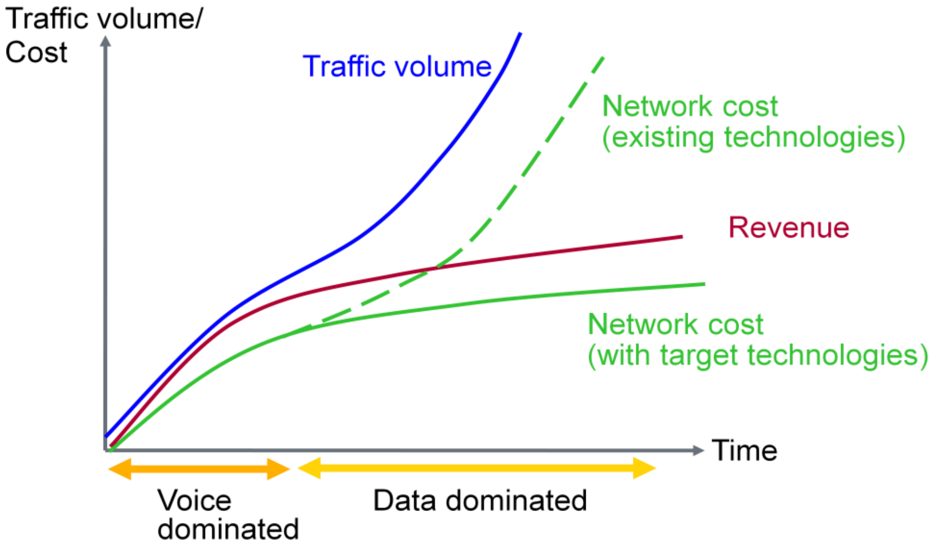

Figure 1 Problem statement of radio access and core network

The operators have to satisfy the demands of the new services and data volume growth, in order to remain competitive. New business models are required and redefining business priorities might also impact the selection of network infrastructure.

1.4 Overview of Evolved Packet Core (EPC) architecture

In 3GPP release 8, LTE and SAE (System Architecture Evolution) work resulted in the specification of the E-UTRAN (Evolved UTRAN) and in the specification of Evolved Packet Core (EPC); both components form the EPS (Evolved Packet System). LTE-EPC, is the name for the long term evolution of UMTS.

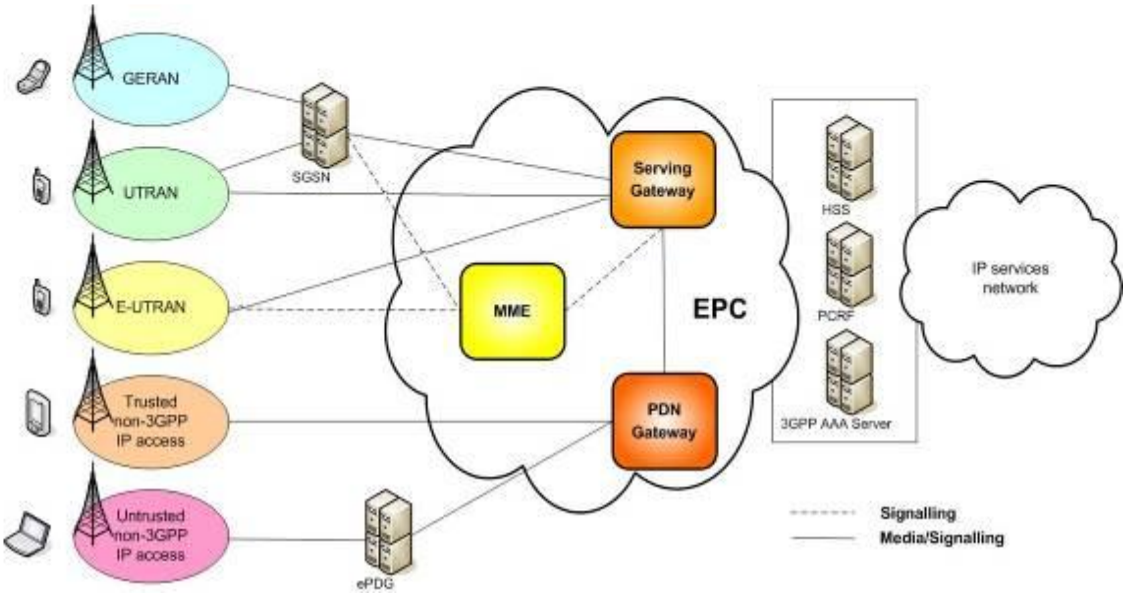

Figure 2 Evolved Packet Core network

The Evolved Packet Core is made of three main network entities, described in [3], [4], and [5]. The user plane consists of two types of nodes, the **Serving Gateway** (S-GW) and the **PDN Gateway** (P-GW). The control plane is made up of a separate **Mobility-Management Entity** (MME :

- The **MME** manages all the signalling (control plane):
- The **S-GW** terminates the user plane interface towards E-UTRAN:
- The **P-GW** terminates the user plane interface towards one or more Packet Data Networks:

1.5 Other network elements linked to the Evolved Packet Core (EPC)

The other network elements linked to EPC are the following:

- Legacy 3GPP access: Serving GPRS Support Node (SGSN),
- Non-3GPP access: Evolved Packet Data Gateway (ePDG), 3GPP AAA server;
- Evolved UTRAN (E-UTRAN);
- Home eNodeB;
- Policy and Charging Control architecture.

2. Network Traffic and Usage Scenarios

2.1 Mobile traffic, service, and technology evolution 2008-2020

This section collects traffic analyses from the last couple of years and provides some traffic forecasts (2008-2020). The evolution scenarios for the growth of traffic volumes and the number of users as well as the impacting application, service and technology evolution scenarios are covered.

2.1.1 Traffic Data evolution

The mobile broadband subscriber and traffic volume increase is inevitable and the future network architecture has to be designed to cope with it. The mobile traffic global increase is a consequence of several factors: growth of the mobile subscriptions (e.g. growth of population, improving living standards), evolution of the mobile networks/ devices and services (e.g. affordability of capable devices, enabled connection speeds, low cost flat rate data plans, easier usage, evolution of communication needs). And there is a huge increase potential of devices/subscriptions/traffic with the Machine-to-Machine (M2M) communications.

The network needs to be optimized to maximize the end-user mobile broadband experience, minimize the mobile device battery consumption and ensure efficient, congestion-free network performance. Because the available mobile network frequency bands are scarce and the utilized spectral efficiencies are tending to the theoretical limit, several other methods to cope with the increasing capacity demand need to be utilized. Even though the regulation is planning to open new Digital Dividend frequencies in the coming years, this alone will not be able to totally solve the problem.

Some regulatory or public funding drivers can have an additional impact on the operator interest to invest to expand the network capacity. There are some guidelines drawn in the European Commission Vision 2020 related to Digital Agenda work [6], for example, guidelines defining the minimum connection speed targets for broadband Internet.

2.1.2 Services and application evolution

The most remarkable mobile user application challenges in the future are expected to come from video, social networking and M2M types of services, which exponentially will increase the traffic volume.

Video: The sum of all forms of video (including Internet TV, Video on Demand, interactive video, and Peer-to-Peer (P2P) video streaming, mobile 3DTV, etc.) will account for close to 90 percent of consumer traffic (fixed and mobile) by 2012 [7]. The evolution of the Content Delivery Networking (CDN) and intelligent data caching technologies in the fixed network side might have impact on the mobile network architecture, mainly by bringing the content lower in the network and enable efficient usage of several parallel flows from different content sources.

Social networking: Consumers are more and more using a variety of services to communicate (e.g. email, instant messaging, twitter, Facebook, video, VoIP, and a host of other social networking applications) that use a mix of voice, video and messaging.

M2M: M2M communications have enormous potential (tens of billions of devices to be connected) to become the leading traffic contributor. These types of services will also generate different traffic time variations than those due to human activity (non-busy hours, strict latency requirements, initialization/synchronization after recovering from a network failure).

Mobile Gaming: As the handheld devices are equipped with better hardware, online mobile gaming traffic is expected to become a significant traffic contributor. Maintaining game stability among several mobile users necessitate the transmission of state updates between each mobile device with low latency.

To efficiently cope with the challenges related to these services, there is a need to consider the mobile network architecture optimization to allow efficient use in heterogeneous network environments and understand the impact of CDN technology evolution.

2.1.3 Evolution-enabling technologies

The main evolutions are related to:

- Bandwidth needs in radio technologies become similar to fixed network;

- Miniaturization of radio technologies facilitates the deployment of devices with only basic connectivity functionality increasing the growth of machine to machine communications; Evolution of processing and networking technologies: there is no clear indication whether in the future the applications will be used on the end device or on the network.

2.1.4 Aspects to be taken into account

The following aspects should be taken into account:

- Flat rate pricing in mobile broadband networks have stimulated many users to change their fixed broadband access to mobile.
- Users should obtain similar bandwidth capacity regardless whether the underneath technology is wireless or wire line.
- Traffic growth in mobile broadband networks is mainly due to the evolution of the mobile networks, devices and services. And to a certain extent it is due to a smooth migration of users from fixed broadband networks.
- New traffic patterns and exponential traffic increase originated from new devices that incorporate mobile broadband connectivity (e.g. sensors, home appliances).
- The evolution of User Interface and ways of interacting with the mobile devices will open up the demand for new applications which require higher bandwidth and generate more traffic.
- The new service levels enabled by the hyperconnectivity will place huge capacity demands on the networks. The four key growing enablers of hyperconnectivity are: (a) the growing penetration of high-speed broadband, (b) the expansion of digital screen surface area and resolution, (c) the proliferation of network-enabled devices, and (d) the increases in the power and speed of computing devices.
- Context-aware mobile computing, in which applications can discover and take advantage of contextual information (such as user location, time of day, nearby people and devices, and user activity), can introduce new challenges to system infrastructure.
- All available capacity will be exploited, with affordable pricing. In mobile networks, different charging models (with respect to fixed broadband) shall/could be exploited, in order to share the limited radio access capacity, since flat rate alone should not be the most suitable model.
- Net neutrality has to be respected in the service delivery and quality.

2.1.5 Key metrics

There are general challenges in the mobile network future trends identified in the studies carried out in MEVICO project: Increase of subscriber amounts (with huge potential of M2M), high increase of the data amounts (driven by video delivery), always on applications, availability of heterogeneous network and multiple types of interfaces in User Equipments (UEs).

The results of the preliminary studies done in MEVICO project show that there are several drivers that will cause network scalability and optimization challenges. The current architecture needs to be scaled according to the growth of:

- data traffic volume per user by about 3-10 times by year 2020 compared to 2010
- number of mobile broadband subscriptions and end users increase by 8-12 times by year 2020 compared to 2010, according to our internal traffic forecasts.
    - when including the M2M devices even to 50 times
- mobility rate (users changing their location during the active broadband communication) will remain around 20-25%, so most of the mobile broadband usage takes place in stationary location.
- number of network nodes, due to densification/frequency overlay/small cell needs, heterogeneous networks
- network signalling load. Even mobile core network signalling load share of the total traffic amount is estimated to increase moderately from 2% to 3% from 2010 to 2020, but with the estimated total traffic, subscription and network node increase the signalling will increase considerably

Therefore, the future mobile architecture should deal among others with the challenges associated with the increase of traffic, mobility and signalling traffic while keeping the OPEX under competitive levels for operators.

2.1.6 Conclusion of the mobile data traffic evolution

Mobile broadband usage has taken off in the last couple of years due to several factors such as improved network capabilities, affordable data plans and the evolution of end-user devices. The main drivers for the future mobile broadband traffic growth are the increase of global subscribers number, the evolution of the user devices that enable easy usage of data hungry services and the evolution of the network functionality that enables operators to provide high speed mobile services with attractive pricing. It is foreseen that these factors, in conjunction with the evolution of the applications in new areas, will tempt more users to utilize new devices and consume more data. Most of the data hungry applications are related to entertainment content, like video streaming, social networking, mobile gaming, and thus the enthusiasm to use them depends largely on the service costs. The end result is that mobile broadband traffic volume will increase in the future and the network architecture evolution has to be optimized to cope with it.

2.2 Mobile usage scenarios

The target of this chapter is to identify the trends, the new technologies and drivers having an impact on mobile core network architecture. Scenarios bringing new requirements to mobile networks in terms of latency, mobility, traffic management, etc. have been identified.

2.2.1 End user service scenarios

Fixed - Mobile Convergence

The Fixed-Mobile convergence section addresses usage scenarios where there is no expected QoE difference for the end user on whether the communications are done over fixed or mobile networks. The following three use cases described are already defined in 3GPP.

- Internet access with Parental control and personal firewall,
- Voice/Multimedia and Charging,
- Video.

Another use case is mass delivery of real-time multimedia content which has specific requirements.

M2M communication and wireless sensor network scenarios

The machine to machine scenarios included the followings:

- Remote healthcare
- Smart metering / industrial monitoring
- Mass monitoring, mass remote control, Tracking objects
- Automotive connectivity traffic scenarios
- Internet of things and future Mobile networks

2.2.2 Network (operator) usage scenarios

The Network (operator) usage scenarios include the following:

- Energy saving improvements
- Virtualization and Cloud computing
- Seamless user experience of mobile Internet over multiple data GWs and multiple interfaces
- Small cell deployment
- Secured access by design to limit unwanted traffic to mobile clients
- Mass event coverage and capacity enabling with wireless mesh transport
- Automatic and Secure Layer 2 Virtual Private Networks

2.2.3 Conclusions on Mobile usage scenarios

The traffic analysis based on aggregation alone is not enough in the future application contexts. Per user and per application analysis is needed for a better understanding and optimization of traffic. Detailed knowledge of traffic patterns, including packet size and time intervals, are needed to improve resource allocations and obtain the required end-user's QoE. In order to manage the increased traffic and new applications with new requirements, LTE-EPC technologies have adopted an all-IP architecture that integrates a more distributed management and QoS strategy. This architecture simplifies the network stack, but makes the management more complex.

3. Architecture Requirements

The architecture in MEVICO focuses on the evolution of the mobile packet core network for the LTE (Long Term Evolution) and LTE-Advanced of 3GPP. MEVICO will study and define system concepts to evolve the Evolved Packet Core (EPC) of 3GPP in the mid-term in 2011-2014 towards the requirements that are challenging the packet connectivity capabilities.

The project will focus on the network aspects to complement the research and standardization (3GPP) already ongoing for defining and standardizing a new radio system LTE-Advanced as the next step of the LTE radio technology in 2010. The project will not address the radio interface aspects, but will rather enhance the network architecture, higher bit rates and higher capacity. Nevertheless, the peculiarities and limitations of the radio portions are reflected into the core network and those impacts will be therefore addressed in the project.

As an example of requirements, we will focus on the following illustrative and challenging video services to show that the architecture covers all the spectrum of potential services, namely Internet TV, VoD, Personal Broadcasting and Interactive Video. The MEVICO network will exploit heterogeneous wireless access to deliver media content to mobile customers. MEVICO will focus on LTE Advanced and Wireless Local Area Network (WLAN) access co-operation as depicted in the figure below:

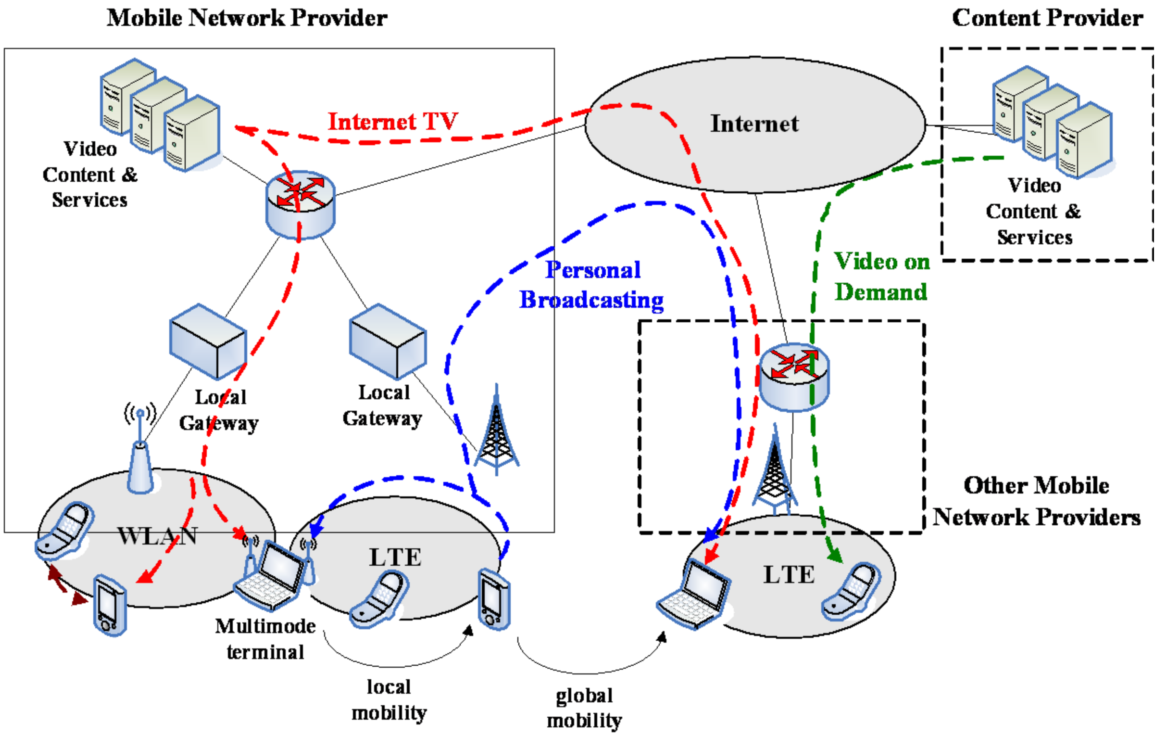

Figure 3 MEVICO possible network vision for improving video service efficiency

The main requirements for the future mobile networks are the following:

- Enabling the efficient use of the heterogeneous network capabilities, like multi-access (several simultaneous parallel paths, fixed-mobile, convergence) and multimode (several overlapping alternative Radio Access Technologies (RATs)).
  - This requires an efficient and optimized way of selecting/utilizing the multiple available paths, because, until now, the mobile device is not able to be active simultaneously on all RATs.
  - Use of multiple interfaces brings new challenges in different functions: Identity Management, security/privacy-preserving methods, charging, lawful interception, etc.
- End user Quality of Experience is the key driver for architecture evolution.
  - Architecture related capacity bottlenecks shall be avoided (i.e., the scalability has to be ensured).

  - Latency and throughput need to be kept optimal when traffic load is high.
  - "Always On" applications need to be handled optimally, without causing extensive load to network signaling.
- Implications from the new, very high capacity radio access network topologies – like LTE and WiFi 802.11n - shall be taken into account.
- Cost optimization needs to be addressed with care since the operator's revenue increase will be modest due to the widespread flat rate model...
- Traffic optimizing concepts under study for the Future Internet, e.g. to access the content cached near to the user, Information Centric Networking/Content Delivery Networking concepts, shall be studied to understand their impact for optimizing the mobile network architecture.

In the subsequent sections, we will list all the requirements. The requirement numbering stems from MEVICO-internal requirement structuring and that for brevity, only the titles of the requirements are listed. Only the requirements with very high and high priority are taken into account:

- A refers to Architectural requirement
- F refers to Functional requirement
- N refers to Non-functional requirement

3.1 High-level requirements – user and operational aspects

The requirements in this chapter shall secure that the network provides mobility functionality within and across the different access systems.

A 5.1.1: Heterogeneous transport technologies

A 5.1.2: Multiple operational domains

A 5.1.4: Topology diversities for network architectures

A 5.2.1: Session mobility support

A 5.3.1: QoS guarantee

A 5.5.3: Support for Disparate Wireless Technologies

A 5.6.1: IPv4 / IPv6 Cross-Family Communication

3.2 Performance requirements

Network performance needs to satisfy the demands of the new services and data volume growth, if operators want to remain competitive.

A 6.1: Optimized architecture for content delivery

A 6.4: Low latency

A 6.7: Synchronization

A 6.8: Mobility type support

A 6.9: Device characteristics

F 6.1: Small cell signalling optimization

F 6.6: Multi-elements connectivity

N 6.1: Dynamic resource allocation

3.3 Network management

Network management deals with operation, administration, maintenance, and provisioning of the network. On top of these normal tasks special attention should be paid to the large number of new device types attached to the network and energy efficiency aspects in the core network although the largest potential here is in the access networks.

F 7.7: Flexible network operation

F 7.8: Efficient network monitoring

3.4 Mobility requirements

Mobile-Fixed network mobility, Multi-radio (LTE, HSPA, WiFi) and Multi-layer (Macro, Micro, Pico, Femto, multifrequency) support in combination with the traffic growth adds complexity to the mobility functionality and this should be reflected in the mobility requirements.

A 8.1: Seamless Handovers

A 8.2: Optimized mobility protocols

A 8.3: Protocol interoperability

A 8.5: Context Transfer

A 8.7: Selection of Mobility protocols

A 8.9: Small cell mobility

A 8.11: Mobility between heterogeneous radio technologies

A 8.12: Support of moving networks

3.5 Scalability requirements

It is important to have a scalable solution to be able to take care of the different traffic growth scenarios.

A 9.1: Small cell support

A 9.4: Signalling scalability

A 9.5: Robust network

A 9.6 Device addressing

N 9.3: Scalability of management solutions

3.6 Reliability and Availability requirements

The increasing possibilities to connect people, things, etc, to distribute functions e.g. cloud computing, add more and more services that rely on the network and thus put increasing requirements on the network reliability in a broad sense. The new architecture should meet the needs these new ways of using the network.

A 10.4: Application agnostic group communications

A 10.6: Interfaces availability information

A 10.9: Support for Multi-homed Devices

A 10.11: Routing loops avoidance

F 10.2: IP flows routing

3.7 Security and privacy requirements

The emergence of the new networks comprising converging technologies, different access technologies and environments mixed of computation and communication, requires new and strong security solutions (including privacy, authentication, need for encryption,…)

A 11.2: Protection against cyber-attacks

A 11.3: Strong Mutual Authentication

A 11.5: Address Ownership

A 11.9: Location and Identity Privacy

A 11.10: Lawful Interception

A 11.15: Network isolation

A 11.17: Node identity

A 11.18: Control layer security

A 11.19: Secure Zone-Based Authorization

A 11.20: User profile based secured zones

F 11.1: Ensure Network neutrality

F 11.5: Device disconnection

F 11.7: Unwanted traffic avoidance

F 11.8: Multipoint VPNs

3.8 Charging Aspects

The network shall support various charging models including all those supported by the 3GPP system contained within TS22.115 and be able to support introduction of new charging schemes including online and offline schemes, and charging schemes for the multi-access system environment.

A 12.1: User profile extension

F 12.1: M2M Charging

N 12.1: Operator legal aspects

3.9 Energy efficiency

The specification of new architecture design must take into account energy-efficiency issues.

F 13.1: Minimize device battery consumption

Sleeping modes of access nodes

3.10 Traffic management

Traffic management and engineering cover all measures to dynamically control and optimize traffic flows in a network domain or in a global view of the Internet, aiming at ensuring a maximum throughput and sufficient QoS/QoE for the users. In order to achieve this goal, traffic management includes methods and schemes for dimensioning, admission control, service and user differentiation and failure resilience as well. The specification of the MEVICO architecture should meet these needs.

F 14.1: Application-awareness

F 14.2: Support for macroscopic traffic management

F 14.3: Support for microscopic traffic management

F 14.4: Improved content resource selection & caching

F 14.5: Support for deployment of new network resources and upgrading processes

4. Architecture Challenges

Based on the requirements (section 3), the architecture challenges, taken into account in MEVICO project, are related to the following aspects:

- Network topology,
- Mobility,
- Network transport,
- Network management,
- Network applications and services.

4.1 Network Topology related challenges

The most important challenges concern the scalability of the network that can be ensured among other ways by an adaptive/flexible topology, against different parameters such as traffic load, subscriber density, number of network connections and signalling transactions.

- Due to LTE radio throughput enhancements and new smart phone applications, the mobile network busy hour data traffic volume is expected to increase up to the ten fold in the next 10 years, so that a new backhaul /core network topology might be required to increase the network throughput capacity.
- In the same time, due to the increase of the mobile broadband subscribers number and due to the introduction of M2M devices with mobile network connection, the network topology will have to be adapted against the density increase of attached User Equipment (UE) and that especially as a default EPS bearer will be systematically created for each new attached LTE UE.
- As the new EPC network aims at supporting both conversational and classical data traffic, Quality of Service mechanisms will have to be able to handle this higher number of network connections per UE. This might induce a change in network topology.
- If the increased need for access network throughput and session handling capability leads to cell density increase (huge number of Femtocells for instance), this will cause potentially bottlenecks for the communication or the signalling transactions within the centralized gateway and servers that handle mobility and service provisioning, once reached the network elements capacity upgrade limits.

The scalability issues of the mobile network will depend on the capacity evolution of the EPC nodes[1] compared to the network load increase for each of the above parameters. An additional challenge is to identify from CAPEX and OPEX point of view the most appropriate EPC nodes localization from a centralized to a distributed architecture, and that will enable to eventually distribute further the EPC architecture. The proper distribution of the EPC nodes will have to take into account the following parameters:

- In the case of heterogeneous access networks, the core nodes optimal positioning for LTE networks and non 3GPP networks might not be the same since the session amount, the traffic bandwidth, the handover frequency or even the service types profile might change when a UE is connected to 3GPP and non 3GPP access network.
- Content and cache servers are getting deployed at the edge of the fixed networks, so that a distribution of the mobile network could permit to merge fixed and mobile content and cache servers.
- Centralized mobile networks permit the use of customized accounting devices in order for a mobile operator to propose offers for its customers, whereas the distribution of the mobile network requires the use of less costly and by the way less accurate and more standard accounting features.
- The complexification of sGi interface/APN management/PCC architecture/company connections…

4.2 Mobility related challenges

The increase in the number of connected devices, diversity of access networks, and the resources limitations pose real challenges on how the network will handle security, users, and flows contexts. This together with the expected data

[1] Note that several vendors already present high capacity figures, there is an order of magnitude of one million for simultaneous active users per MME and/or S-P Gateway.

traffic growth will have a serious impact on flows performances when considering the current centralized architecture approach and existing mobility management and routing procedures (e.g., bottlenecks, overloaded access networks). Specifically, those following challenges are of crucial importance:

- Anchor-based mobility management protocols (Mobile IP, Proxy Mobile IP, etc.) for non-3GPP accesses rely on centralization of all traffics towards a unique anchor wherever UEs are currently attached (potentially far from the anchor). Mobility of UEs will lead to considerable amount of traffic routed throughout the core network to the centralized anchor.
- Some UEs applications need regular access to the network, which is often referred to as the “always-on” mode even if no user data is to be transmitted. This means that some traffic still passes through the P-GW and P-GW load balancing is difficult to perform.
- Current network devices may have several interfaces able to get access to different types of network (3G, WiFi, etc.). When one access network is overloaded, it might be possible to redirect traffic to other access networks or to perform multipath routing.

A decentralized architecture with multiple external gateways is a relevant approach to distribute network resources and to handle scalability issues. However, it is expected that the multiplication of GWs will also lead to more frequent inter-GW handovers. Therefore, mobility management solutions and security mechanisms have to be adapted to cope with this phenomenon.

- In some cases, UEs communications do not require: (1) the support of a specific (L2, L3) mobility management protocol because it is handled at another layer (e.g., application layer with SIP) or because UEs are mainly static (e.g., M2M devices, sensors, home location, etc.), (2) seamless mobility support as transport protocols are able to handle packet loss (e.g., TCP). However, at attachment all UEs are handled automatically by most current mobility management protocols, leading to wasted network resources for the above depicted.
- When moving from one un-trusted access system to another (like between two different WLAN networks) a considerable delay is introduced by setting up a new security association. Furthermore, the current use of IKEv2 in EPC can lead to overlapping (encapsulated) IPsec connections. E.g., in case of initializing an IMS session through a 3GPP WiFi access, an IPsec association is established both on the network level and on the SIP signalling level, resulting in overprotection and signalling overhead between the UE and ePDG.
- Paging enables to reduce energy consumption as it is not necessary for the MN to be permanently connected to the network. In EPC the paging requires that the MN has a specific allocated ID (the GUTI most of the time). In a distributed architecture, this ID might change frequently. Current paging procedures – operating when packets are coming in - would be non optimal with distributed MME or even unfeasible. Paging and Location Update (LU) procedures should take into account the upcoming multiplicity of gateways and interfaces per active UE to extend and improve the performances of idle mode management procedures.
- To improve user experiences, the EPC might propose/enforce its policy of vertical handovers towards networks with higher available resources. Some UEs are able to connect to several types of accesses or networks (LTE, Femto, Pico, WLAN, etc.) and so, the operator could have in the core network, functions to support smart vertical handover. Some of these functions already exist (e.g..802.21, and for selection/reselection: ANDSF) but they need to be upgraded accordingly.

To leverage all those benefits, mobility management protocol should be extended to support new types of UEs (moving networks, M2M, etc.) and optimized to reduce routing path lengths. Meanwhile, new routing solutions should be overseen to better handle UEs mobility.

- Existing mobility management protocols do not all support moving networks (train, bus, aircraft, cars, boats, etc.). Those networks are interfaced by one or more mobile routers and provide connectivity to several UEs
- Anchor-based mobility management solutions suffer from triangular routing (the routing towards the anchor when two UEs are close to each other). Such sub-optimal routing has to be handled to improve network resources usage.
- Future routing solutions may require new locator namespaces and routing mechanisms. Introduction of new locator types and routing mechanisms specific to the intra-domain should be supported independently from the identifiers used in the service stratum, and without influencing inter-domain routing.

4.3 Network Transport related challenges

Due to Internet and peer to peer services, the traffic has increased heavily but the flat rate tariff prevents revenues to grow in similar pace to cover the increasing costs. New innovative Mobile Transport solutions, possibly optimized together with future mobile systems (LTE-Advanced and beyond), are needed to overcome the cost crisis from transport point of view. These changes consist of certain architecture changes as follows.

A **flat architecture** of LTE, i.e. moving radio controller functions to the BTS, affects a lot the quality and performance requirements to MBH transport. The delay sensitive loops do not necessarily exist anymore between a BTS and its controller. On the other hand new synchronized air interfaces may need very strict mutual timing requirements (microsecond level of phase/time synchronization accuracy) between base stations.

Some **transport node functionalities can be integrated to the base station** (e.g. Ethernet switching), when BTS functions like a part of normal MBH solution. Switching and some network management functions are physically inside a BTS but they are part of E2E MBH concept.

Some BTS internal interfaces can be brought out and extended by fibre where available; i.e. **BTS is split in two parts** – centralized base band processing node (BB Hotel) and distributed antenna RF heads.

The **new LTE X2 interface** between the adjacent base stations (eNBs) of LTE architecture is used for handover (HO) negotiations (control plane) and data forwarding (user plane) caused by the handover process.

The energy consumption, CAPEX and OPEX will increase with the bandwidth.

The size of addressing and routing tables will increase with the number of end points, thus the signalling overhead will also increase.

The transport network needs to support various migration paths with different technologies (e.g. Carrier Grade Ethernet, MPLS-TP, IP/MPLS, PBB-TE, etc.

Multiple topologies (i.e. centralized or distributed) should be supported by the transport network.

The security concept management should be developed in a more dynamic environment (re-negotiation of security parameters etc.)

The importance of the horizontal X2 interface might increase

- o Some RAN related new features might need to increase the transported data and set strict latency requirements for the X2 interfaces between the eNBs
- o The amount of X2 peers increases with the increasing amount of eNBs

The Network Transport related challenges are the following

- The transport network should allow the possibility to be shared with co-sited base station from different operators.
- New synchronization requirements have to be supported.
- In order to share the transport network between multiple operators or in order to differentiate various traffic flows both IP addresses and VLANs play an important role so the transport network should be able to handle them efficiently.
- Ethernet is widely deployed in core network and mobile backhaul so it does not only require point to point or point to multipoint but also multipoint to multipoint connections.
- The sharing of functionalities between L2 (i.e. Ethernet switch) and L3 (i.e. IP router): routing and addressing are based on IP at L3 but Ethernet is implementing part of that functionality.
- The transport network needs to provide secure communications so Ethernet needs to address security or IPSec is handling the security, thus it has some performance impact.
- The Transport network needs to differentiate multiple traffic flows with different QoS.
- The transport network needs to support the required capacity according to the expected traffic as well as maintaining acceptable delay.
- The transport network should provide plug and play functionality to allow integrating new eNB when needed due to capacity or when changing technologies in the existing eNB via SW upgrades.

4.4 Network Management related challenges

The evolution of the RAN introduces new requirements (for instance, CoMP and strict requirement on the X2 interface) and increases the complexity of network management that needs to deal with the co-existence of different technologies, e.g., RATs (HSPA, HSPA+, LTE, LTE-A, WIFI). In addition, distribution of the architecture increases the number of network elements to be managed. SON (Self Organizing Network) for LTE RAN provides several features optimizing radio resource usage and automating radio network setups. SON will help simplifying transport setup for network elements (e.g. femto/micro cells). Related to SON's features, the following items bring some challenges to the mobile network architecture:

- The radio SON features are transport agnostic, thus their operation may unintentionally conflict with transport network management.
- The number of nodes in the EPS (eNBs, GWs, IMS servers, content servers, routers, switches, Femtocell GWs) is very high and the management solutions should enable scalable operations. Furthermore, the increase of M2M subscriptions and M2M generated traffic monitoring requirements might set new challenges to scalability.
- With the common EPC for multiple radio accesses, intra-3GPP inter-RAT handovers need to be improved for balancing load between RATs coexisting over the same coverage area. In the same way, hand-overs between 3GPP and non 3GPP accesses (e.g. LTE – WiFi) can provide boost in performance of the network as perceived by the user and improve resource usage (see load balancing).
- Different RA technologies (HSPA, HSPA+, LTE, LTE-A) will coexist for a significant time and it would be uneconomical to build separate mobile backhaul or management system for each of them, thus these should be able to handle the traffic and management functions of the RA technologies efficiently and allow sharing of the bandwidth without unwontedly privileging one customer over another.

4.5 Traffic management related challenges

As mobile and wireless communication architectures evolve toward broadband multiplay and multimedia networks, the demands for solutions on the infrastructure increase. Legacy voice, and novel data, video and other applications are to be served on the same network, simultaneously. Advanced terminals (i.e., smart phones, tablet PCs and other mobile devices) are spreading and consuming more and more network resources by running their multimedia applications and services. Consequently, the needs for available wireless bandwidth will constantly increase and LTE/LTE-A networks will likely follow the same path as wireline networks in the past resulting in a significant expansion of CDN and P2P traffic volumes.

In such a fast development it is essential that the network must be aware of each application's traffic type and enforce traffic management and control (i.e. priority, routing, bandwidth, etc.) required for ensuring improved Quality of Experience for every user anytime and anywhere. Assuring that mobile and wireless communication systems are application-aware, operators can achieve flexible adaptation to any new application and traffic pattern as soon as they emerge in the future. Operators need to install effective management tools to control every traffic component using QoS policies, prioritised access and admission control, bandwidth allocation schemes, traffic shaping and rate control, and flow based processing. Only such an active, advanced traffic management will ensure that operators can provide cost-effective data transfer with real-time multimedia information over heterogeneous access architectures of future networking schemes.

Based on these considerations the main traffic management related challenges identified in project MEVICO are the following:

- Satisfy user experience with minimum of infrastructure resources and still be flexible to handle the possible large variation of traffic patterns over time.
- Initiate handovers of sessions and/or flows not only based on signal degradation, costs, etc. but also based on a possible threat of congestion or any other threat on the QoS-QoE conditions.
- Provide QoS differentiation based on both applications and user profiles and ensure an appropriate scheme of user and application prioritization and differentiation which is not limited to forwarding behaviour but may consider access control as well.
- Split and manage connections (e.g., TCP sessions) over multiple flows inside the network.
- Optimize P2P and massive multimedia transmissions over the network
- Solve the problems of existing combinations of link layer ARQ and TCP (unnecessary TCP retransmission causes unwanted traffic through the network and reduces application throughput and response times).
- Optimal design and efficient management of Content Delivery Networks in an operator's infrastructure (e.g., identify suitable locations for caching, select suitable locations for content, detect unfavourable resource usage, redirect requesting node to alternative resource, etc.)

- Implement efficient offloading techniques, access network/core network elements (re)selection schemes in order to effectively distribute users' data traffic through localized wireless access points (femtocells or WLAN) and to locate service gateways (breakout points) near to those access points (aiming to avoid non-optimal routing and overloading of the network elements).
- Supply switch on / off schemes of networking equipments with traffic management aware decision algorithms.
- Anticipate to apply an intelligent planning process for extending the available resources (i.e., design optimal or near-optimal capacity extension procedures which are able to cope with the enormous traffic volume evolution).
- Enable fast re-active mechanisms based on detection of application and network layer events to accomplish rate adaptation for multimedia streaming application and synchronization with resource management in EPS networks.

Traffic management functions tackling the above challenges usually require access to higher layer user plane data, i.e. IP packets, TCP segments and application layer protocols. Placement of such functions at SGi interface (co-located with PDN-GW) or S5 interface (co-located with S-GW) are possible options, since GTP tunnelling is terminated at these locations. In the following, a brief analysis is done on possible impacts. Positioning of TM functions at S-GW implies that all user plane data can be managed by the considered function unless there is handover between 3GPP and non 3GPP access network. In such case user plane traffic could not be processed or handled by the same node, hosting the TM function. If this can't be avoided, possibly different instances of the TM function have to coordinate in order to ensure continuous TM operation, in case such feature is supported by the TM function in consideration. Positioning the TM function at SGi interface – e.g. co-located with PDN-GW – may cause problems if user data is transferred using different access point names (APN). This usually implies that data paths stretch along different SGi interfaces. It is common practice in currently deployed networks to allocate the same APN to a user for all OTT services. However managed operator services may use different APNs. As a consequence the same TM function may not be used for connection via different APNs. This situation would increase equipment cost (CAPEX) as well as operational cost (OPEX). As a consequence, the suitable location of TM functions depends on mobility aspects (whether a TM function needs to be supported after 3GPP-non3GPP handover) or connectivity aspects (whether the same TM function shall be in usage for services using different APN).

4.6 Network applications and services related challenges

This section describes some of the challenges that new applications and services will bring to the mobile architecture based on their specific needs and requirements.

4.6.1 M2M related challenges

Machine to Machine (M2M) service evolution will set challenges for mobile network functionalities:

- Individual M2M device addressing, global addressing, group addressing, device vendor based provisioning.
- Network selection mechanisms, location tracking, steering of roaming for MTC devices.
- M2M specific properties: MTC group concept, MTC monitoring, MTC time controlled and time tolerant functions, MTC low mobility, MTC small data transmission.
- Charging mechanism specific for M2M communications.

4.6.2 Energy efficiency related challenges

Heterogeneous overlapping networks and potentially more distributed architecture might increase the total energy consumption and might be underutilized with the lower traffic times or unevenly utilized. Then the optimum and controlled resource utilization can provide some energy savings.

Network controlled reduction of energy consumption in the devices for extending the battery life might be challenging, in the heterogeneous radio network environment. Scanning of the possible radio interfaces might be able to be optimized based on the network delivered information about availability of other 3GPP or non-3GPP accesses.

4.6.3 Improved user experience and efficient resource usage

This set of challenges is associated on one hand with poor quality of experience for running multimedia (streaming) applications on mobile networks. Secondly unexpected traffic patterns, like caused by flash mobs and other events may significantly contribute to decline the amount of potentially available resources. Some of the following aspects are not restricted to streaming applications but it is assumed that this class of applications needs a special focus in the project with respect to traffic management.

- How to achieve acceptable QoE for OTT (Over The Top) content (located in external CDN / network)?
    - Some content may not be cached within the domain of the MNO, but inserted from a 3rd party content / CDN provider.
- Detection and localization of high traffic load within local domain or external network:
    - timely detection of problems and proper reaction mechanisms.
- Increasing amount of traffic in upstream direction:
    - usually there is less capacity on the path in upstream direction – some user hosted content might be shifted into the network,
    - some applications (like video conferencing) require QoS support in both directions of flow.
- Align resource selection principles from application with constraints from the network:
    - detection of and reaction to unfavourable selection,
    - how to manipulate resource location information (DNS, etc.) based on resource selection principles,
    - influence resource selection in external network / CDN.
- Improve QoE by caching popular content from Internet:
    - analysis of promising caching strategies especially for mobile access according to content popularity over time, of the possibility for content partitioning and other factors.
- Inefficient content delivery to mobile devices:
    - optimization for unicast and multicast streaming applications,
    - take into consideration capabilities of mobile devices, wireless access and subscriber profile.
- How to go from required QoE for the user to QoS support in the network?
- Support of QoE in roaming case:
    - transfer of content without dedicated QoS control like the one provided in the home network.

5. Proposed Technology Solutions

This section describes the MEVICO Proposed Technology Solutions to cover the Architecture Challenges identified in section 4. Each sub-section includes high level description of the related problem statement and focuses on the aspects that MEVICO project will address.

5.1 Mobility

Facing the traffic evolution trends, higher network throughput and better scalability and flexibility of the core network functions are required as was concluded in the network topology related challenges in Section 4.1. All challenges under the mobility topic described in Section 4.2 are connected to this previous goal. The main challenges are to elaborate appropriate mobility management and path selection strategies facing the foreseen trends of traffic demands and user behaviours. This topic focuses on user terminal and EPC element aspects. The proposed solutions in the focus of this project for the above mentioned challenges are the following.

Smart traffic steering

Smart traffic steering with multi-access terminals and multipath protocols will enable better load distribution considering user, network and application preferences. The functions needed for smart traffic steering are the followings.

Smart traffic steering decisions: the most important selection problems considered are access interface selection, gateway selection, source address selection during terminal attachment and session establishment. For the support of terminal and flow mobility, mobility anchor selection in a distributed mobility management scenario requires novel classification algorithms as well. Enabling technologies investigated in the project will be the IEEE 802.21 Media Independent Handover protocol which provides a framework for transverse information services, physical and link layer resource monitoring, reservation and release. The 3GPP Rel-10 Access Network Discovery and Selection Function (ANDSF) describing the access interface selection policies must be further improved to provide an optimized set of rules to the UE.

Multipath technologies: the Multipath TCP (MPTCP) can transmit one TCP flow over multiple interfaces, and can balance the load between subflows. Stream Control Transmission Protocol (SCTP) supports multistreaming, i.e. several streams related to the same application can be handled by one SCTP stream, and backup SCTP associations. Both technologies will be analyzed and further improved to enable multipath communication.

Flow mobility: The performance of 3GPP Rel-10 IP Flow Mobility (IFOM) will be evaluated. IFOM enables smart IP flow allocation.

Offloading techniques: offload the EPC and LTE through non-3GPP networks could further improve the overall network throughput and quality of service. Access offload through IEEE 802.11 managed by the 3GPP operator will be evaluated.

Distributed and dynamic mobility management

These solutions cover terminal and flow mobility, and reachability of multi-access devices on L2 and IP level. The technologies developed in this project aim to achieve the following properties: increase network throughput by the support of a dynamic activation of mobility signaling and by providing distributed, anchorless or partially anchorless solutions.

Mobility management technologies include Session Initiation Protocol for SIP-based services, and SCTP for non SIP-based applications. These technologies basically can provide end-to-end, anchorless mobility, but they will be applied in a flat or distributed approach in the EPS.

Proxy Mobile IPv6 (PMIPv6) will be extended with route optimization procedure among the Mobility Anchor Gateways.

The project also covers how to adapt the Host Identity Protocol to provide distributed mobility management in EPS. HIP by default follows an end-to-end approach, hence could provide an anchorless solution.

A new Ethernet-level mobility management solution will be developed and evaluated, that could replace the GTP concept of EPC by Ethernet VLAN tunneling, hence reduce the overhead.

Evolution of the current 3GPP based model (GTP tunnels) with the dynamic and distributed mobility principles will be studied.

An anchorless mobility solution for TCP sessions called NMIP (TCP rehash) will be evaluated.

DMA (Distributed Mobility Anchoring) has been initially discussed in IETF to improve MIP/PMIP by distributing mobility anchors and use as much as possible a local, not tunnelled addresses, see also DDMM.

Here the technology intends to optimize the EPC based on the ideas of the DMA, but utilizing existing 3GPP protocols like GTP with as less as possible changes, to enabling SW upgrades to optimize the usage of existing resources. An underlying assumption is that a GW distribution brings certain benefits. A proposal is to change PGWs using intelligence in the PGW or changing SGWs for routing optimization.

The first DMA solution applies after a UE has moved into a new “gateway area”. The PGW selects IP (PDN) connections for what a new IP address and service interruption may be acceptable from application point of view and forces a reconnection that allocates a new more optimal PGW and new IP Address. This leads to more optimal routing and savings in transport networks.

A second DMA solution proposes to relocate the SGW to achieve maximal SGW-PGW collocation in a distributed architecture when UEs use different PDN connections. This saves at the end GW capacity. Different gateway locations may result from the fact that a UE may connect to local and/or central networks or Internet providers.

Access and network security

The existing user access security procedures must be revised, and the communication protocol might be further optimized to the new distributed EPC architecture, aiming to reduce overprotection and decrease L2 and L3 re-authentication times during handover. The investigated technologies will be the Internet Key Exchange v2 protocol, HIP and HIP Diet Exchange that is a lightweight version of HIP.

Bootstrapping

Configuration of multi-access terminals might lead to conflicts in case of parameters that have wider than interface-level scope. These conflicts must be discovered and resolved. ANDSF policies will be investigated from that aspect.

5.2 Network Transport

The next billion Internet users will connect primarily through mobile networks. Therefore, mobile networks have to support constant growth of traffic and increase the throughput from 1Gbps to tens or hundreds of Gbps already in the near future. Ethernet-based technologies have several features that make them especially interesting. Therefore, Ethernet is a natural solution for replacing legacy SDH and other older technologies, and the energy consumption of L2 switching is an order of magnitude lower than IP routing.

The proposed solutions to cover the above mentioned research paradigms that will be deployed in this project are the following:

Carrier Grade Ethernet with inbuilt O&M

The objective is to provide Carrier Grade Ethernet and overcome the limitations of using Ethernet for large scale networks. In order to provide reliability and robustness required for Carrier Grade Networking an O&M mechanism is required. Therefore, the goal is to enable routed based Ethernet where the O&M functions will provide the necessary routing optimizations and bootstrapping algorithms, as well as the link break detection and route recovery mechanism.

Ethernet Mobility to the Edges based on TRILL

TRILL leverages IS-IS routing protocol to achieve Ethernet frame shortest path routing with arbitrary topologies. In this research item the goal is to utilize TRILL extended with DHT to deploy mobility in the network edges. The goal is to combine the advantages of bridging and routing and fully distributed mobility mechanism implemented in the Link layer (i.e. Ethernet). In order to increase the available throughput we consider that is necessary to move towards lower layer switching and minimize processing per packet.

Consumer Edge security

The unwanted traffic is one of the reasons for inefficient usage of resources (i.e. radio spectrum, routers, bandwidth, etc). The spam or unwanted traffic has to be filtered from the sources to avoid waste transmission (unnecessary data, resources usage). In this objective we propose to deploy Consumer Edge Security element that will enable the setup of end to end trust connection for traffic where the sources has been verified.

Automatic and Secure L2 Virtual Private Networks (VPNs)

Virtual Private Networks (VPNs) are popular among network providers that wish to separate multiple LAN domains across a single network infrastructure. One VPN technique is Virtual Private LAN Service (VPLS), a layer 2 (L2) solution that connects several physically separated LAN segments in to one logical LAN segment, i.e. emulated LAN or VPN overlay. This research item is interested in investigating so called “bump-in-the-wire” customer VPLS solutions in which the VPN service is overlaid on top of a provider network combined of IPv4 and/or IPv6 hybrid segments. In particular, the research item studies how identities can be utilized to mutually authenticate the PEs as belonging to a certain overlay and facilitating the renumbering of the PE devices.

Wireless mesh networks for mobile backhaul first mile access

With the introduction of WiMAX and LTE the need for mobile backhaul transport capacity grows rapidly, to the level of Gbps. Fibre media is able to provide the high data rates but fibre is not available everywhere either for technical or commercial reasons. Additionally, more base station sites, with different cell sizes, must be provided to meet the capacity and coverage requirements. Therefore, new wireless solutions are needed for the backhaul, especially in crowded areas characterized often by lack of available frequencies. One feasible solution providing sufficient transport bandwidth and capacity is E-band (71-76 GHz, 81-86 GHz) microwave radio with Ethernet connectivity. A wireless mesh backhaul can be used for small cell, high capacity base station first mile access and for other high capacity packet connections (e.g. office and home access), traffic management.

Relaying

Relaying techniques are considered as an alternative solution to enhance capacity for the cellular networks, to extend coverage in specific locations, to increase throughput in hotspots and to overcome excessive shadowing. It gives important advantages such as ease of deployment and reduced deployment costs and decreased output power compared to deploying regular Base Station (BS). Moreover, there is no need to install a specific backhaul in the network. It is an important aspect and one of the key technologies taken into consideration during the standardization process of 4G technology LTE-Advanced. RNs are also envisioned to be transparent to UE. In other words, the UE is not aware of whether it is connected to RN or a conventional base station. This ensures backward compatibility with previous LTE releases 8/9. Therefore, gradual introduction of relays without affecting the existing structure of UE's can be ensured.

Relaying promises coverage-area extensions and high data rates for the cell edge users. This is especially useful because LTE will operate on high carrier frequencies, i.e. 2.6 Ghz which will result in ultra-dense deployment of network nodes, the transmit power is limited when transmitting broadband at the cell edge and the most of the traffic is generated indoor. It can also be used as a capacity improvement with load balancing and cooperative relaying techniques.

Current relay architecture in 3GPP LTE Release 10 assumes fixed relays. However, handover of a relay from one donor eNodeB to another donor eNodeB should also be supported in future network architectures and releases which will be a consequence of mobile relaying.

Radio over Fiber (ROF) for enhanced macro coverage

In the ROF technology light is modulated by a radio signal and transmitted through an optical medium. The optical medium is more robust against several impairments, such as attenuation losses, electric discharges etc. The optical transmission technology can support any bit-rate the current and future wireless systems may offer. Fiber To The Antenna (FTTA) is one of the key applications of ROF that will enable enhanced macro coverage where base station and antenna has to be separated by a long distance.

5.3 Traffic Management

Section 4.5 introduces the main traffic management related challenges which are connected to the MEVICO goals and motivations. In order to tackle these challenges, techniques operating within different traffic management building blocks must be considered.

- First, mechanisms with the primary objective is to improve performance of individual flows based on application type, user profile and other policy related information must be incorporated. Such solutions are belonging to the microscopic traffic management (1) building block.
- Second, the macroscopic traffic management (2) must also be introduced in the network with the primary objective to improve efficient usage of network resources. Parameters for optimization in this case describe traffic patterns without detailed knowledge of individual flow attributes.
- In addition to microscopic and macroscopic traffic management, a third group is improved resource selection and caching (3). The associated mechanisms address the selection of resources in distributed data management systems (P2P, CDN, caching), if necessary. This building block may rely on services of both microscopic and macroscopic traffic management. These could be in place without dependence to other traffic management building blocks. Cross-layer P2P is a novel technique where the ISPs can have control over the non-optimized

P2P traffic. Proactive Network Provider Participation for P2P (P4P) is a promising solution to non-optimized and self-organizing P2P.

- The fourth building block is called as application supported traffic management (4) which tries to optimize performance from end user perspective of certain, widespread applications (e.g., based on CDN and P2P) without getting support from network elements.
- The fifth building block is more relevant from business perspective without too many technical aspects: steering user behaviour (5) is mainly used by network operators and by possibly other stake holders as well in order to influence user behaviour by defining certain constraints for usage of networks / services and certain incentive to comply with the usage constraints.
- The last building block in this enumeration is about capacity extension in case the available network is regularly in high load conditions. It is the challenge to apply an intelligent planning process for extending the available resource (6). In addition to the building blocks there are some common functions like policy control and traffic monitoring.

Even though the above building blocks and the associated mechanisms / possible technical solutions should represent a functional decomposition of traffic management in EPS on high level it is assumed that some of the mechanisms to some extent are dependent on each other. It is one of the most important efforts of MEVICO to map the diverse mechanisms into functional components for the traffic management architecture and elaborate the dependencies between the building blocks and the containing functional components based on the various design options that have been identified.

5.4 Network Management

As indicated in section 4.4, the network management related challenges identified by MEVICO are mainly: avoiding conflicts when introducing SON features, finding solutions to scalability and heterogeneity requirements and managing intra 3GPP handovers and handovers between 3GPP-non 3GPP accesses for optimizations and load-balancing.

The common characteristic of all the alternative/complementary transport solutions is that, in each case, long-lasting connections (OC-x, LSP, EVC) are configured between the EPC nodes and that QoS schemes are applied either at the IP/MPLS or CET layer. These connections are configured either manually or via management tools that provide some level of automation when, for instance, the network is extended with new eNBs. In addition to the route/path of the connections and parameters needed for the connectivity (e.g., VLAN tags, IP addresses), other technology specific transport level parameters must be configured to define the amount of reserved/granted resources (as in the case of CET: CIR, EIR, CBS, EBS) or the level of service granted to a specific traffic class: scheduling weights, buffer allocations, etc.

To cope with the network management issues in EPC and heterogeneous networks, several topics and technologies have been identified that need to be addressed to make management more efficient.

Managing heterogeneity:

Heterogeneity involves managing, using the same management system, different co-existing network technologies; for instance, CET/DWDM, IP/Ethernet/NG SDH and also different radio technologies sharing the same transport network, such as HSPA, HSPA+, LTE, LTE-A. In order to maintain heterogeneity in the network an open standard layered network architecture for co-existing network technologies can be introduced.

Adapting EPC/Network Management to LTE-A features:

LTE-Advanced introduces several new features for enhancing the peak rates and service quality experienced by the user. Managing these features requires more strict synchronization and lower delays in the network. The scalable system bandwidth would put the requirement for more flexible resource allocation solutions and new management solutions might be required, for instance, to avoid introducing serious load on the X2 interface.

To validate that SON functionality and policies are working correctly it will be necessary to examining SON related messages from the S1-MME/S10/S11 interfaces to determine that the appropriate Network Elements are selected by the eNBs and S-GW. For validating the SON functionality and eNB algorithms and for optimizing the core element usage, normal signalling KPI’s shall be used to determine if the network is equally loaded.

SON features in radio and transport Network Management

SON in LTE can efficiently improve the management and resource utilization of RAN but it is necessary to investigate the impact of radio SON features on transport network management and to find efficient global solutions.

The Mobility Robustness Optimization (MRO) SON feature aims, first, to reduce the number of handover-related radio link failures (Too Early HO (handover), Too Late HO) and, second, to automatically adjust the HO parameters to avoid incorrect HO parameter setting that can lead to inefficient use of network resources due to unnecessary or missed handovers. For this, neighboring eNBs need to exchange certain information, e.g., through Radio Link Failure (RLF) reports.

Load balancing algorithms for EPC

The Load Balancing (MLB) SON feature aims to dynamically and automatically balance the traffic. The different hand-overs should be considered: hand-overs between eNBs, hand-overs between eNBs and HeNBs, inter-RAT hand-overs (e.g., LTE – HSPA, LTE – WiFi). An eNB must know its own load and the load of its neighbouring cells. This information is exchanged through the X2 interface.

Network monitoring for EPC

To be able to correctly manage the networks using classical or SON techniques, precise information is needed at all times on the state of the network and estimations of traffic evolution for different types of traffic should be rendered possible. For this, network monitoring needs to be adapted to EPC and heterogeneous network constraints and several topics need to be addressed including : traffic analysis and capturing performance, time-stamp accuracy, protocol stack support, interface requirements and SON support as well as satisfying all the dependability requirements.

As defined in WP5, the main monitoring techniques that will be used in MEVICO are:

- **End-to-end monitoring** to evaluate the QoS/QoE of applications and services,
- **Deep Packet Inspection (DPI)** for the identification and the classification of protocols and applications, and,
- **Monitoring of SON activities** for both testing the SON features and verifying them during operation.
-

Energy saving and impact on network monitoring and management

The Energy saving needs to be taken into consideration and, from a monitoring point of view, it is very important to provide network measurements to optimize the energy saving policy, and test that the implemented energy saving policy behaves as expected.

5.5 Network applications and services

5.5.1 Network functionality virtualization and realization with cloud computing

Scalability and optimization of the mobile network architecture for the high traffic demand are the major challenge in the future. Virtualization and cloud computing methods have showed their potential in IT industry, like data center applications and have potential to be utilized to functionalities of mobile networks. Potential use cases are in mobile networks virtual operator concepts, network sharing principles and core network user plane distribution related virtualization and centralizing of the NW control overlay functions. This could enable better resource utilization, E2E QoS policy control and Heterogeneous Network (HetNet) control e.g. for load balancing.

5.5.2 Network Energy Efficiency improvements by efficient capability utilization

There is a need for optimizing the efficient usage of the mobile networks, because the number and the capacity of the network elements increase due to high capacity demand.

Depending on the changed user activity (like time of date, weekend etc.) some redundant capacity of the network can be switched off (like hot spot layers). The control and optimizing this functionality needs some more study, where also the network control virtualization could help.

In core network energy efficiency could be achieved by flattering the network protocol stack i.e. removing some of the network layers and use transport layer directly (i.e. Ethernet), to reduce the processing per packet. An improved traffic engineering to reduce the traffic to flood the network unnecessarily also would help in the energy efficiency.

The mobility management could be optimized, by reducing paging areas according to the terminal based parameters, such as mobility profile. Analyze the tradeoffs between location update and paging based on network hierarchy structure and where to store the paging information is needed.

RAN energy efficiency is more important issue due to high amount of the nodes, but this is considered in other projects than MEVICO.

5.5.3 Network efficiency improvement for Video/Multimedia Applications

Video content delivery will set the biggest challenges to the mobile network scalability due to its high demand for bandwidth.

The further evolution need is studied for the mobile network architecture to support CDN, torrent-type of delivery and other intelligent data caching methods. It is important to identify the optimum location of CDN servers within the EPC network elements and possible impact to the architecture. These together with converging networks (fixed, mobile, home, enterprise) might indicate evolving sources of the data and thus needs for changes in the network topology and thus for architecture optimization.

It is necessary to look at transfer optimization protocols that are specific to information centric applications to enable controlling of traffic flows/content delivery according to content and access availability. One of these is the IETF Application Layer Transport Optimization (ALTO), which needs some further adaptation to wireless networks, where maintaining the QoE is even more challenging. There is also related work ongoing in EU Future Internet (FI) projects (like SAIL).

5.5.4 Network improvement for M2M Applications

The foreseen high increase of the Machine-to-Machine (M2M) type of devices and applications might cause impacts or improvement needs to mobile network architecture or functional requirements. Most of the identified challenges are investigated in 3GPP Releases 11 and 12 work items NIMTC and SIMTC as well as ETSI Technical Committee M2M work.

The M2M application special characteristic issues to be studied for network architecture evolution:

- Improvements on network/interface selection mechanisms, location tracking, steering of roaming for MTC devices – Possibilities for signaling optimization
- The network impacts of the M2M specific properties: MTC group concept (MTC Devices that are co-located with other MTC Devices), MTC monitoring, MTC time controlled and time tolerant functions, MTC low mobility, MTC small data transmission.

5.5.5 Application based network traffic analysis and engineering

In order to improve network QoS and application QoE , there is need for modeling the selected Internet applications traffic characteristics and their adaptive behavior at times of congestion. Based on this input there are two types of traffic engineering mechanisms to be investigated:

- The macroscopic traffic engineering relates to adaptive routing, gateway selection mechanisms, multi-path transmission and mobility support, described in more details in Mobility chapter.
- The microscopic traffic engineering relates to mechanisms for rate reduction of active traffic flows under QoE constraints, based on the behavioral models of application operation, QoS mechanisms, measurement and control functions.

The application generated traffic mix observation at the short time scales is needed to test Traffic Engineering solutions, as well as for admission control of traffic for short term network dimensioning. Target is to utilize method to derive short term traffic mix estimations from long term ones given as input. The proposed method is non-linear unlike usual methods and thus more reliable.

5.6 Network Topology

From the perspective of operators, the challenges in the coming years are to be able to apply the required network evolutions for becoming mobile broadband integrated providers.

The current networks topology is centralized which creates a set of bottlenecks in the communication with the servers that handle mobility, service provisioning, etc. The objective is to propose and describe different architecture scenarios over the same network topology model and to study different protocol scenarios to be compared under different angles, for examples:

- Handover performances (establishment, delays)

- Flexibility in integrating and making use of various access technologies (typically LTE/WiFi),
- Packet transport,
- Economical comparison according to traffic forecasts.

Different options for enabling multipath through Heterogeneous accesses and optimize the multipath management, should be analyzed. Moreover, the various transport protocols (GTP, MIP based, other …) and the load balancing with heterogeneous access should be studied.

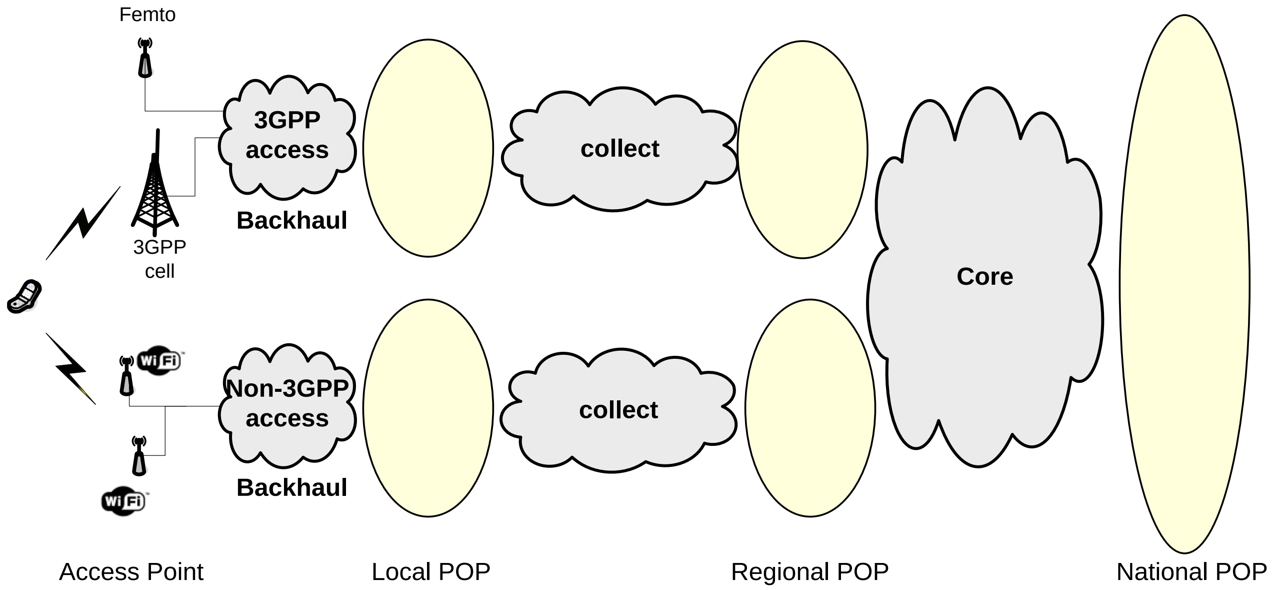

Figure 4 Network topology model

The architecture scenarios are divided into three families, depending on the distribution level of EPC and IMS nodes in network topology model presented in Figure 4.

- Scenario 1: progressive distribution from the national PoPs to the regional PoPs
- Scenario 2: progressive distribution from the regional PoPs to the local PoPs
- Scenario 3: maximum distribution to the Access points on the antenna sites

6. Architecture Approaches

The results of the studies done in MEVICO related to mobile scenarios and traffic demands show that there are several parameters that will cause network scalability and performance problems. Some of the most relevant parameters are related to changes in traffic - growth in data volume (i.e. increase by 3-10 times by year 2020 at busy hour) and number of subscribers (i.e. increase of mobile broadband subscribers by 8-12 times by year 2020), and heterogeneity aspects: multiple traffic patterns handling (e.g. MTM), multi-technology smart and seamless support, including mobility.

Based on the new traffic scenarios and the growth of broadband users a set of requirements and challenges have been identified in section 3 and section 4. The requirements and challenges are grouped under three high level objectives i) backhaul capacity improvements, ii) improving the usage of heterogeneous network resources, and iii) improving core network capacity.

MEVICO proposes a set of technology solutions listed in section 5 to address the traffic and user growth demands as well as the requirements and challenges identified in section 3 and 4. This section proposes a set of architecture approaches to address the above mentioned three high level objectives. For each high level objective three topology evolution scenarios (centralized, distributed and flat) are considered. Finally the new technology solutions proposed in section 5 are mapped to the appropriate topology models, indicating the functionality enhancement location in the existing network elements or need for the new elements. Rationales about the benefits and performance impacts as well as possible co-existence of the technologies are also discussed.

Note: the extensions to MEVICO architecture approaches assume that IP layer will persist in end host (and part of network elements) and IPv6 will be adopted in UEs. Due to the uncertainty of future technology development and network deployment scenarios a set of open and contradicting trends can be foreseen. These include questions such as whether the separation between backhaul, aggregation and core network will persist or will they converge into single transport network. Will multiple radio interfaces converge into a single entity in the eNodeB? Will the SON functionality of network management be able to handle large number of Femtocells? And will the WiFi/fixed convergence happen or will it be blocked by regulation? And will it drive the need for separate backhaul or core networks to handle the offloaded traffic separately?

Trends in EPC functionality future requirements

Networks need to enable capacity, coverage and services where needed

- Whole communications ecosystem has become very dynamic and the new demands might emerge quickly
- Core functions need to become self-aware and self-adapting to react to and fulfil user expectations dynamically to deliver services and content, and dynamically provide the capacity needed

Intelligent/integrated traffic management need to ensure that the correct QoE is achieved throughout the network to manage the excessive traffic explosion

- Ensures that the available transport and radio network resources are used the most efficient to provide the best possible customer experience at all times
- QoS methods, intelligent data caching/CDNs and co-operation with the current network status (like Policy Control)

HetNet (multiple technologies, frequencies and cell layers) needs to be optimized to be managed according to current service need

- SON functionalities are essential to enable automation in resource utilization optimization – and SON needs to operate E2E (not only sub-optimizing)

Architecture Future trends

MEVICO architecture will address the traffic demands and requirements identified based on current forecasts with a set of technologies currently under development. However, a set of future technologies not covered in MEVICO project scope will also influence the future architecture. The most important trends are the following.

- Hardware commoditization: Several EPC functions are able to run on single generic telecom hardware, due to SW virtualization.
- Cloud computing: Some network components/functions will be moved to commodity data centers.

- Transport simplification: SW defined networks appear to be a mechanism to dynamically change network infrastructure.
- Optimized network monitoring for intelligent support of QoS (Celtic Ipnqsis Project [8])
- Deployment of mobile networks for efficient transport of video applications (Celtic MEDIEVAL project [10])

6.1 Topological models

This section defines the following concepts of centralised, distributed and flat architectures as baselines for the different topologies that MEVICO architecture approaches will consider when addressing the three high level objectives.

6.1.1 Centralized architecture

The centralized architecture is considered as the current 3GPP Rel.10 architecture with the enabled functionalities (S-GW localization, SIPTO offload…). In the centralized architecture S/P-GW and IMS components are located in the National PoP.

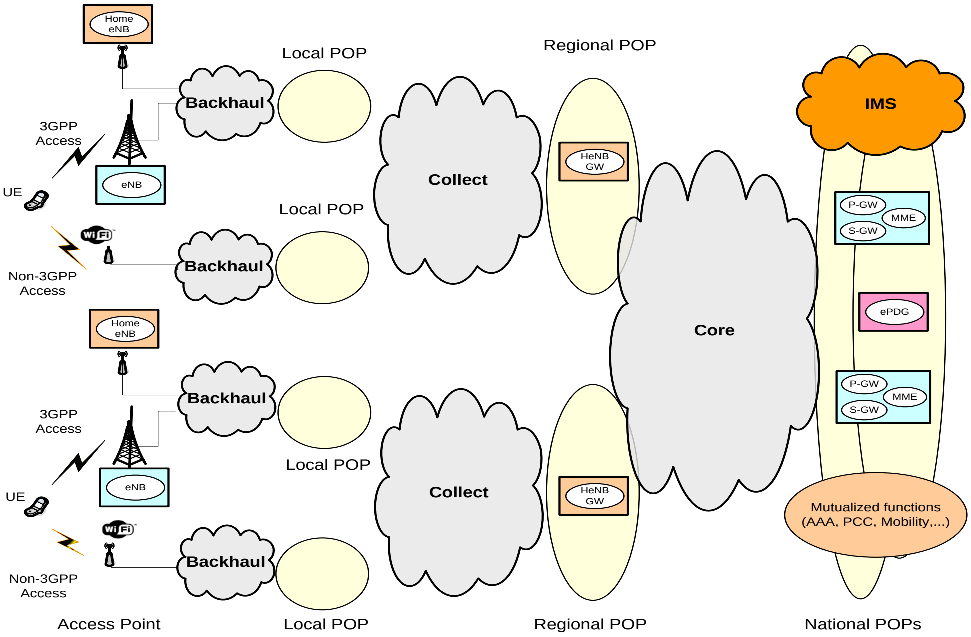

Figure 5 Centralized architecture model

6.1.2 Distributed architecture

The distributed architecture will include multiple gateways (S/P-GW functionality) located in the regional POPs. The distributed architecture is assuming that the functionalities are enabled to be distributable (optimized, and other EPC functionalities might still be centralized).

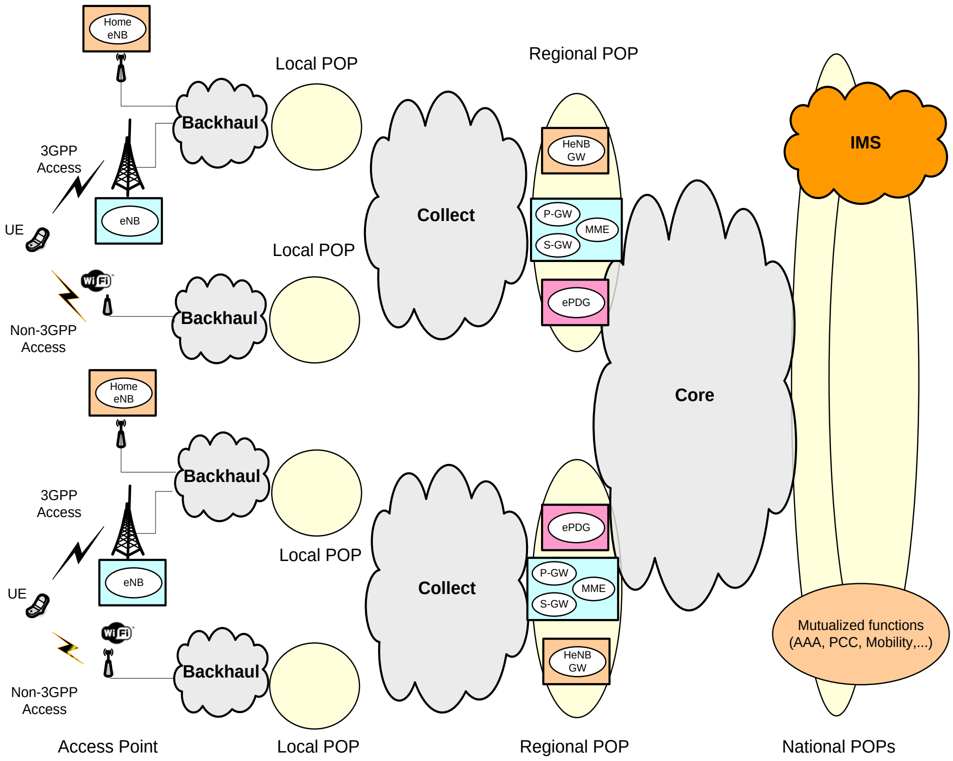

Figure 6 Distributed architecture model

6.1.3 Flat architecture

The flat architecture also referred as ultra flat architecture consists of the architecture where S/P-GW, MME and possibly (part of) IMS functionalities are in the local PoP. The legacy IP routed network is expected up to the eNodeB side (it will be operator controllable).

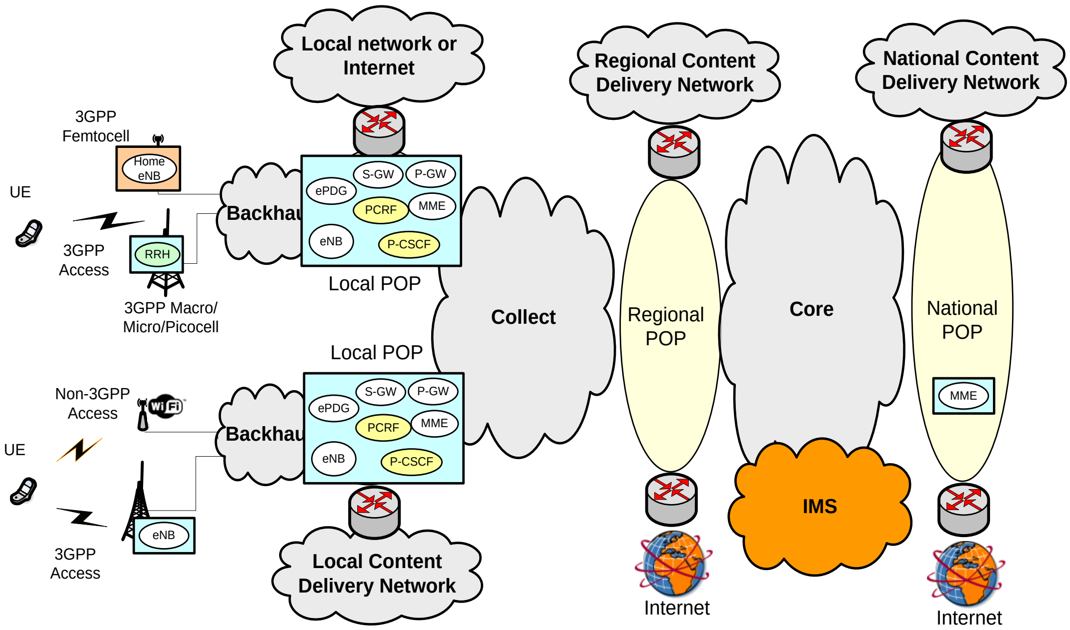

Figure 7 Flat architecture model

6.2 Transport Architecture approach

The MEVICO architecture will address different layers from transport up to networking, traffic and network management. The transport architecture will provide service to the upper layers and should address the three high level objectives a) improve backhaul capacity, b) improve the usage of heterogeneous network resources, and c) improve core network capacity. In this section the transport architecture is described for centralized, distributed and flat topology. The proposed distributed and flat architecture moves functionality to the edge which might allow simplification of transport network. The end result might be the merging of core and aggregation networks into single network. This topology change is assumed to make the aggregation network to move towards more meshed networks in order to enable faster mobility and traffic between eNodeBs (i.e. X2 interface). The end result might be the logical merging of core and aggregation networks into single network. Therefore, the logical merging of aggregation and core network where the hierarchical distribution might disappear but physical deployment of aggregation and core tiers will remain. Additional short-cut connections are needed to connect together the local aggregation networks.

Moreover, mobile networks are moving towards harmonized transport network based on Ethernet, MPLS-TP, L2/L3 VPN and Pseudowire. The proposed technologies for improving the transport architecture are TRILL, Wireless Mesh Networks (WMN), HIP-Femto and HIP-VPLS, which are described next.

6.2.1 Rationale for using the involved technologies

6.2.1.1. Wireless Mesh Network (WMN)

WMN is a high bandwidth communications network made up of point-to-point communications links organized in a mesh topology providing a virtual transport service for a set of eNBs. The technology is not sensitive to the EPC topology scenarios presented in this document. It is compliant with lots of transport technologies and combinations of technologies. It is a complimentary solution to the existing wireless and wireline backhaul access solutions for LTE and LTE-A.

WMN technology provides many economic and technical advantages for backhauling LTE and LTE-A base stations. The level of utilization of the transport resources can be greatly improved within the mesh coverage area. Overall data throughput and transport connectivity is increased by sharing transport capacity flexibly between the client nodes in the mesh network. Other advantages include operational easiness, high reliability and flexibility/scalability to adapt to traffic fluctuations and network changes since the network throughput can be dynamically and autonomously optimized. The technology enables a flexible way to enlarge and build the network according to the transport capacity need. It also enables horizontal connections between base stations for fast X2 connections. The in-built SON features simplify deployment & installation, maintenance and network management processes.

6.2.1.2. Transparent Interconnection of Lots of Links (TRILL)

From a transport point of view, Ethernet-based technologies have several appealing features, notably it allows increasing the available throughput by moving towards lower layer switching and minimizing processing per packet. TRILL combines the advantages of bridging and routing and fully distributed mobility mechanism implemented in the Link layer (i.e. Ethernet). TRILL enables handling mobility in Ethernet when nodes are moving between eNodeBs, which reduces the amount of mobility requests that have to be handled in upper layers (i.e. IP, Transport or Session layers).

TRILL reduces the signaling traffic and reduces the latency to manage mobility, thus increasing the overall capacity of the backhaul network.

6.2.1.3. Carrier grade Ethernet with inbuilt O&M (O&M)

O&M provides bootstrapping mechanism where optimal routing information is transferred to the forwarding elements (i.e. Ethernet switches). This reduces the delay in finding new routes in case of link failure, thus improving the reliability and capacity of the core network.

O&M aims at providing the minimum recovery delay necessary to provide carrier grade transport in the core network.

6.2.1.4. Host Identity Protocol for Femtocells (HIP-FEMTO)

The HIP-based L3 handover procedure has the following advantages for 3GPP architectures: in case of flat or distributed 3GPP architecture, a considerable part of the operator traffic will go through untrusted IP networks. HIP provides automatic and flexible VPN configuration to securely reach femtocells and distributed GWs in the access networks.

6.2.2 Centralized Transport architecture

The access networks are normally connecting the eNodeBs to the Local PoP and they are typically based on star, point-to-point or small ring structures and in the future also on partial mesh topologies. The physical deployments could be based on Ethernet, IP/MPLS, microwave radios, PON or xDSL among others.

The aggregation networks are connecting the local PoPs to the regional PoP and they are based on hierarchical ring structures in most of the cases but also mesh topologies are used. The physical deployments could be based on Ethernet, IP/MPLS or SDH/WDM.

The core networks are connecting the regional PoPs with the national PoPs and they are based on mesh networks most of the cases. The physical deployment is based on IP/SDH, IP/MPLS, Ethernet or SDH/WDM.

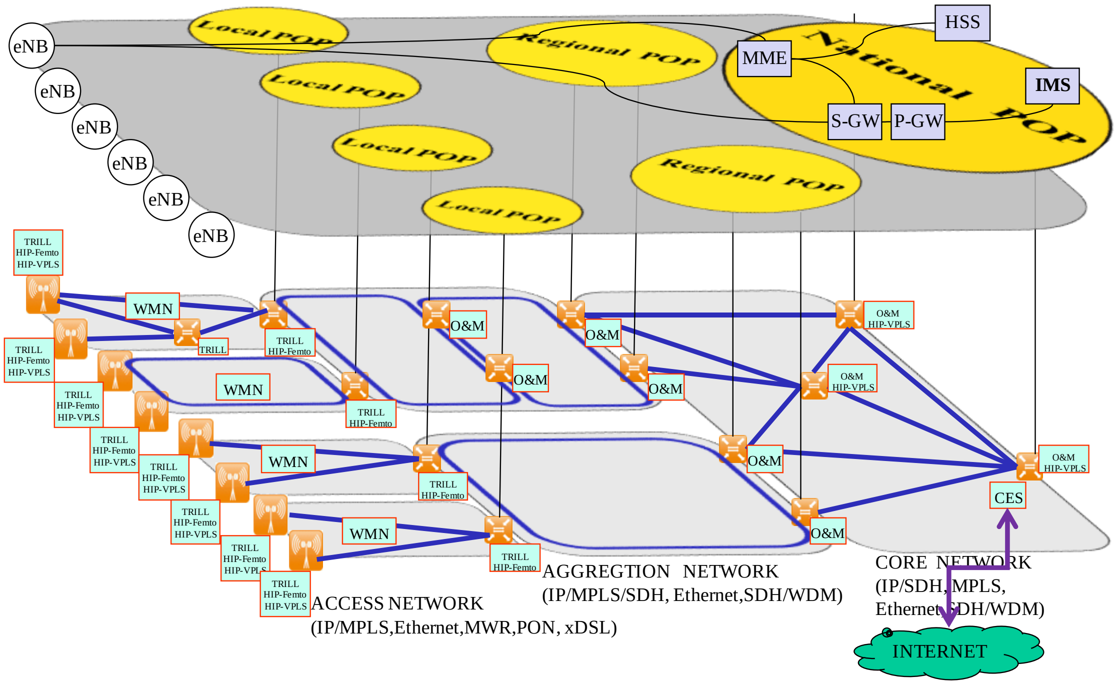

Figure 8 Centralized Transport architecture

6.2.3 Distributed Transport architecture

The transport network might require additional connectivity between Local PoPs to allow the traffic to be offloaded to public Internet.

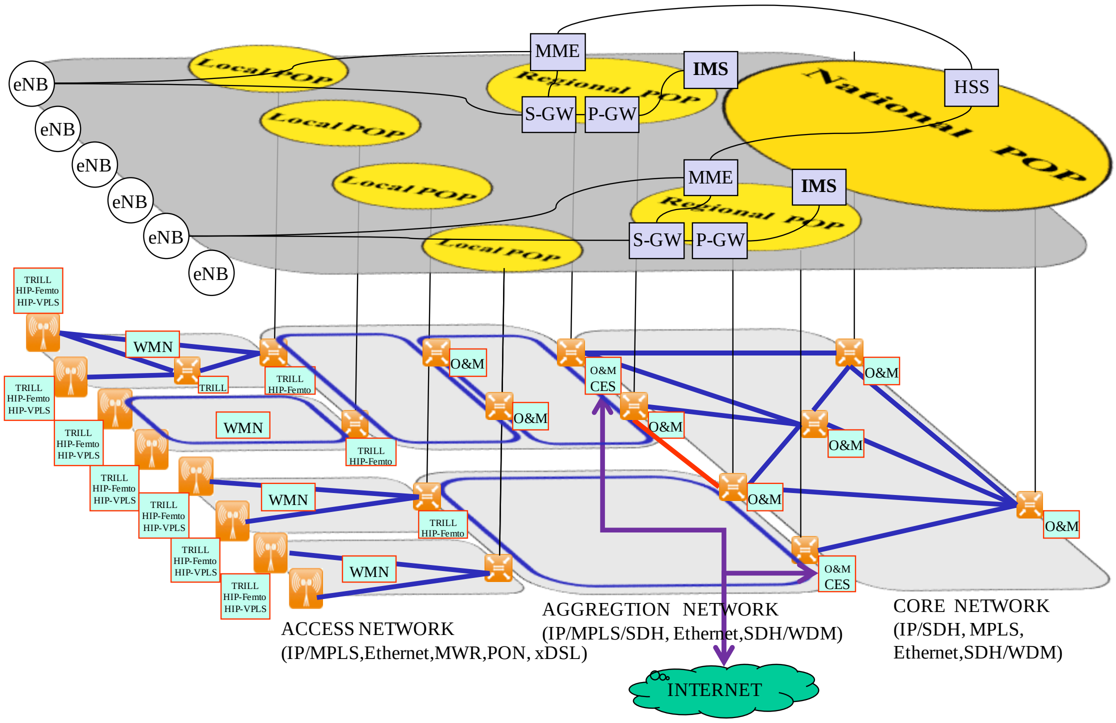

Figure 9 Distributed Transport architecture

6.2.4 Flat Transport architecture

In case of flat topology the transport network might require additional connectivity links between regional PoPs to allow traffic offloading to public Internet. The need of full mesh topology in the aggregation network might end up in the merging of aggregation and core networks at logical level.

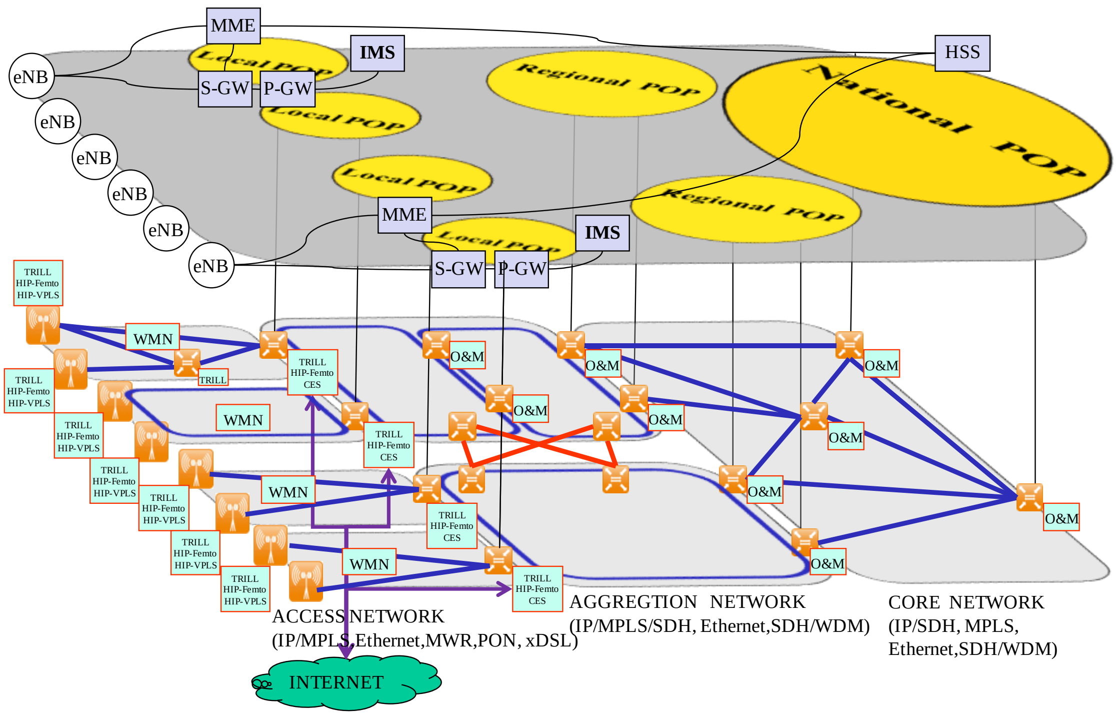

Figure 10 Flat Transport architecture

6.3 Roaming problematic

International roaming is a cornerstone of mobile networks. MEVICO new architecture proposals, specifically the distributed/flat approaches, raise some questions about it. In the following diagram, you will find a state of the art of the 3GPP roaming reference points depending on the various potential cases, knowing that the standardisation is not complete so far, and neither the operators' usages.

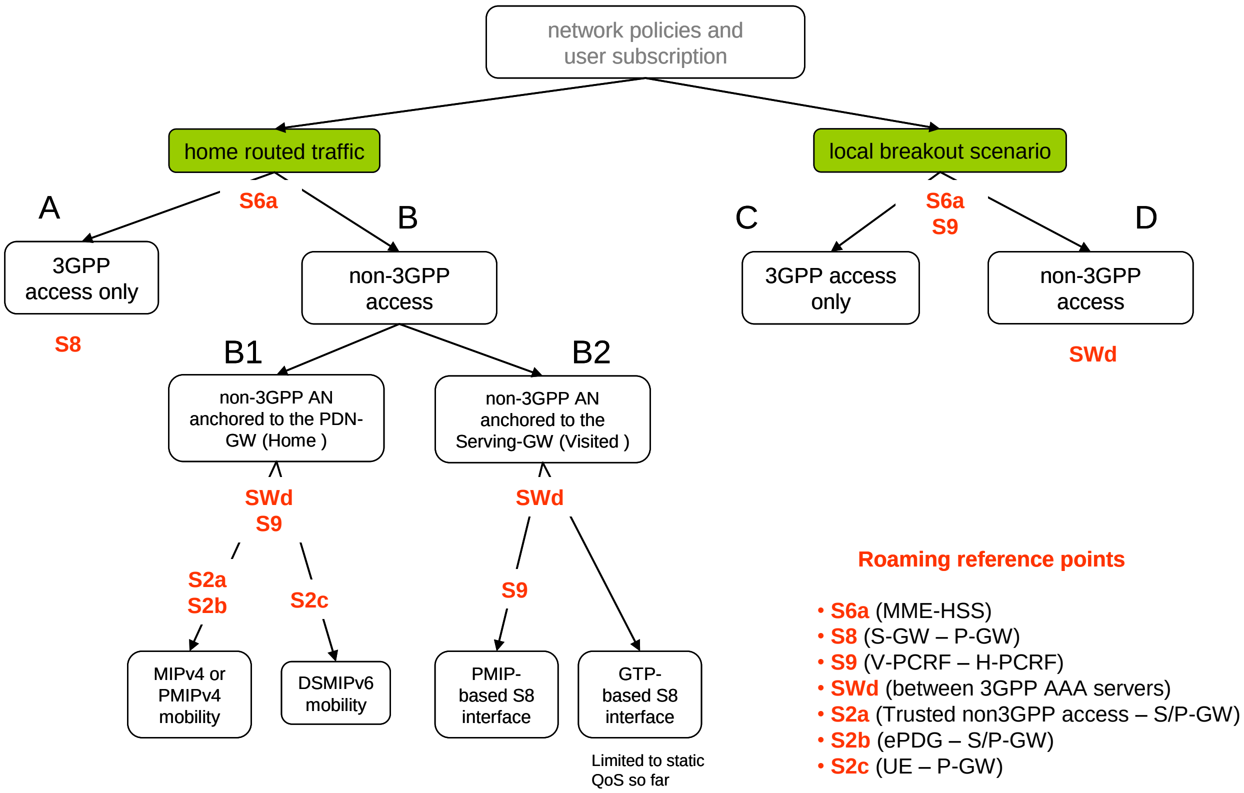

Figure 11 Roaming reference points

For a certain number of cases, including the abroad local calls and the web browsing, the local breakout scenario suits better for a question of optimised routing. Nevertheless, as long as the operators will provide walled garden services, the home routed scenario will be needed, to the cost of a suboptimal routing. A mix of both scenarios could apply depending on subscriptions, services, involved operators' policies and agreements between the operators.

As a first approach it seems the main point is to minimize the number of roaming reference points. That is why case B1 above is not recommended at all.

In all the cases, the MME distribution appears to be an issue.

In addition for 3GPP accesses, in case A, the serving-GW distribution is an issue. In case C, it is the PCRF distribution that is a challenge.

For non 3GPP accesses, both cases B2 and D show that PCRF distribution is an issue. 3GPP AAA servers are not concerned by distribution.

This requires further analysis to be done when evaluating the different architecture options and the evolution of the network topology.

6.4 Architecture approach for backhaul capacity improvement

The MEVICO architecture will improve the backhaul capacity as the mean to address the traffic and user demands as well as the challenges and requirements identified in the preliminary studies. The architecture will address the improvement of capacity and efficient utilization of the access part of mobile networks (i.e. backhaul) by reducing the signaling overhead, improving the capacity of the physical transport available as well as improving the reliability. Moreover, the MEVICO architecture will enable load balancing (with WMN, SON), will provide an efficient resource allocation (SON, GW selection) and will enable the traffic offloading/ FemtoCells. This section will define three architecture approaches based on centralized, distributed and flat topology deployment. MEVICO is proposing following technologies to improve the backhaul capacity; WiFi, TRILL, Mobile Relaying, WMN (Wireless Mesh NW), GW selection, MTM (Microscopic traffic management), MCCS (Multi-Criteria Cell Selection), SON and HIP/Femtocell.

6.4.1 Centralized architecture

Functional view of centralized architecture for backhaul capacity improvements

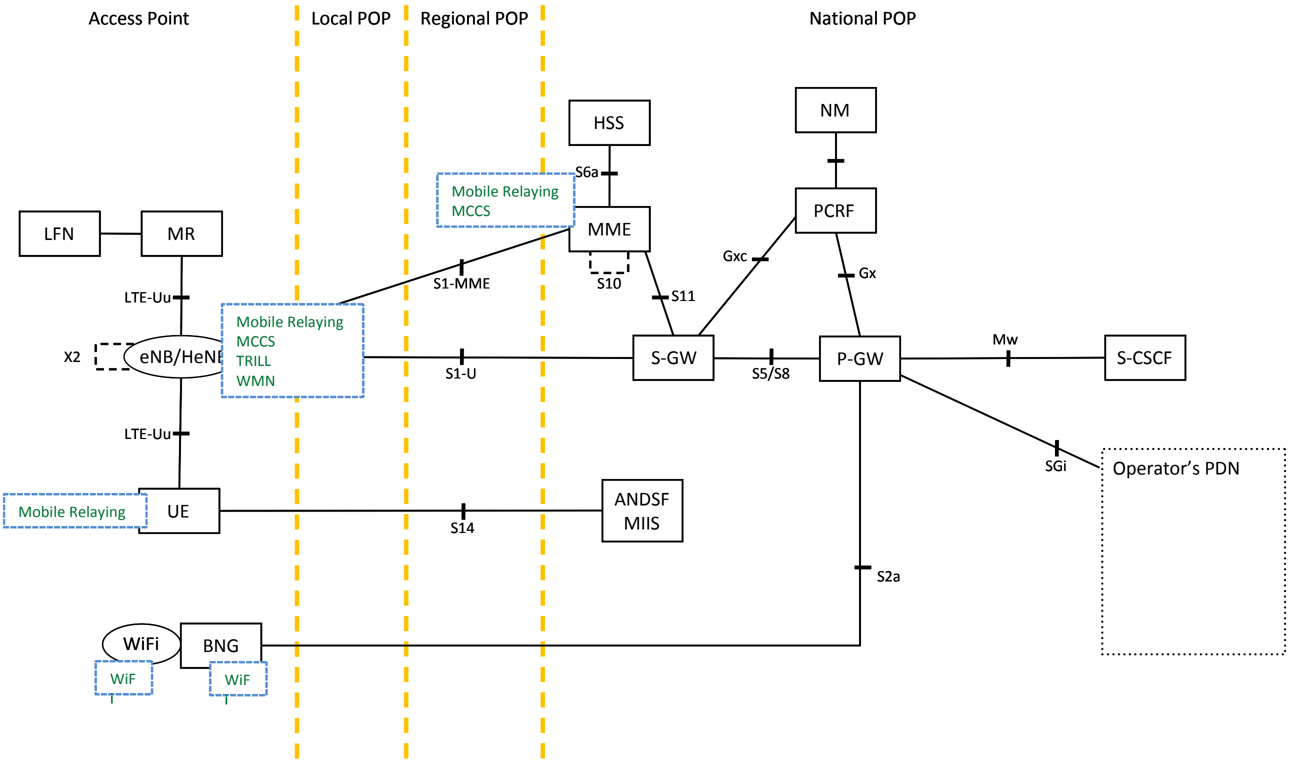

Figure 12 Centralized architecture for backhaul capacity improvement

6.4.2 Distributed architecture

Functional view of distributed architecture for backhaul capacity improvements

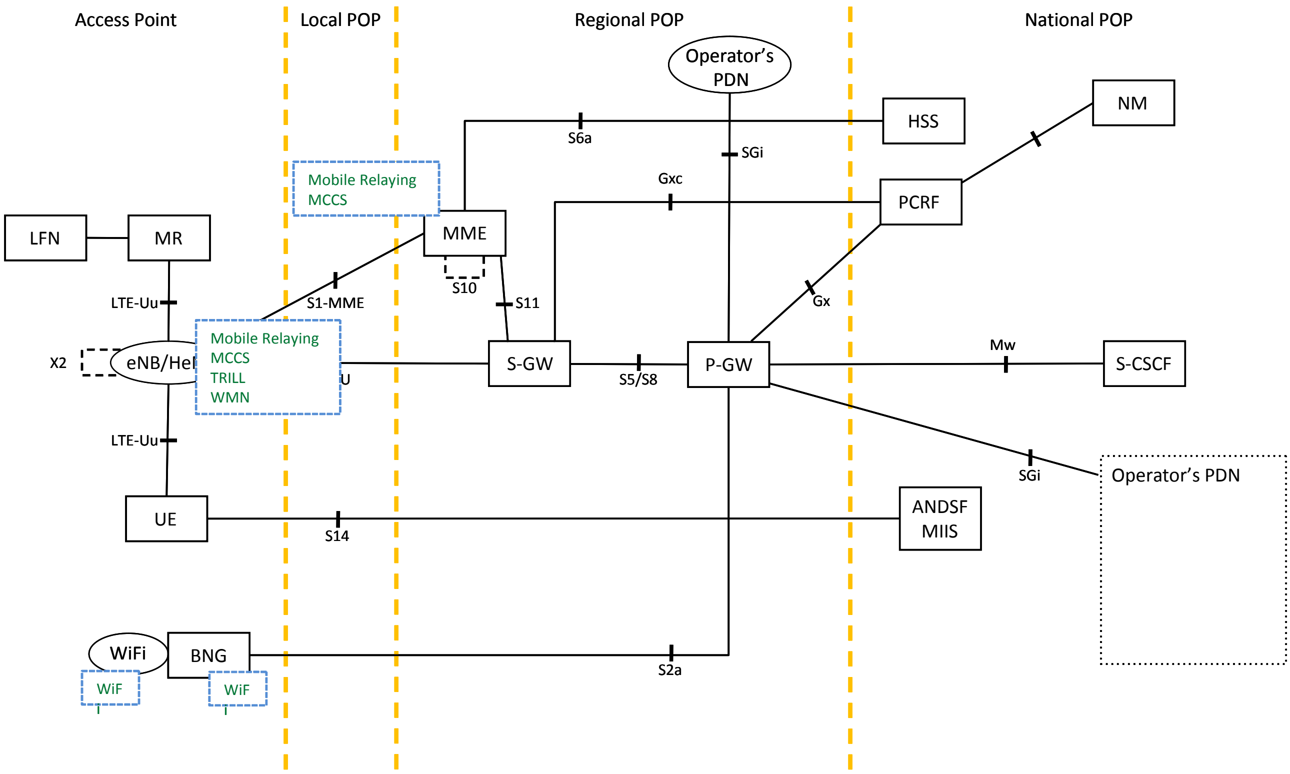

Figure 13 Distributed architecture for backhaul capacity improvement

6.4.3 Flat architecture

Functional view of flat architecture for backhaul capacity improvements

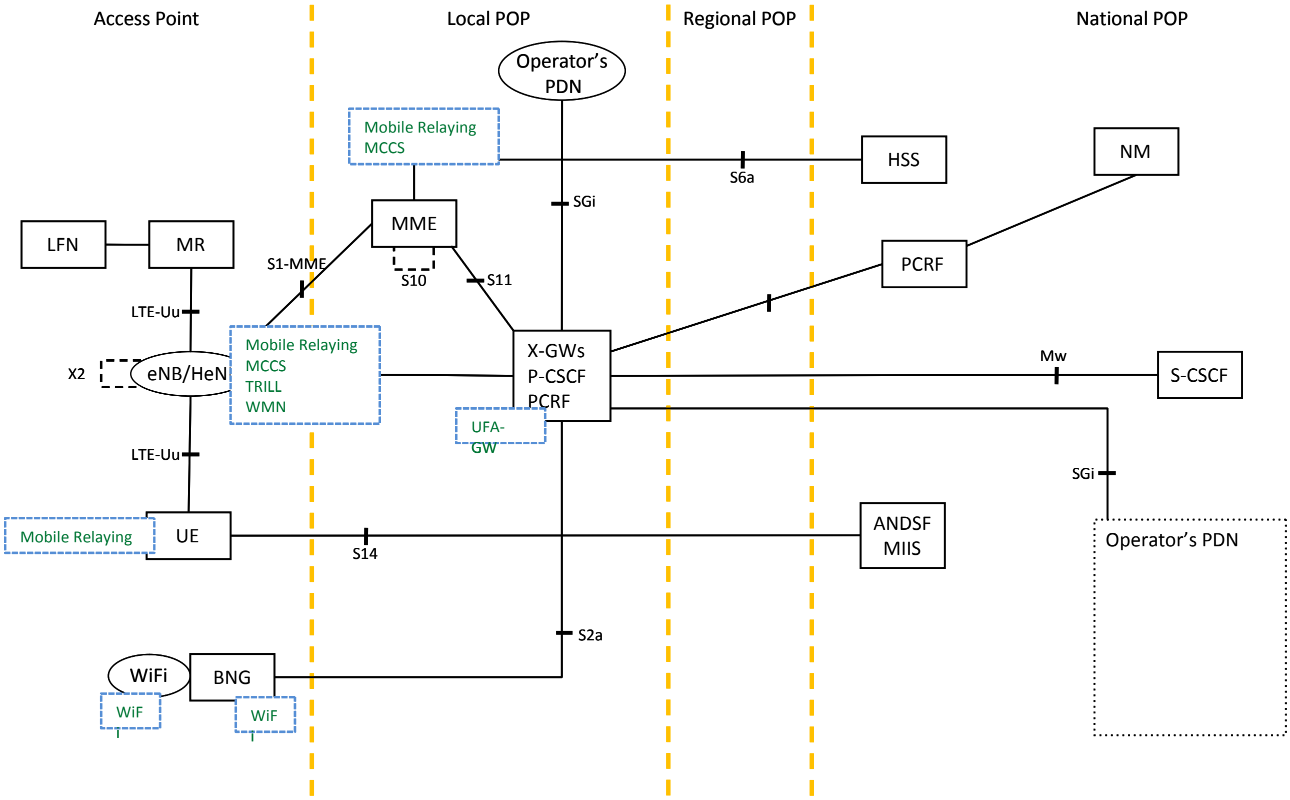

Figure 14 Flat architecture for backhaul capacity improvement

6.4.4 Rationale for using the involved technologies

6.4.4.1. WiFi

WiFi represents the most appropriate technology for offloading the wide area radio network towards the fixed wireless network. The technology is not topology sensitive in the sense that as long as there is a defined anchor point it will work. Operator can provide personal connectivity services for devices in residential network, e.g. firewall, content filtering, secure authentication. Operator partner services tied to mobile subscription can be provided also over WLAN behind RGW, e.g. Spotify. Operator (Fixed, Mobile or ISP) manages the WiFi AP. Mobile Operator can provide better indoor coverage for WiFi enabled devices.

6.4.4.2. Relaying

The Mobile relaying will add several tasks to the core network procedures to handle mobile relays that are moving over time. We will investigate signaling procedures, latency issues and different traffic management techniques between the core network elements and donor eNodeB, and between RN and donor eNodeB in detail. Moreover, we will also investigate the radio resource managements and the throughput gains in the radio core network side and its impact on EPC structure for mobile relaying support in LTE-Advanced. In LTE EPC, mobile relaying is proposed to be related to MME and eNodeB structures.

6.4.4.3. Ultra Flat Architecture (UFA)

The fully distributed architecture scenario suggests PCC and IMS nodes are distributed in addition to S/P GW and MME. The mobility implies the change of all these nodes. In order there is low impact on handover performance, the UFA concept simplifies the concatenated equipments into a single GW, and adds a proactive step and a network controlled handover execution.

UFA is flat and introduces distributed signaling and data anchors, which are the UFA_GWs and the SxS_GWs. This enables to better distribute the traffic load, unlike the centralized anchors. UFA_GWs distribution enables to distribute the S-CSCF and the Application Servers, which enhances their scalability and reduces the delay for accessing Application Servers content.

UFA is a flat architecture based on SIP. As for current mobile networks, it implements IMS and policy control functions. However, it is constituted of a single layer implementing these functions. It reduces the number of node types and interfaces, and only requires distributed and temporary anchors, instead of centralized ones.
The main idea of UFA is to gather as much information as possible into one Gateway and exchange information with another Gateway. UFA contains the I-CSCF, S-CSCF and the HSS nodes and two new nodes that are the UFA Gateway (UFA_GW) and the SIPcrossSCTP (SxS_GW) Gateway. The basic UFA_GW function is providing physical connectivity to users (capacity, coverage). Connectivity is a vital network part and one of the most dimensioning criteria.
The UFA_GW is the main UFA node; it gathers classical IP-AN nodes functions (e.g. NB, RNC, SGSN and GGSN functions for UMTS), policy control functions, P-CSCF functions, SCC AS functions and new functions. The SxS_GW is in some cases necessary to handle the case of non-SIP native services. UFA performances are measured for services transported over SCTP. When data is lost due to handover, SCTP considers that these losses are due to congestion and retransmits them after a timeout, causing high handover delay and resource use degradation. UFA solves these issues. Its performances are compared to the most known solution handling the mobility of these applications.
In case of data growth, UFA_GWs will be duplicated to satisfy the connectivity criteria.
Most technologies should be applicable on UFA, especially PMIP, because we have worked on its adaption on UFA.

6.4.4.4. Multi-Criteria Cell Selection (MCCS)

In next generation networks the architectures are evolving to include cells of different coverage to increase the end user data rate. Mainly, there are two types of cells deployed in hierarchical manner. First type is a wide area cell (macro cell) that provides moderate date rates for users with above average mobility, and the second type is a local area cell (micro/femto/small cell) that covers a limited area for limited mobility or nomadic users. In case of such deployments end users can access more than one cellular coverage with different load levels, and they have the opportunity to select not only the local cell but also the wide area cells. Previous UE initiated SNR based techniques cannot provide optimal cell selection since those techniques do not consider utilization of the candidate cells.

Using small cells in the network increases the end-user throughput but it brings extra handoff. Each handoff requires extra signalling and may cause connections failures. Many metrics such as signal strength, distance, signal-to-noise ratio (SNR), bit error rate (BER), traffic load, quality indicator and some combination of these indicators can be used in order to decide if the handoff is required or not. Therefore, proper design of cell selection criteria is a must to achieve efficient load balancing and minimize the number of handoffs in next generation networks.

We investigate several MCCS techniques and try to devise improvements or new algorithms. We further investigate the architectural impact of using MCCS in next generation core networks. In LTE EPC, MCCS is proposed to be deployed in MME and eNB.

6.4.4.5. Self Organizing Network (SON) solutions in EPC

SON collects automated network management solutions that are operating autonomously in the network. This includes Self-Healing, Self-Configuration and Self-Optimization features. The goal of these automated management solutions is to decrease the costs related to operating of the mobile network. Definition of SON solutions for LTE radio access network is considered as a major issue in 3GPP since Release 8. However, the transport impact of the radio SON solutions and SON solutions for mobile backhaul/transport networks is out of scope in those investigations.

Transport SON solutions and solutions complementary to the radio SON features enables the efficient utilization of the mobile backhaul resources. For example, an additional feature making the Mobility Load Balancing algorithm's operation transport aware can avoid radio SON actions being optimal for the radio part but at the same time being non-optimal from the backhaul point of view. The self configuring and self optimization algorithms operating on the EPC nodes and/or in the mobile backhaul aim at improving resource utilization and shall cope with issues like:

- Multitude of alternative transport solutions and options
- High number of transport and radio parameters
- Dimensioning/planning based on measured or predicted traffic/load
- Parameter configuration based on guidelines and recommendations
- Static parameters not capable to adapt to the changing conditions

These listed examples result in non-optimal system operation, inefficient resource usage and difficult management. To overcome these issues the SON features defined for transport networks shall provide solutions for the automated tuning of transport related parameters and automated connection setups in the radio access and transport devices. The scope of the automated self-configuration and optimization algorithms in centralized and distributed SON operation scenarios is to reconfigure or adapt the system configuration in order to follow the changes in traffic/load. These algorithms and solutions should ensure the consistent, efficient, adaptive and optimized configuration of mobile backhaul.

Different SON architecture variants

Centralized SON architecture

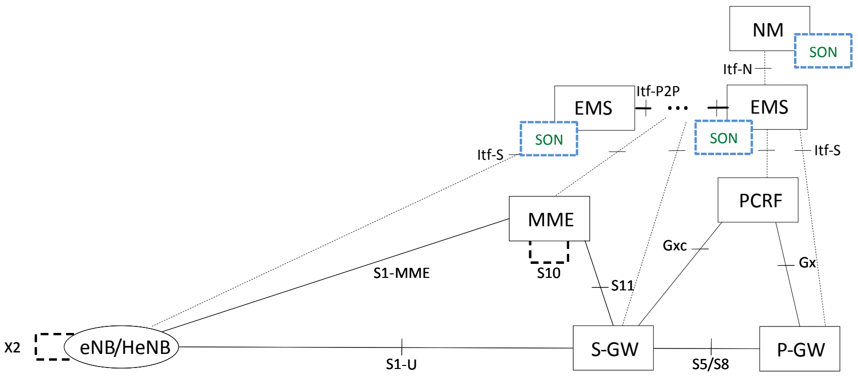

Figure 15 Centralized SON architecture

Distributed SON architecture

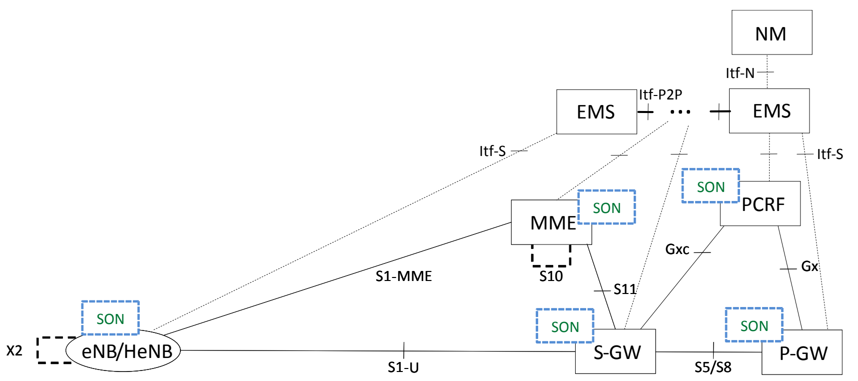

Figure 16 Distributed SON architecture

Hybrid SON architecture

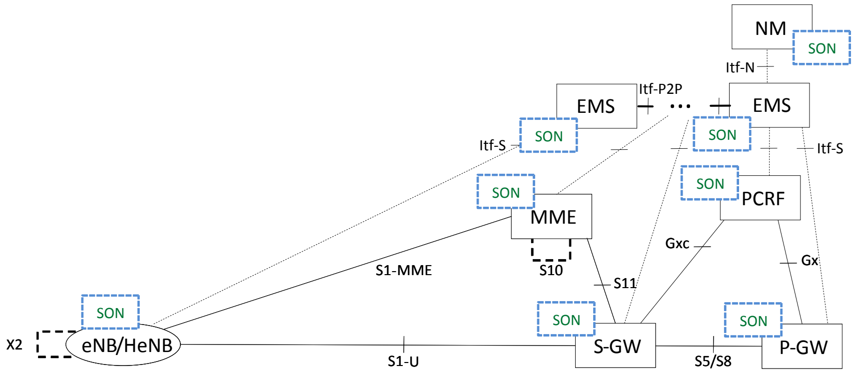

Figure 17 Hybrid SON architecture

6.4.5 Performance validation

Performance validation is performed in two ways in MEVICO. Firstly, there are individual technology validations which are described in the validation documents of the work packages. Those validations are important because they will show the benefits of technologies compared to current EPC Release-10 technologies.

Secondly, in WP1 the objective is to investigate common performance gains of technology combinations, and as a result of the analysis, provide directives for technology choices in different scenarios of the future 3GPP architecture.

We have investigated each technology we are dealing with in the MEVICO project and have defined dependencies between the technologies. Dependency means that the performance gains in terms of a given KPI of a technology, benefits other technologies as well.

If the common usage of several technologies results in enchained performance gains, those will form a specific group that should be evaluated in detail during system validation regarding performance gains and common behavior. We defined evaluation topics where all technologies are present that form dependency groups and contribute to chosen KPIs.

On the other hand, certain technologies may have incompatibility problems or may cover the same functionalities. This leads either to possible architecture alternatives or to different use cases depending on which combination of technologies is in use for a set of functionalities.

6.4.6 KPIs

This section focuses on improving backhaul capacity.

KPI 1.1: throughput gain in 3GPP access and backhaul

For this KPI we will need some assumptions on the topology of the backhaul network. For example n is the number of eNodeBs/APs, m is the number of SGWs/ePDGs/PGWs, one eNodeB connected to multiple GWs (GWs' polling areas overlap), neighboring eNodeBs and neighboring GWs have "direct"/"optimized" connections in the transport network)

Involved technologies:

- ANDSF (WP2): access network selection. It influences that on which of the n base stations/access points the demand arrives. Leads to better usage of access networks
- MCCS (WP4) and SON Mobility load balancing (WP5). They influence that on which of the n base stations the demand arrives. Lead to better usage of micro cells and higher RAN throughput, hence traffic demand increase for the backhaul per base station.
- DDMM, DMA-GTP, GW selection, Inter PGW. They select the PGW (and one of the m SGWs) where the demands exit from the backhaul
- Mobile relaying (WP2). Causes cell throughput gain, hence demand increase for backhaul per base station
- WMN, TRILL: The involved transport technologies determine the capacity of backhaul and aggregation links
- CES: decreases demands per backhaul/aggregation network edge point because it filters out unwanted traffic

Femtocells:

- HIP-FEMTO: Adds IPsec overhead between HomeNb-s and GWs in the aggregation or core network. It causes a constant multiplication factor for the throughput demand on the backhaul.

Multimedia broadcast:

- Multicast support in RAN and backhaul. It decreases traffic demands in case of multimedia broadcasting or other point to multi-point communication
- MSO: in case of quality problems, decrease the throughput demand of media streams to a lower rate by cross-layer interaction with the application and monitoring the available network resources.

P2P:

- mP4P has influence on P2P traffic demands in the backhaul part. Demands are to be localized to one network part.

multi-access devices:

- offload from the RAN and backhaul. Significantly decreases traffic demand on the 3GPP-access and backhaul. see KPI 2.1 for technologies and details, this is analyzed there.

HIP-auth, HIP-UFA:

- Adds IPsec overhead to user traffic

KPI 1.2: backhaul and RAN influence on E-E delay

We plan to analyze expected gains in typical scenarios. The gain

- mainly depends on the connectivity and backhaul technology: WMN, TRILL, SON self-management
- the influence of Mobile relaying (there are multiple wireless hops!)
- Path length in the backhaul and aggregation network depends on
  - network selection: ANDSF, MCCS, Mobile Load balancing, Mobile relaying
  - GW selection influenced by: GW sel, Inter PGW, DDMM, DMA with GTP

KPI 1.3: reliability (response time to link failures, bootstrap time)

- WMN, SON, TRILL (these are part of validation plan in WP3)

KPI 1.4: fair load distribution

Fairness definitions are needed

What is the common influence of these technologies on balancing the backhaul load?

- SON self management in the backhaul and RAN (how the parameters will be tuned, response time)
- Mobile relaying (adaptation time, maximum influence of technology on the load balance when very asymmetric load distribution arrives)
- MCCS and SON: Mobility load balancing (maximum influence of these technologies on the load balance)
- DDMM, DMA-GTP, GW selection, Inter PGW: (maximum influence of the technologies on the load balance)

6.4.7. Miscellaneous technologies

6.4.7.1 Cell on/off

The increasing amount of traffic is not uniformly distributed in time and spatial dimension. For energy saving purposes but also for traffic management solutions, the equipment on/off switching meets the requirement of improving efficient usage of network resources. A future objective is to consider in the decision of cell on/off switching the impact on the core network, i.e. backhaul utilization, load of distributed S-GWs.

6.4.7.2. Cell breathing

Cell breathing in the RAN causes redirections of traffic between neighboring cells. The variance of the traffic is mainly affected due to cell breathing in the S-GW. A future objective is to elaborate RAN TEHO decision strategies which also consider the impact of cell breathing on the core network.

6.4.7.3. NETwork CONFiguration Protocol (NETCONF)

Two protocols were standardised in the 80's and are widely used: Simple Network Management Protocol (SNMP) and Common Management Information Protocol (CMIP). The main problems with SNMP are that it has an elementary information model, uses unreliable UDP for transport and lacks transaction support. On the other hand, CMIP has more features than SNMP but the tendency is to adopt new approaches such as IETF Network Configuration Protocol (NETCONF). This confirms the trend towards standardised Web Services and XML/HTTP-based management that is generally more and more preferred by the network management community and industry.

NETCONF (RFC6241, RFC6242) provides mechanisms to install, manipulate, and delete the configuration of network devices. It uses an Extensible Markup Language (XML)-based data encoding for the configuration data as well as the protocol messages. The protocol operations are realized as Remote Procedure Calls (RPCs) that facilitate communication between a client and a server. The client can be a script or an application typically running as part of a network manager and the server is typically a network device.

Technologies such as an extension of DHCP (Dynamic Host Configuration Protocol) can be used for auto-configuration and EAP-SIM (Extensible Authentication Protocol - Subscriber Identification Module) for security parameter configurations. Configuration files can then be downloaded from a configuration server by using the NETCONF protocol, which provides an initial configuration for the network element.

6.4.7.4. Policy-Based Management (PBM)

PBM is seen as encapsulating business objectives which in turn are automatically applied to the managed systems, with reduced human intervention. Even though it is not as easy as it sounds, research and tests using PBM has gradually verified its enormous potential and showed that it can simplify complex management tasks of large-scale systems (see [9]).
In PBM, high-level policies (defined by the NM operator) are translated to low-level element operations for monitoring the network and automatically enforcing appropriate actions (through the policy control server located at the PCRF), such as modifying Traffic Handling Priority (THP), Allocation and Retention Priority (ARP), Maximum Bit Rate (MBR), Guaranteed Bit Rate (GBR) and QoS class identifier (QCI). Policies capture high-level management objectives and are the means to integrate self-management capabilities. The first standardisation efforts have been carried out by IETF: Policy Framework (RFC3198, RFC3460), Resource Allocation Protocol (RFC2753).

6.5 Architecture approach for improving the usage of heterogeneous network resources

The MEVICO Architecture will meet a need of harmonization of network with heterogeneous access and use. The architecture must allow for features such as selection of the most appropriate access network depending on the service / user / network requirements (instant or chosen by the operator). It should implement a seamless mobility between networks, relieve the traffic to the most appropriate network on the basis of relevant criteria. Solutions for increased efficiency must be integrated: multihoming, traffic grouping... The most relevant technologies introduced in section 5 will be integrated in the best way for that purpose. This section will define three architecture approaches based on centralized, distributed and flat topology deployment. MEVICO proposes the following technologies to improve the usage of heterogeneous network resources; WiFi, SCTP, Multicast, IFOM, ANDSF, UFA, NMIP, CES, MPTCP /UE. MPTCP/Proxy, HIP/Femto, PMIP-RO (Route Optimization) and PMIP-NEMO.

6.5.1 Centralized architecture

Functional view of centralized architecture for improving usage of heterogeneous network resources

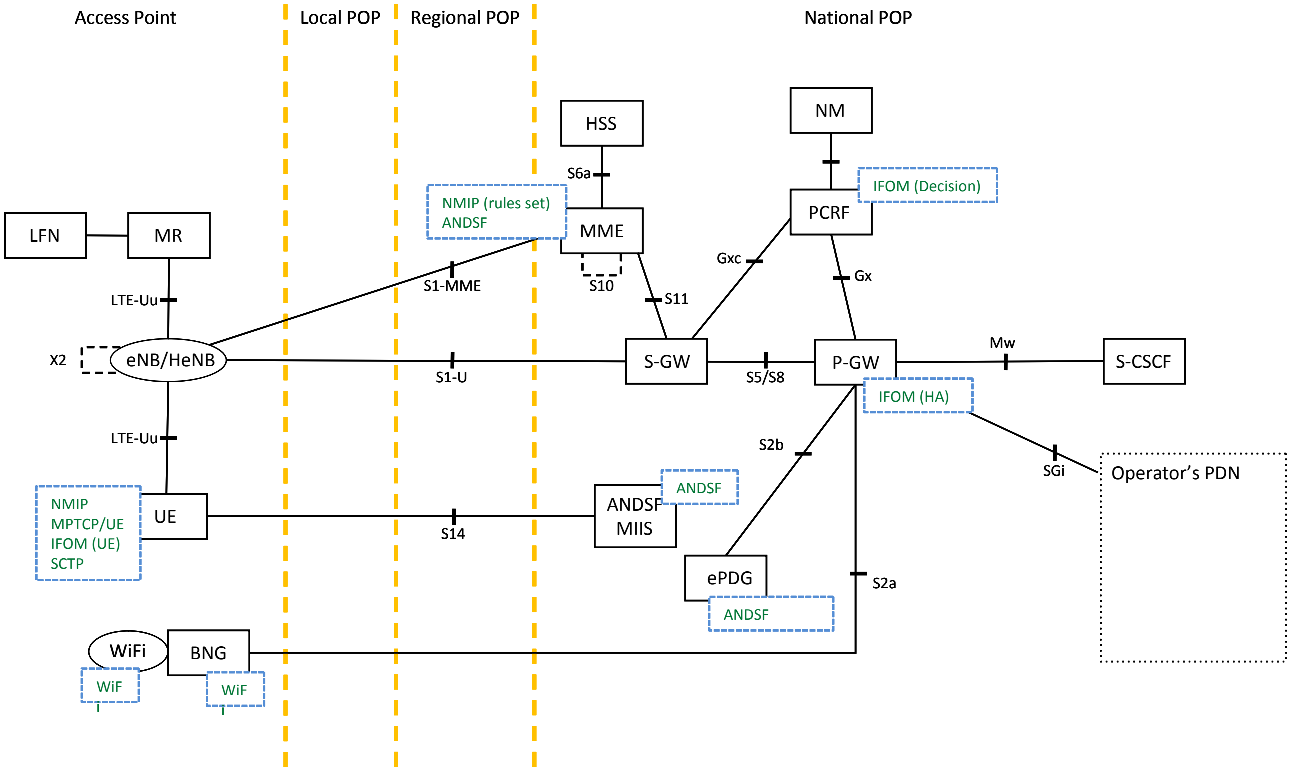

Figure 18 Centralized architecture for heterogeneous network resources improvement

6.5.2 Distributed architecture

Functional view of distributed architecture for improving usage of heterogeneous network resources

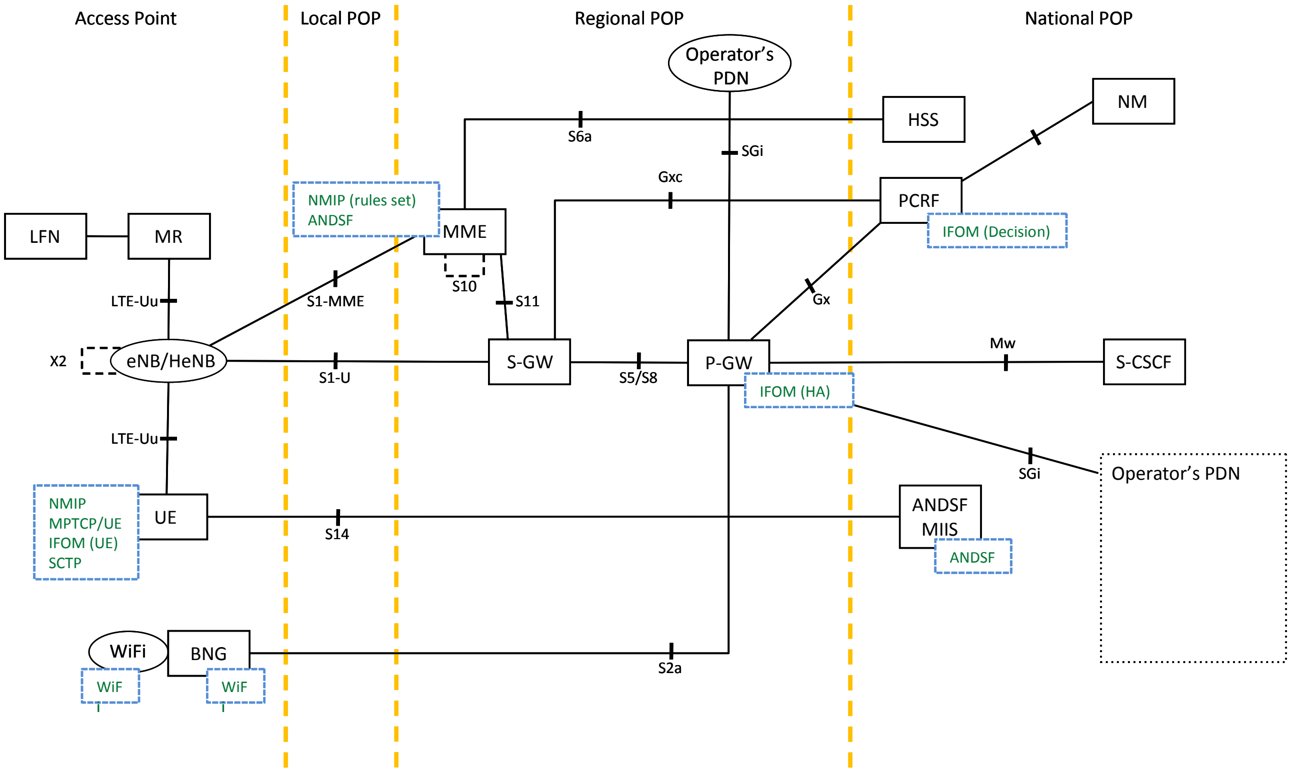

Figure 19 Distributed architecture for heterogeneous network resources improvement

6.5.3 Flat architecture

Functional view of flat architecture for improving usage of heterogeneous network resources

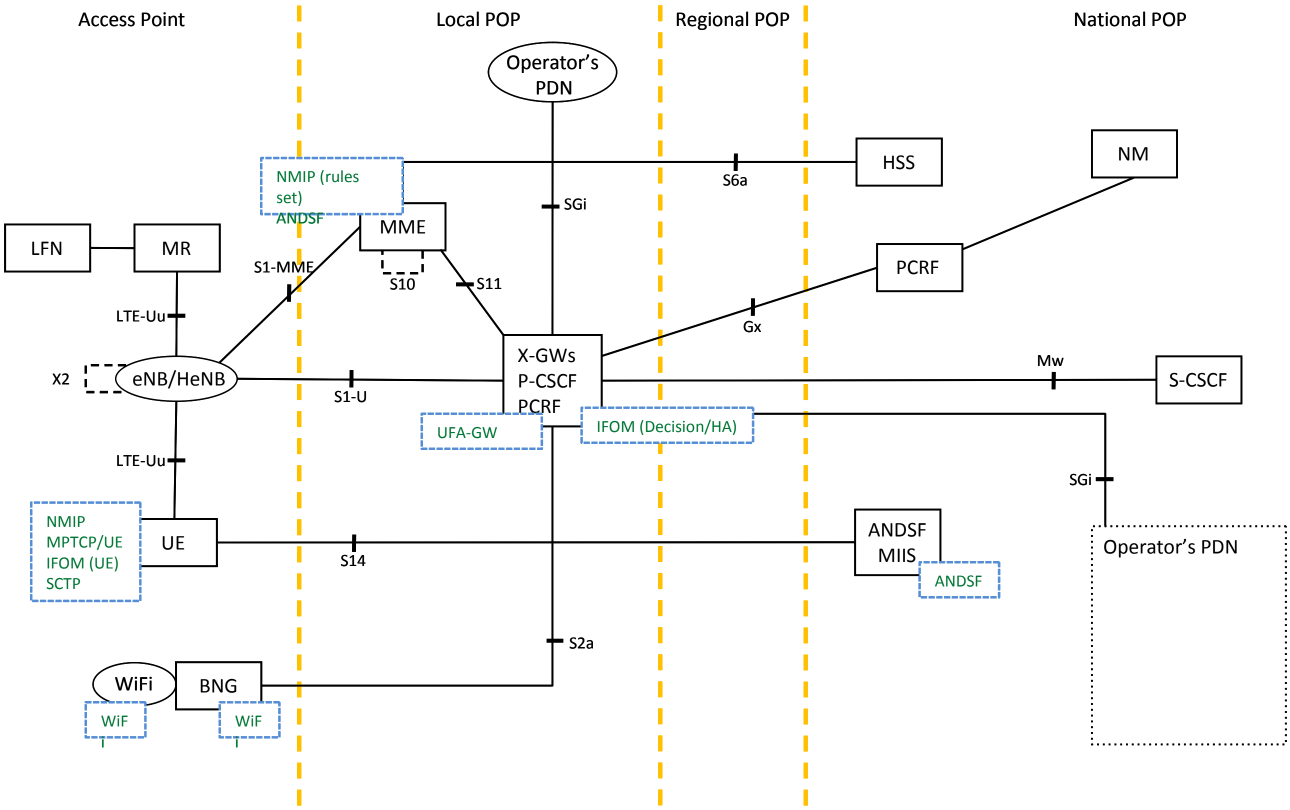

Figure 20 Flat architecture for heterogeneous network resources improvement

6.5.4 Rationale for using the involved technologies

6.5.4.1. Stream Control Transmission Protocol (SCTP)

SCTP provides a connection oriented reliable service and congestion control services, like TCP. Additionally it provides multistreaming and multi-homing that provides resiliency in case of path failure. It could be useful in case of data losses due to a mobility handled by EPC.

An extension m-SCTP could replace the PMIP, MIP or DSMIP protocols used in the EPC architecture, to handle user mobility, to bring a gain on the signalling plane.

It does not use any centralized server, and has few impacts from the topological choices but works better for flat architecture.

The main rationale for using SCTP as the transport protocol is the main features of the protocol. SCTP inherits most of its features from the most predominant reliable transport protocol on the Internet: the Transmission Control Protocol (TCP). Like TCP, SCTP provides a reliable, ordered transport service ensuring that data is transmitted across a network without error and in sequence. Furthermore, like TCP, SCTP provides connection-oriented communication and mechanisms to control network congestion. Prior to data transmission, a connection or, as it is called in SCTP parlance, association, is setup between the two communicating endpoints, and it is maintained during their entire communication.

One of SCTP's novel features is multi-homing. Multi-homing enables the endpoints of a single association to support multiple IP addresses. Each IP address is equivalent to a different network path towards the communicating peer, for sending and receiving data through the network. Currently, SCTP uses multi-homing as a means for path-level redundancy to provide uninterrupted service during resource failures, and not for load balancing.

Another novel feature that SCTP provides is the Dynamic Address Reconfiguration extension which leverages SCTP with mobility support. The Dynamic Address Reconfiguration extension enables SCTP to dynamically add an IP address to an already existing association, dynamically remove an IP address from an association, or dynamically request to change the primary destination address of the peer endpoint during an active SCTP association. Moreover, a local SCTP endpoint can influence its incoming network interface, by advising the peer endpoint about the destination IP address it should use for data transmission.

SCTP is to our knowledge the only transport layer protocol that actually supports a layer other than the application on top of it, and has a special field in its header that indicates the next protocol to receive the data. Other transport protocols like TCP and UDP do not provide this option and always assume that the next layer is the application layer.

SCTP requires support on the UE or the application, and do not need modifications on the network side. It can provide end-to-end anchorless mobility. Its performance is independent of the architecture.

6.5.4.2. IP Flow Mobility (IFOM)

IFOM is a special extension of IP mobility. It allows moving selected ongoing communication flows from one network to another, without any interruptions of the modified flows while keeping the other flows on their actual network. In 3GPP, UEs are able to simultaneously connect to 3GPP and WLAN accesses and are allowed to exchange different IP flows belonging to the same PDN connection through different access networks. The scheme makes possible to selectively offload some traffic (such as best effort traffic) to e.g., the WLAN segment while using UTRAN or E-UTRAN access systems for other traffic (such as traffic with higher QoS requirements). Some policies are introduced and the UE can be informed of them by ANDSF.

6.5.4.3. Access Network Discovery and Selection Function (ANDSF)

ANDSF can be used when either both 3GPP and non-3GPP accesses are available or when multiple non-3GPP accesses are available. The usage of ANDSF capabilities is intended for scenarios where access network level solutions are not sufficient for the UE to perform Network Discovery and selection of non-3GPP technologies according to operator policies.

The ANDSF contains data management and control functionality necessary to provide network discovery and selection assistance data as per operators' policy. It consists in sending information to UEs about available access networks and inter-system mobility policy.

Access Network Discovery and Selection Function features help to optimize the selection of the various radio accesses. This can be either 3GPP or non 3GPP (mainly Wifi) accesses. The scan which is the process to determine the available radio accesses and the quality of radio links is costly in time and energy. The radio accesses knowledge may be used to shorten the scan process and prioritized which radio accesses should be scanned first. As example, it is useless to scan Wifi Access Point where the user has no right to access.

In the 3GPP ANDSF version the operator provides rules, mainly dependent on the location of the UE, giving the radio accesses preferences of the operator. These rules are intended to be merely static.

The IEEE ANDSF version named MIH is based on an information service that allows the same kind of features. Furthermore the event service permits to have a network controlled HO.

This is a key feature for offloading techniques to be used by connection manager.

6.5.4.4. Not Mobile IP (NMIP)

NMIP links terminals such as mobile phones directly with servers, cutting out the need for tunnelling and reducing the network itself to simple switches. NMIP is designed as a light mechanism to provide an anchorless mobility management. It does not use any centralized server, and has few impacts from the topological choices. NMIP implementation will use a rules database that we have put on the MME.

NMIP is an extension of the TCP protocol that solves the problem of the connection break when the IP address of one of the correspondents changes. In a mobility context, it allows to manage the radio interface change. As the reconnection is a fast process, the HO is realized seamlessly. Events such as interface up/down are caught and may trigger change of TCP connections. A rule system is implemented to determine which TCP connections may migrate. This technique may be used to realize automatic offloading when one radio interface is a 3GPP one and the other is Wifi.

6.5.4.5. Customer Edge Security (CES)

CES analyzes incoming traffic to the mobile network and perform source based authorization based on trust information from the originating network. This allows reducing the unwanted traffic that reaches the mobile networks, thus improving the available capacity for legitimate traffic. t. CES allows users to utilize IP-based devices and applications without modification in a network that does not provide routed IP. Furthermore, CES adds security enhancements and allows incoming access to services (e.g. communication applications) in a private network.

6.5.4.6. Multipath TCP (MPTCP)

MPTCP is a modified version of the TCP protocol that supports the simultaneous use of multiple paths between endpoints and no centralized anchor is needed. As a consequence of the multipath the traffic is balanced on the available paths. Fairness is ensured on each path to avoid any starvation on one link. The throughput of the connection is then improved as transmitted data is sent simultaneously on several links. The reliability of the connection is increased as even if one radio link fails, the other links can cope with the data transmission.

6.5.5 Rationale for technology coexistence

Comparison between HIP, SCTP, NMIP and MPTCP protocols:

topic	HIP	SCTP	NMIP	MPTCP
Sec: Message Sender authentication	yes	no	no	no
Sec: Message weak authentication	yes	no	yes (nonce)	yes (tbc)
Sec: Message encryption	yes	no	no	no
Reliability: link fail over	no	yes	no	no
Locator/Identifier split	yes	no	no	no
Throughput optimization: link aggregation	yes (GW's are the aggregation points, anchorless in user plane if distributed or flat network)	no	no	yes
Seamless communication: HO delay	only L2 re-attachment delay to next L2 PoA in case of single-interface device		IP @ change + 1RTT	
Efficiency: payload data/total size (*)	IPsec BEAT mode: no overhead	An extra session layer header is added		
IP address change delay (**)	sent during handover preparation through the previous GW to the MN.			
HO delay after IP @ change	L2 attachment delay		1 RTT	
Multicast support	no	no	no	no
Middleware: NAT/FW traversal issues	yes, rfc 5207 sec 5	yes	yes	yes (if only one path is active)
Multihoming support	yes, rfc 5206 provides possibility, but policy-based flow mapping is not part of standard [1].	yes	For all flows (one interface per flow)	yes

Table 1 Comparison between HIP, SCTP, NMIP and MPTCP protocols

This section focuses on improving the usage of heterogenous access networks

6.5.6. KPIs

KPI 2.1 Offload gain due to the usage of multi-access capabilities:

Given proportion of UEs are multi-access devices.

- offload through WiFi: in function of the WiFi coverage part of the traffic demands can be offloaded from the 3GPP-access and backhaul. It decreases demands on the LTE/EPC. This validation is part of WP2 validation plan. It could be interesting to see the maximum gains of offloading. E.g., are there any traffic demands types which should never be offloaded?
- IFOM, NB-IFOM: enabler technology, part of the IP flows can be moved to non-3GPP access and backhaul. Hence it decreases traffic demands on EPC/LTE. However it only profits from the gains of non-3GPP coverage, in our case the offload through WiFi, so it may not be so interesting for this KPI only if it introduces further constraints to the effective offloaded traffic throughput.
- MPTCP (MPTCP-Pr in case of NEMO): enabler technology, offloads part of TCP traffic from 3GPP access and backhaul. It also profits from Wifi or othe offloading, so it may not be so interesting for this KPI. Only if it introduces further constraints to the effective offloaded TCP traffic throughput.

KPI 2.2 capacity aggregation

- MPTCP: enables high bitrate TCP sessions by capacity aggregation.
- MPTCP-Pr: in NEMO use case, MR could use MPTCP-Pr for the previous reason.

KPI 2.3 Handover delay

- ANDSF: influence scan duration
- GW selection, Inter PGW, DDMM, DMA-GTP: influences GW selection time and path length for signaling (e.g., re-authentication delay)
- Mobile relaying: influences signaling delay
- MCCS, SON Mobility load balancing influences path in the RAN and backhaul, and signaling delay
- TRILL, WMN: path length and delay to MME/GW
- L2 handover delay: TRILL (provides intra-GW mobility?)
- L3 handover, inter-GW mobility:

  depends on the technology combination:
  - PMIP-RO (1st alternative)
  - m-SCTP, NMIP (2nd alternative: part of traffic demands solve mobility on transport level. In that case mobility management should not hide the interfaces from the transport level)
  - UFA (3rd alternative)
  - HIP-UFA (4th alternative)
  - PMIP-NEMO (NEMO extension for 1st alternative)
  - HIP-NEMO (NEMO extension for 4th alternative)

Different technologies for the same purpose and different use cases will define the technology combination alternatives

Use cases can be:

- intra-GW mobility, inter-GW mobility;
- intra-access technology (3GPP or non-3GPP)
- inter-technology(within 3GPP-access, between 3GPP and non-3GPP accesses

KPI 2.4 Service interruption delay due to handover

- Investigation of which part of the handover signaling influences the service interruption time (relevant in case of real-time services)

- Investigation of handover preparation delays (to see if a preliminary stated maximum frequency of handovers is supported or not)

Use cases may be based on device type

- in case of single-interface device
- in case of multi-interface device (this is more important)

KPI 2.5 handover related signaling load on the network

- given in terms of signaling messages mapped to which part of the network

6.6 Architecture approach for improving Core network capacity

In order to address some of the bottlenecks identified in requirements and usage scenarios, the third high level objective to accomplish with the new mobile architecture is to improve the network capacity. The basic mechanisms identified towards capacity increase consist of firstly to optimize usage of network capacity so reducing unwanted traffic but also to decrease the signaling that is expected to increase when growing the mobility and number of users. Secondly, the routing should be optimized to reduce latency and having unnecessary signaling with large round trip time traversing the network. Instead, some of the functionality could be handled closer to the edge will reduce the traffic in the core network, thus increasing the overall capacity. MEVICO proposes using following technologies to improve the backhaul capacity; HIP/Femto NMIP, PMIP/RO, DMA, HIP/NEMO, Multicast, MPTCP/UE, MPTCP/Proxy, HIP/Auth.P4P (Proactive Provider Particip. for P2P), O&M and Inter P-GW.

6.6.1 Centralized architecture

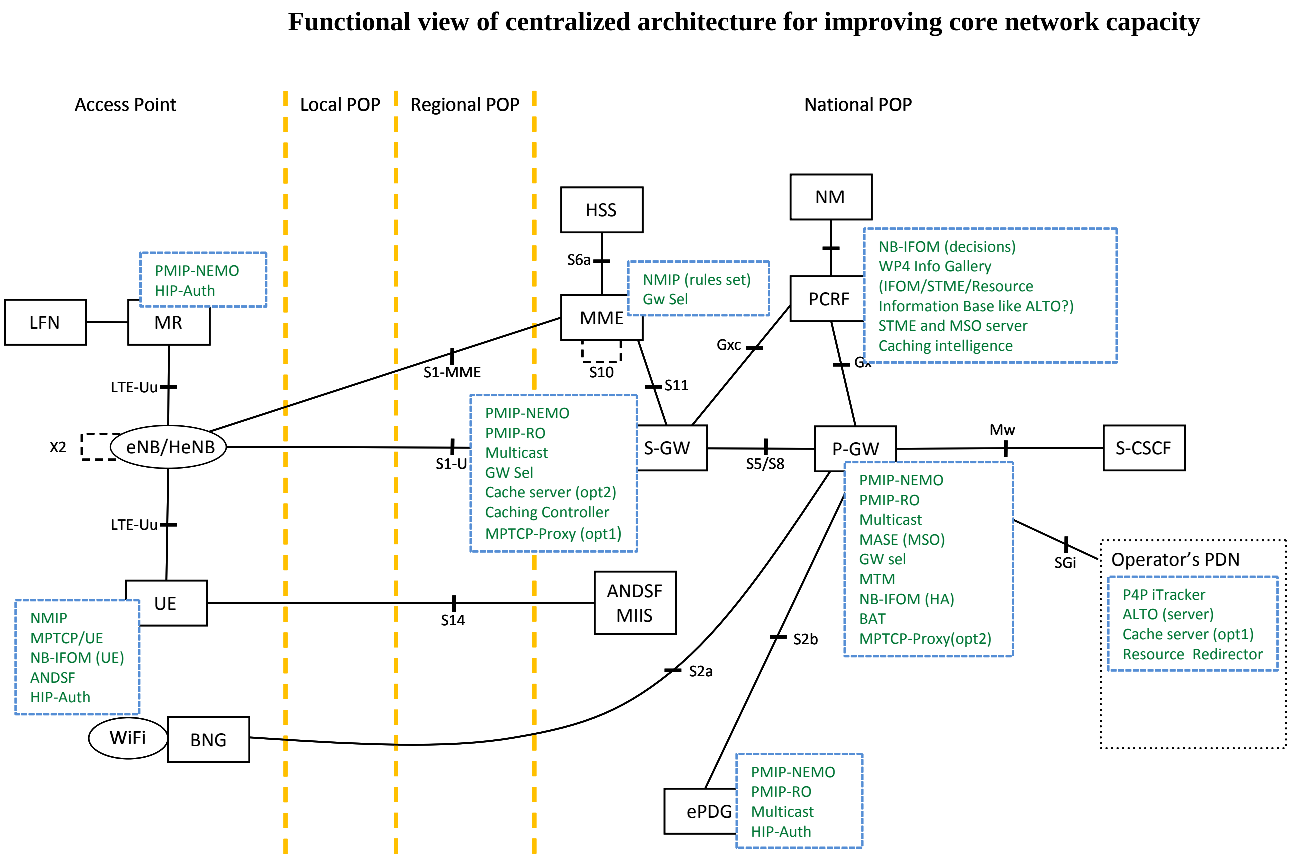

Figure 21 Centralized architecture for core network capacity improvement

6.6.2 Distributed architecture

Functional view of distributed architecture for improving core network capacity

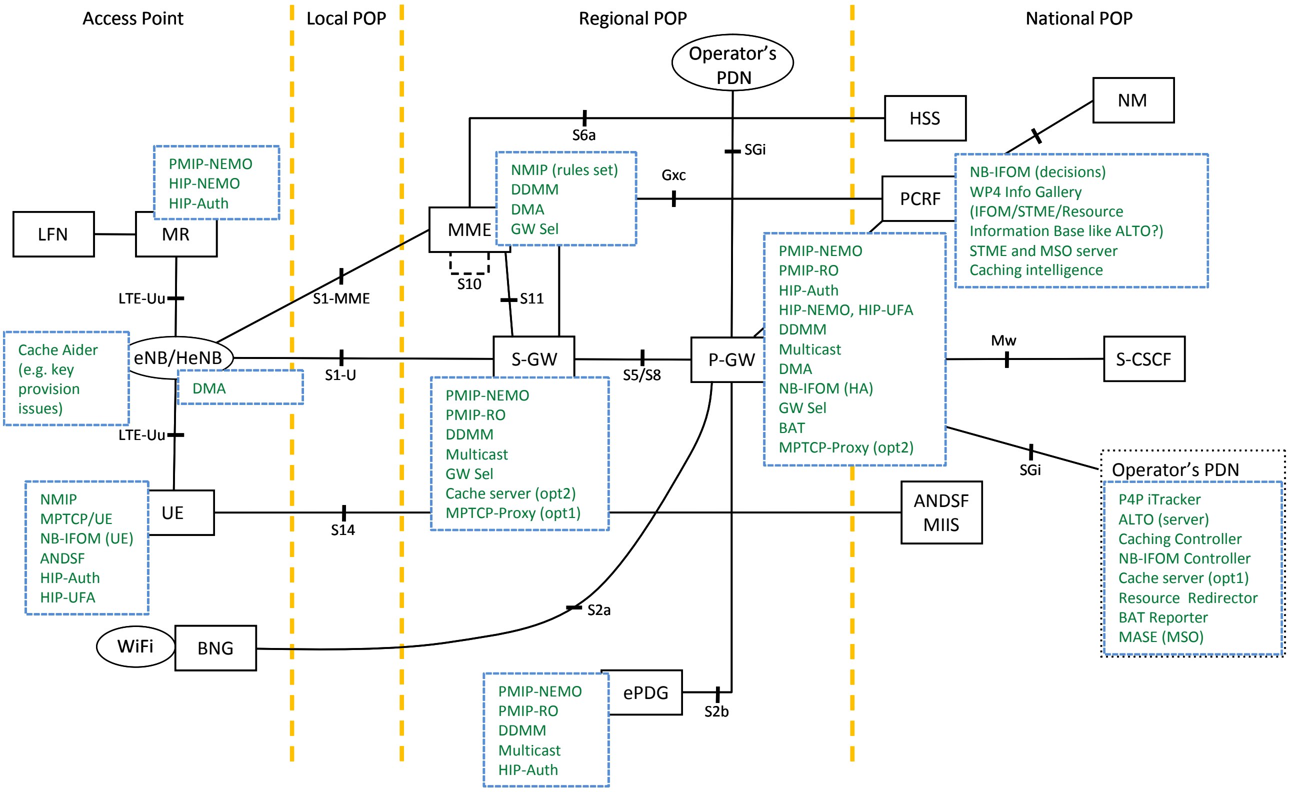

Figure 22 Distributed architecture for core network capacity improvement

6.6.3 Flat architecture

Functional view of flat architecture for improving core network capacity

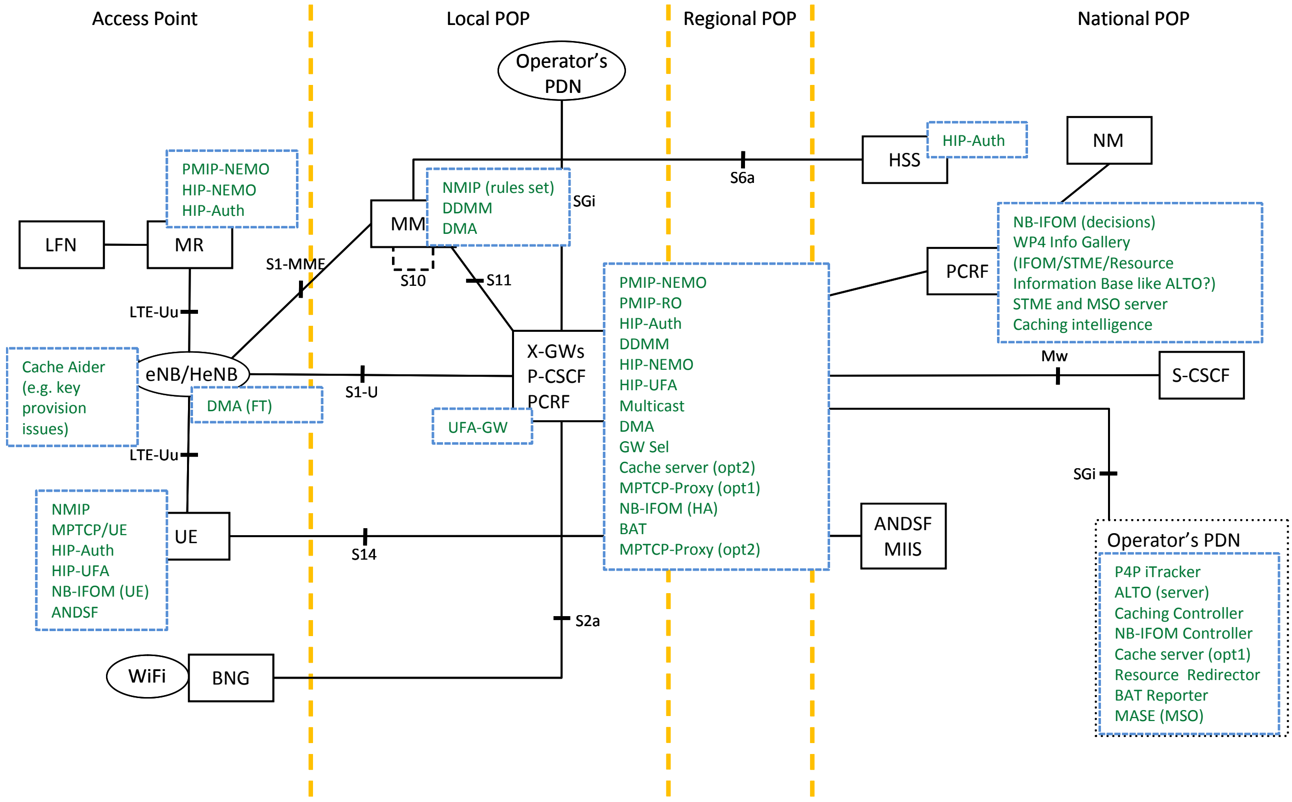

Figure 23 Flat architecture for core network capacity improvement

6.6.4 Rationale for using the involved technologies

6.6.4.1. Gateway selection

The optimum selection and reselection of core network elements might be dependent on the currently selected access network. A function which coordinates the different selection procedures might improve the overall system performance. To support traffic management also in distributed gateway scenarios it could be beneficial to include additional criteria into the GW selection process like:

Load of the transport network, mobility behavior of users (e.g. low mobile, high mobile), access networks supported by the GWs and the UEs…

6.6.4.2. Proactive Network Provider Participation for P2P (P4P)

P4P aims reducing the load on the network caused by regular P2P traffic. In the regular P2P transmission peers are selected randomly regardless of their location or the link costs. When using P4P, the network provider shares the network topology and corresponding cost map with the P2P application server. With this approach optimal routing (in the view of network provider) and lower end-to-end delay is obtained.

Mobile P4P (mP4P) is the next step in P4P research where the mobile end users share content in a LTE cellular network. LTE mobile environment causes further constraints to P4P such as existence of heterogeneous access networks, frequent joins and leaves of nodes, expectation of efficient signalling and lower complexity algorithms. mP4P will address these issues for optimal routing of content sharing within the LTE/EPC.

Reducing the backbone traffic and lowering the network operation costs are the two main rationales for deploying mobile P4P in LTE/EPC. The P4P iTracker can be easily deployed in the operator’s PDN without any hardware modifications in the EPC. The required network map is tracked and stored by the iTracker which shares this information upon the request of the P2P application server.

6.6.4.3. Network-based IP Flow Mobility (NB-IFOM)

The currently standardized IFOM (IP Flow Mobility) solution in 3GPP is strictly UE centric as the operator must firstly deliver the flow routing policies to the UE, and then the UE must provide these policies to the PDN Gateway. Also the ANDSF has no interface to the PCC system, therefore requires other ways to get informed about the updated flow routing policy for a particular UE. NB-IFOM (Network-based IP Flow Mobility) tries to eliminate the above limitations and create an operator centric flow management framework. The advantages of NB-IFOM enable operators to enforce IP flow routing policies without involving the UE first, such making able the PCRF (the central policy control entity) to decide on the flow routing policy based on e.g., the available resources in the network, before signalling the policies to the UE. The network-based solution is more efficient than the ones that rely on the UE to perform policy acquisition and enforcement: in the current, UE centric standard it is possible that the network context and resource availability may have changed by the time the UE provides the routing policies to the network; therefore the PCRF will not be able to authorize the new flow policies anymore. Such situations can be avoided if NB-IFOM is applied in the architecture.

6.6.4.1 Multimedia Streaming Optimizations

An important challenge is the transport of streaming applications in 3GPP mobile networks. In addition to the large amount of traffic generated, streaming applications may show a significantly varying video bitrate over playtime. New elements MASE (Media Aware Serving Entity) and STME (Steering Traffic Management Entity) enable the coordination of streaming application requirements and conditions with the bearer resource management in LTE. In addition bottleneck situation in the network can be communicated with the application higher in order to enable adaptation of the video playback to the changed conditions. The introduced components can contribute to a sustained QoE for the user and efficient management of network resources at the same time.

6.6.4.2 Caching improvements

In addition to “off the shelf” caching solutions, it is necessary to analyze optimal deployment strategies for LTE based on modeling and simulation of network, caching and content popularity parameters. Further optimization of caching can be achieved by considering chunk based storage, content diversity aspects and intelligent replacement strategies for the caching algorithms.

6.6.4.3 Resource redirection

Most widely used re-direction mechanisms used in current CDNs (server-client delivery) is based on DNS and HTTP protocols. These mechanisms use individual data sets. The envisioned technology aims to facilitate usage of common data sets over different re-direction mechanisms, influence of selection beyond the local MNO domain and detection of unfavourable resource usage.

6.6.4.4 Application Layer Transport Optimization (ALTO)

The IETF ALTO protocol provides guidance to content delivery applications in networks such as P2P or Content Delivery Networks (CDNs), which have to select one or several hosts or endpoints from a set of candidates that are able to provide a desired data resource. This guidance shall be based on parameters that affect performance and efficiency of the data transmission between the hosts, e.g., the topological distance. The ultimate goal is to improve QoE of the application while reducing resource consumption in the underlying network infrastructure.

While flow movements within the EPS can have an impact on the e2e path and its performance, there is no current way for decision elements within an EPS to anticipate it. Therefore it is necessary to find a way to integrate decision functions in the EPS with knowledge at the e2e scope. To improve its QoE for applications such as video download or streaming, the UE may use the ALTO protocol to jointly optimize the user QoE and the usage of EPS resources by providing the UE with information helping it to choose the best possible location from which to download the whole or piece of content while considering path changes within the EPS.

6.6.4.5 Bulk Analysis Tool (BAT)

With the introduction of LTE and smart phones, network management for data traffic is becoming a harder problem every day. Data traffic is increasing day by day, it is not easy to keep up with such fast change and the network operator cannot increase the capacity so fast. It is more important than ever before to observe the network and take necessary actions in terms of QoS per applications, hence optimize the network usage for improved customer satisfaction and still remain profitable.

Deep packet inspection mechanisms are being deployed at the operators, however the amount of investment required to inspect all the traffic in detail is huge. Usually only the traffic which may be important for the operator's business is detected and the rest is not identified.

This shows that some kind of Bulk Data Analysis may prove to be very useful to classify the total data into applications so that the network operator can get to see big picture. This does not need to be done in real time and not the whole traffic needs to be analyzed, but some time periods can be selected to reflect to the whole week.

One way of doing this is configuring an offline DPI tool and running the bulk samples over this, however this would require very frequent configurations for

- new sources of existing applications
- new application types

Moreover the DPI may not be ready to handle all these changes. Therefore a method is required where traffic is classified into applications;

- VoIP
- VideoStreaming
- P2P
- Instant Messaging (IM)
- Gaming
- Web Surfing
- Other

by utilizing traffic characteristics which are common within the classes. Bulk Analysis Tool (BAT) is mainly proposed to carry this functionality.

6.6.4.4. Host Identity Protocol – Ultra Flat Architecture (HIP-UFA)

HIP-UFA refers to HIP-delegation service based IP mobility management. It provides secure inter-GW mobility. It provides mechanism for GW relocation (similar to P-GW relocation with DMA solutions).

The HIP-UFA technology fits well the use cases where mobility between distributed ePDGs (distributed architecture), X-GWs (flat architecture) must be supported, and HIP-auth technology is used for user access authorization. This technology is expected to decrease the number of HIP Diet Exchanges or HIP Base Exchanges in case of frequent inter-X-GW handovers when using HIP for user access authorization. It also removes data traffic anchors, the only anchors are the distributed GWs (first IP hop) of the Ues/MRs.

It could also be introduced in the centralized and distributed architecture on the P-GW. However in those cases L3 access authorization and the security overhead caused by HIP Base Exchange between the UE and the P-GW is unnecessary. Indeed, in that case there is no need for L3 authentication and IPsec SAs between the UE and the P-GW, since 3GPP EPC already has its AKA / SIM protocol to authenticate the user through by the MME.

The coexistence of PMIP and HIP-UFA has no benefits, but also has no technological constraints. The same UE could support both technologies in different network domains.

HIP-UFA and DMA technologies are alternative solutions for optimal GW locations but can’t be applied simultaneously.

6.6.4.5. HIP DEX AKA based authentication (HIP-auth)

The HIP-auth abbreviation refers to HIP DEX AKA based authentication of the UE or MR. The HIP DEX AKA method has the same role as IKEv2 EAP-AKA method in non-3GPP non-trusted access. It is responsible for L3 user access authentication between the UE/MR (initiator) and the ePDG (responder). Besides authentication a tunnel mode IPsec SA pair is established between the initiator and the responder.

The HIP DEX AKA authentication method is intended to be used by low-end devices that perform e.g. M2M communication via the EPS. Hence, it is intended to cause less signaling and computational overhead then IKEv2 EAP-

AKA. It also has slightly different security features than IKEv2 EAP AKA. From deployment point of view, both need support on the UE/MR and the ePDG.

In EPC Release 10 the user access authorization in 3GPP-access is performed by the eNodeB and MME as authenticator and authentication server, respectively. The authentication method is AKA or SIM. Access to IMS may be controlled by IMS AKA method, where the S-CSCF is the authentication server, P-CSCF is the authenticator. If IMS AKA is used, then an IPsec SA pair in transport mode is established to protect SIP signaling between the UE/MR and the P-CSCF. Access authorization from non-trusted non-3GPP access is performed using IKEv2 EAP-AKA or EAP-SIM. In that case the ePDG is the authenticator and the AAA server is the authentication server.

In long-term, in flat topology several functional entities, such as the eNodeB, MME, ePDG, P-CSCF might be co-located. Hence all the above mentioned authentication procedures could be realized using one unique L3 access authorization. In that case the X-GW could be the authenticator and it might provide AAA-server or proxy functionality as well. HIP-DEX AKA could be a solution for low-end devices for unique user access authorization.

However, we must note that when L2 access authorization is missing, then attackers are able to exploit L2 control messages for attacks and might be able to attach to the access point. This is true in current non-3GPP non-trusted access as well, if the operator of non-3GPP access does not protect its access point. In case of 3GPP-access and trusted non-3GPP access L2 protection is obligatory in current standards.

It depends on the operator’s trust model whether a uniform L3 access authorization suits his needs or it is better to defend against L2 attacks and to deploy AKA/SIM authentication for each access technology.

6.6.4.6. Host Identity Protocol – Network Mobility (HIP-NEMO)

HIP-NEMO provides mobility management for moving networks with reduced signaling compared to the case when each LFN should update their IP address. It provides similar functionality to PMIP-NEMO.

HIP-NEMO introduces a HIP extension with the concept of mobile Rendezvous Server (mRVS). The mRVS is required in order to handle moving networks. It is expected to reduce HO related signaling for LFNs in the moving network. This technology should be used when HIP-UFA provides inter-GW mobility management (see section on HIP-UFA) or when the MR is HIP-enabled and uses HIP for user access authorization (see section HIP-auth).

The coexistence of PMIP-NEMO and HIP-NEMO has no benefits, but also has no technological constraints. The same MR could support both technologies in different network domains.

6.6.4.7. Proxy Mobile IP – Route Optimisation (PMIP-RO)

The route optimization solution addresses the problem of centralized mobility anchoring in PMIPv6 to reduce the impact of triangular routing by using intermediate anchors closer to UEs. The objectives are twofold: reduce unnecessary load at the LMA and provide a set of methods that allows transferring the data anchoring role from the LMA to distributed servers. The transfer of role would allow having a moving functionality that would optimize routing within a PMIPv6 domain.

From a centralized to a distributed architecture, the number of heterogeneous RANs should increase leading to a more intensive use of PMIPv6 in the core network, i.e., S5, S8, S2a, S2b, interfaces. Optimized communications between RANs will be improved by using this technology.

6.6.4.8. Proxy Mobile IP – Network Mobility (PMIP-NEMO)

PMIP-NEMO is an extension to PMIP to support the movement of prefixes (not just single addresses) allocated to MRs (and subsequently used for configuring addresses on LFNs) at MR’s request. The current procedure of PMIP allows to allocate a “Home Network Prefix” (HNP) to a MH directly connected to the infrastructure. This HNP has to be used on conceptual “home link”, i.e., the network link between the MH and the MAG. The typical implementation of NEMO assumes that the MR requests a specific (set of) prefix(es) to be announced to the LFNs. Hence, the requested prefixes are not directly used on the “home link”. The support of such procedure is not specified in PMIP, which thus requires an extension.

Besides PMIP-NEMO, HIP-NEMO is another alternative to support moving networks in MEVICO. HIP-NEMO makes use of the infrastructure and procedures of HIP to support LFNs mobility. Here, the modification of the IPv6 address is a trigger to update the mapping between the locator and the identifier of the considered UE. Generally speaking, a PMIP vs HIP comparison can be made: HIP (with or without NEMO extension) can be considered as a solution supporting mobility by involving both ends of a communication flow (MN and CN needs to be modified for HIP), whereas PMIP does not require modifications on CN to support mobility.

PMIP-NEMO and HIP-NEMO can be used on the same network infrastructure but the coexistence may not show any benefit. The support of NEMO by PMIP will hide modification of LFNs' IPv6 prefixes needed by HIP to update its mapping. However, both solutions are strongly tied to the mobility management protocol they inherit. Therefore, a core mobile infrastructure running PMIP will find advantages relying on PMIP-NEMO in the same way than an infrastructure supporting HIP will find benefit relying on HIP-NEMO.

6.6.4.9. Distributed and Dynamic Mobility Management (DDMM)

The IETF is currently working on deployment of MIP/PMIP in distributed architecture: DDMM work item. One of the purposes of dynamic mobility management is to provide mobility support only to those applications and to those MNs that really need it. Another purpose of DMM is to avoid providing mobility support for IP session beginning and ending while the mobile node remains attached to the same point of attachment.

6.6.4.10. Distributed Mobility Anchoring (DMA)

DMA technology in here intends to optimize the EPC based on the ideas of the IETF DMA, but utilizing existing 3GPP protocols like GTP with as less as possible changes in distributed architecture.

In a centralized only architecture these optimizations are not needed. In the flat architecture the UFA GW can be seen as a central GW. If applying GW functions in the flat architecture in the eNodeB the proposed solutions may result in un-proportional high signalling overhead. Hence it works best in a distributed architecture with distributed GWs.

Further benefits can be achieved if the network control functions are even more centralized and the distributed GWs contain less functionality and can contribute to further savings in HW spending.

In general 3GPP has assured the coexistence of PMIP and GTP. The question of co-existence has to be looked more for the suggested technologies and concepts rather than looking for PMIP-GTP co-existence only.

There is the proposal of tunnelling traffic between MAGs (SGWs) bypassing the LMAs (PGWs) and traversing over an intermediate anchor developed for the PMIP protocol (PMIP-RO). Another proposal is to change PGWs using intelligence in the PGW or changing SGW for routing optimization (DMA).

The DMA proposals with PGW relocation may not coexist with PMIP-RO as they provide different solutions for the routing problem: The PMIP-RO solution provides tunnel modification while keeping the UE IP address, the other solution is to select IP (PDN) connections in the PGW for what a new IP address and service interruption may be acceptable from application point of view and force a reconnection that allocates a new more optimal PGW and new IP Address.

It is clear that these are alternative solutions for optimal GW locations but can't be applied simultaneously.

The proposal to relocate the SGW to achieve maximal SGW-PGW collocation and optimal routing (DMA) could also coexist with PMIP-RO, if MAG changes are possible and may also result in MAG-LMA collocations.

6.6.4.11. Multicast

We oversee a multiplication of different types of RANs interconnected by distributed EPC entities closed to mobile users. However, some RANs or even mobile users' terminals do not support natively multicast, which is a key technology to optimize core network performance for group communications. The proposed technology aims at overcoming such limitations by providing the necessary network transport functionalities to optimally support group communications throughout the network despite de partial deployment of multicast in such a network. These functionalities do not introduce any modification on user terminal.

6.6.5 Rationale for technology coexistence

Here is described what is the reasoning or benefit of having all the technologies together and whether there are some dependencies or coexistence problems. E.g. if network supports NEMO then NEMO based mobility might be preferred to HIP mobility, etc).

6.6.5.1 Integration of Traffic Management Building Blocks

To achieve resource utilization efficiency for every possible scenario, network operators are on the verge of deploying an integrated traffic management framework aiming to cover all the traffic management building blocks and to organically integrate them into the architecture. In MEVICO the main goal of WP4 is to define the main elements of this framework and to investigate the possible schemes for their efficient interoperation. The analysis of the cooperation of different traffic management solutions and schemes is of particular importance as in some scenarios they might react counteractively. The work of WP4 already started to investigate these aspects. In the following sections we provide the first insights regarding the potential co-existence of several traffic management techniques.

6.6.5.1.1 ALTO and Resource Selection Service

Application Layer Traffic Optimization (ALTO) is the title of a working group within IETF. The motivation is to support an application with a set of candidate nodes, which host requested content. Even though the original purpose was to provide relevant information to P2P applications, the framework might be applicable to other context as well, such as CDN networks based on a client-server delivery concept. Preference for a specific location and the associated cost can be queried from an ALTO server by an ALTO client. That way the network view can be reflected within the selection process. The resource selection framework followed within the MEVICO project also addresses the problem of un-favourable selection of network resources. Hence the ALTO approach can represent a specific solution but the resource selection framework takes other aspects into consideration as well. Extending the concept of resource query to existing re-direction mechanisms, these have to support an ALTO client implementation. Understanding the impacts of specific extensions for well-known protocols, e.g. ALTO client support integrated into DNS server implementations need further analysis. On the other hand not all re-direction mechanisms follow the query model, which is suggested by ALTO. For instance anycast routers advertise information about preferred routes. Thus ALTO can be considered as a specific solution for the overall problem space of resource selection.

6.6.5.1.2 Possible conflicts of caching with other traffic management schemes in LTE

One possible issue is related to the candidate locations of caching servers inside the MNO domain and is strongly connected to the problem space of different MEVICO architecture proposals (i.e., Centralized, Distributed and Flat architectures). The most apparent location of caching nodes is beyond the GTP tunnel endpoint at the SGi interface (e.g., co-located with PDN-GW) or at the S5 interface. The problem here is that placing cache servers there could result in loss of connectivity to the cached material after handovers between 3GPP and non 3GPP accesses. Applying cache nodes at the S1 interface (e.g., co-located with eNodeB) might also be considered, but the benefit of accessing content closer to the users comes along with potential security issues (operators commonly use IPSec) or other limitations (e.g., deployment costs, problems of outdoor deployment). The above motivations make co-location of cache nodes within gateways to be also a promising option.

Besides to the possible architectural impacts, the co-location of caching and gateway nodes may also affect gateway re-selection based traffic engineering and/or mobility management mechanisms: if a cache server can be accessed only via a specific gateway node, gateway re-selection and/or mobility management would break the connectivity to the serving cache. Solving this problem would require a complex and presumably costly cache node session transfer protocol. Therefore these mechanisms can co-exist and work simultaneously, if gateway re-selection and mobility management does not force cache server re-selection of a running session.

Microscopic traffic management schemes and techniques (such as support of multipath communication based on IFOM and/or MPTCP) address operator-centric handling of individual flows. Even though content accessed from a cache node normally doesn’t have strict real-time requirements (in terms of latency) there still might be certain quality of experience requirements for the traffic flow associated with the requested cached content. It is therefore a challenge to synchronize the access to the cached content with EPS bearer management and possibly consider additional aspects such as user device characteristics, radio access conditions and other network parameters. Therefore a tighter co-operation might be required between caching and microscopic traffic management solutions at least for some type of cached content.

6.6.6 Performance validation

This section focuses on the improvements of core network capacity.

6.6.7. KPIs

KPI 3.1 Failure recovery time, reliability

- O&M technology improves reliability and failure recovery time of Carrier Grade Ethernet in core transport

KPI 3.2 E-E delay between UE and content or path lengths in terms of transport hops / IP hops

- all technologies which influence network selection and gateway selection influence path lengths from the UE until the GW (see KPI 1.2 for technologies). However note that
  - GW selection is highly constrained by backhaul connectivity between base stations and GWs even if polling areas of GWs overlap
  - path length is influenced by the topology (centralized, flat, distributed))
  - Hence GW selection mechanisms may not be important in this KPI
- IP mobility management protocol influences path length as well, but should not be investigated for this KPI because
  - for all IP mobility management technologies investigated in MEVICO (PMIP, HIP-UFA, UFA, NMIP, SCTP) the traffic will go through the distributed GW the UE is connected to on IP level.
  - hence only GW selection is important to analyze, IP mobility management technologies will not influence path length because they are distributed or anchorless.
- path length from GW to content: content location influenced by caching, redirection, ALTO, mP4P (different use cases for different applications)

Hence the most interesting question regarding KPI 3.2 is that is it possible to harmonize GW selection and content location selection

- In fact, this would not work just integrating them because they occur in different phases of UE procedures
- Currently, before session establishment there is APN selection, then based on the application type there is a redirection to content location.
- GW selection phase should already take into consideration application type and redirection possibilities.

KPI 3.3 analyze offload gains for core network equipments: e.g., number of active user contexts per network equipment

- might not be interesting because only depends on core network dimensioning)
- in fact this mainly depends on the number of network elements and active users, i.e. it is a network dimensioning question which would need very realistic and detailed examples.

- regarding IP-mobility management contexts:
  - there are multiple alternatives for mobility management.
  - all need contexts in the core network, except m-SCTP and NMIP
  - m-SCTP and NMIP will "offload" mobility signaling from the core network, BUT
    - these users have PDN contexts as well, hence the number of active user contexts per GW is not really influenced by m-SCTP, NMIP
    - m-SCTP and NMIP result "mobility management signaling offload" gain for the IP-mobility management network functions.
    - Initial reachability of UEs is not guaranteed through a home network prefix but other mechanisms are needed for a correspondent node to reach the MN through SCTP or TCP. This needs a functionality in the core network.
    - m-SCTP and NMIP enhance user experience, enable lower service interruption delay and enable shorter paths due to their anchorless property. These benefits are reflected in KPI 3.2, KPI 2.3 and KPI 2.4

There are different use cases for technologies (and different architecture alternatives). Maybe more specific KPIs could be analyzed as well related to different technology combinations.

7. Conclusion

Mobile broadband usage has exploded in the last couple of years due to improved network capabilities and affordable data plans. The mobile traffic increase globally has and will further be a consequence of several factors including: Growth of the mobile subscriptions due to population growth, improving living standards, evolution of the mobile networks, devices and services (affordability of capable/smart devices, enabled connection speeds, low cost flat rate data plans, easier usage, evolution of communication needs) and huge device increase potential due to new M2M connected subscriptions. People want to communicate, socialize, obtain/diffuse news and information and purchase digital goods with their devices on the fly, independent of the location and time. The evolution of the digital lifestyle drives consumers to use mobile subscriptions for the communication of data, the access to applications from their laptop, smart phone, tablet computer or even transparently based on M2M communications.

As the digital lifestyle goes mainstream, mobile broadband traffic volume will continue to increase making it necessary that the future network architecture evolution allow improved optimization to cope with these new demands. Video and other applications that depend on a high Quality of Service (QoS) are growing rapidly, placing new demands on the network capabilities. In order to manage the increased traffic and new applications with new requirements, LTE-EPC technologies have adopted an all-IP architecture that integrates a more distributed management and QoS strategy. This architecture simplifies the network stack, but can make the efficient operability more complex. Operators seeking to successfully deliver robust rich media data, voice and video services will need to measure and assure QoS in the all-IP network and this requires not only proper planning and network engineering but also a system that is robust, optimized and designed to handle future mobile data demand.

However, it is difficult to foresee the exact impact on the network due to new evolving applications; as it is impossible to foresee the discovery of new technologies that may improve the capacity of the current technologies and equipment. In MEVICO work there are not identified clear bottlenecks in the LTE architecture, but there are needs for high scalability and flexibility of the network capabilities due to potentially quickly evolving demand. Network needs to adapt and optimize itself to meet the ever-changing needs of subscribers, the services they use and the operational state of the network itself. Network shall become continuously aware of user traffic demands and the network resources that are available to serve those demands dynamically. Thus, improved architectural optimizations (both with respect to CAPEX and OPEX) have to be identified in order to ensure the sustainability of future mobile networks. For this reason, the MEVICO project intends to specify a network model optimized to maximize the end-user mobile broadband experience and ensure efficient congestion-free network performance.

MEVICO architecture document has firstly described the problem statement that mobile networks will face in the future. The challenges and requirements for the next generation of mobile networks are identified.

Secondly, MEVICO proposes a set of technologies to address those challenges. Those technologies are under study by the MEVICO partners and more detailed analysis of their benefits and performance are expected in the later phase of the project.

Thirdly, MEVICO architecture approaches describe how those technologies are planned to be deployed in the current Evolved Packet Core (EPC) network and how they could co-operate to provide efficient architecture evolution. Different network topology evolution scenarios and key network capability improvement items are used as basis of the analysis.

MEVICO architecture work will collect the results of the evaluation of the proposed technologies and the performance results when integrating the proposed technologies. Moreover, we will analyze the benefits of the new architecture compared to the implementation efforts (CAPEX/OPEX/evolution disruption etc.). Thus, based on the results of the performance improvements and CAPEX/OPEX evaluation we will conclude the potential network evolution visions.